Can we optimize an Early Warning System based on earthquakes to improve investments?

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis Econometrics and Operations Research

Name student: Menno Boor
Student ID number: 456065

Supervisor: dr. A.J. Koning
Second assessor: dr. P. Wan

Date final version: July 3, 2019

Abstract

In this paper we investigate whether we can optimize an Early Warning System based on earthquakes, proposed in Gresnigt et al. (2015), to improve investments. We simulate 1000 asset return series with a GARCH(1,1) model and with a GJR(1,1) model and use the most negative returns in each series to estimate parameters in an Epidemic-Type Aftershock Sequence model. These parameter estimates are used to estimate the probability of a crash in the next five days. We then propose an investment strategy that only invests in the assets for which the estimated crash probability is smaller than some threshold. We explore several ways to determine the optimal threshold and find that it is indeed possible to optimize our Early Warning System. This is done by defining crashes as the 95% quantile of negative returns and choosing the threshold that maximizes the average Hanssen-Kuiper Skill Score over all simulated returns. In particular, this strategy yields a Sharpe ratio that is approximately twice as high as the Sharpe ratio of a strategy that remains invested in all assets.

Contents

1 Introduction

Financial assets do not grow linearly over time, but at times they experience a rapid growth (known as a bubble) and at other times a sharp decrease (known as a crash). If an investors holds on to a large group of assets for a long time period, it is observed that she gets rewarded for the risk that she takes as those assets will grow on average. However, the gains that she makes from investing in those assets would be larger if she would be able to sell these assets when they are about to decline in value and buy them back thereafter.

It is thus beneficial for investors to know when an asset will decline in value, as it might allow them to take less risk and increase returns. On a larger scale, it can also be useful for policymakers to predict accurately when assets will decline in value, as they can change the policy in order to stabilize the economy when a crash is predicted to happen. It would further be an addition to the current literature if we are able to find an investment strategy that is able to sell assets when they are more likely to decline in value, because this would imply that the Efficient Market Hypothesis (a theory that states that all available information is already incorporated in asset prices) might not hold. Furthermore, being able to predict crashes of assets could provide insight into the market structure, which is also interesting from an academic point of view.

In the current literature, there are some Early Warning Systems (EWSs) introduced that indicate when a certain asset is likely to crash. One example is the EWS studied in Gerlach et al. (2018), in which the Log-Periodic Power Law Singularity (LPPLS) model is used to predict crash dates. Another example is the EWS introduced in Gresnigt et al. (2015), which makes use of the Epidemic-Type Aftershock Sequence (ETAS) model, a model that is often used for earthquake sequences. We will focus in this research on the latter EWS and we will use it to predict when a crash will happen. In Gresnigt et al. (2015), this is done by estimating the probability that a crash will happen in the next five days and then predicting a crash if the estimated probability is bigger than a threshold value of 0.5. In this research, we will explore whether it is possible to give more accurate predictions of the crash date by varying the threshold value that should be exceeded to predict a crash, and picking the threshold value that results in the most accurate predictions. We also want to investigate possible investment strategies that use the optimal threshold values that we hope to obtain. We thus want to find an answer to the following research question: can we optimize an Early Warning System based on earthquakes to improve investments?

To answer this question, we will adapt the EWS of Gresnigt et al. (2015) and apply it to different datasets; we first apply the EWS to S&P 500 data and later apply it to simulated asset return series. The latter allows us to draw more general conclusions. The main adaptation that we make to the EWS is that we explore several ways to find optimal thresholds for giving a warning. Using the simulated return series, we determine thresholds that (i) maximize the average Hanssen-Kuiper Skill Score (KSS, a measure of predictive power) over the evaluation period for all simulations, (ii) maximize the KSS of each individual return series over the evaluation period, and (iii) maximize the value of an investment in each return series over the evaluation period. Based on these different thresholds, we investigate investment strategies for individual assets and for portfolios of assets. The investment strategies only invest in the assets for which the estimated crash probability is smaller than the threshold. These strategies are then evaluated by the Sharpe ratio and Value at Risk, two common performance measures in the financial industry.

We hereby find that it is indeed possible to optimize our EWS. It is preferred to define crashes as the 95% quantile of negative returns and to determine the threshold for giving a warning signal as the threshold that maximizes the average KSS of all assets that an investor wants to invest in. We conclude this as this combination yields the highest average Sharpe ratio and highest average Value at Risk over all simulations. In particular, this investment strategy significantly outperforms an investment strategy that remains invested in all assets, as it yields a Sharpe ratio that is approximately twice as high.

Our research contributes to the current literature in two ways. First, it proposes a new method to rebalance asset holdings or construct portfolios. This method specifically looks at the possibility to sell certain assets when they are more likely to crash and buy them back when they are not as likely to crash anymore. Second, our research gives an indication that the Efficient Market Hypothesis does not hold, as we obtain significantly higher Sharpe ratios (thus significantly higher returns given the amount of risk taken) by only using information contained in past asset prices.

This paper is constructed as follows. Section 2 gives an overview of the literature that motivates why we will use an ETAS model to model crashes. In Section 3 we discuss which data we use in our research. We then motivate and explain the methods that we use in our research in Section 4, and describe our findings of applying these methods in Section 5. In Section 6 we give a conclusion of our research and answer our research question.

2 Literature

The EWS based on earthquakes is introduced in Gresnigt et al. (2015), as similarities between earthquakes and crashes in financial markets exist. They model extreme (negative) returns of S&P 500 data by an ETAS model. In this model, the arrivals of crashes are modeled by a Hawkes process, which is a nonhomogeneous Poisson process. This arrival process of crashes seems appropriate, as it is described in Sornette (2003) that crashes are often preceded by a positive feedback mechanism of noise traders following each other, and therefore crashes might be self-exciting. This self-enforcing component is captured by the Hawkes process, which would not be possible if we used a homogeneous Poisson process to model the arrivals of crashes. To test whether an ETAS model is appropriate to model financial crashes, it is tested in Gresnigt et al. (2015) whether the arrival times $\{t_i\}$ of crashes are generated by a Hawkes process. This evaluation technique is referred to as a 'residual analysis technique', as described in Ogata (1988). This technique states that if the arrival times $\{t_i\}$ of crashes are generated with a conditional intensity $\lambda(t)$, then the transformed times $\tau_i = \int_0^{t_i} \lambda(t)dt$ (referred to as the residual process) have the same distribution as a homogeneous Poisson process with intensity 1. A formal proof of this is left to the interested reader, and can be found in Papangelou (1972). It can therefore be derived that the transformed interarrival times $\tau_i - \tau_{i-1}$ follow an exponential distribution with intensity 1, if the arrival times $\{t_i\}$ of crashes are indeed generated by a Hawkes process with conditional intensity $\lambda(t)$. It is tested in Gresnigt et al. (2015) whether this is the case by a Kolmogorov-Smirnov test, with as null hypothesis that the transformed interarrival times are exponentially distributed with intensity 1. In their paper, they do not reject the null hypothesis that an ETAS model is appropriate to model financial crashes, when they define crashes as the 95% or 97% quantile of extreme (negative) returns.

The EWS of Gresnigt et al. (2015) produces positive Hanssen-Kuiper Skill Scores (KSSs) when applied to the S&P 500 data during the recent financial crisis, meaning that the rate of correct predictions is higher than the rate of false predictions. Moreover, they show in their paper that their modeling framework is able to predict the probability of a crash at least as well as GARCH-type models. Furthermore, they show that their framework is able to capture information from the stock returns that well known GARCH-type models cannot capture. They conclude this by applying the ETAS model to the standardized residuals of the GARCH-type models, which yields significant parameter estimates. Taking all this information into account, we conclude that it seems to be promising to model crashes of financial assets using an ETAS model.

3 Data

As we want to reproduce the results found in Gresnigt et al. (2015), we want to use the same dataset as they use in their paper. That is, we want to use the daily returns of the S&P 500 index over a period from 2 January 1957 to 1 January 2013. Here a problem arises however, as it is not mentioned in Gresnigt et al. (2015) which datasource is used to obtain these data. We downloaded the S&P 500 data from Thomson Reuters Datastream, from Bloomberg and from Yahoo Finance. However, these datasets differ from the dataset mentioned in Gresnigt et al. (2015); the dataset of Thomson Reuters Datastream starts not earlier than 31 December 1963, whereas the datasets of both Bloomberg and Yahoo Finance miss some observations in comparison to the data as described in Gresnigt et al. (2015). We make the choice to use the data of Yahoo Finance (daily closing prices, 14097 observations in total), as these are available from 2 January 1957 (until now) and do not contain 'NaN' values for particular time points. We then transform these closing prices into simple daily percentage returns $r_t = \frac{p_t - p_{t-1}}{p_t} \cdot 100\%$, where p_t denotes the index price at time t, so that we obtain 14096 return observations. To draw more general conclusions about results that we find, we will also simulate data. Which models we use to simulate data and how we choose those models will both be explained in more detail in Subsection 4.4.1.

4 Methodology

This section will describe and explain the methods and techniques that we will use to build, adapt and apply an Early Warning System based on the ETAS model. We will first explain what a conditional intensity is (in Subsection 4.1) and how parameters can be estimated with maximum likelihood estimation (in Subsection 4.2), so that the reader understands the underlying theory. We will then explain in Subsection 4.3 how to predict when a crash will happen and how to measure the accuracy of those predictions. We finally describe in Subsection 4.4 how we want to optimize the accuracy of the predictions and how we want to use those optimal predictions for investing.

4.1 Conditional intensity

The intensity at which crashes arrive at time t for a Hawkes process depends on the arrival history up to time t. We therefore say that the Hawkes process has a conditional intensity $\lambda(t|\boldsymbol{\theta};\mathcal{H}_t)$, where $\mathcal{H}_t = \{t_i : t_i < t\}$ represents the history of arrival times of events until time t. We will first explain what a conditional intensity exactly is. The conditional intensity (at time t) has to do with the probability that an event will happen in a small period of time after time t, given the arrival history up to time t (so with $\mathrm{P}(\text{event in } (t, t+\delta t)|\mathcal{H}_t))$. Taking the time increment δt small enough, this probability can be approximated by the conditional intensity $\lambda(t|\boldsymbol{\theta};\mathcal{H}_t)$ times the length of the increment δt. More formally, letting $\delta t \to 0$, we can define the conditional intensity at time t as

$$\lambda(t|\boldsymbol{\theta};\mathcal{H}_t) = \lim_{\delta t \to 0} \frac{\mathrm{P}(\text{event in } (t, t+\delta t)|\mathcal{H}_t))}{\delta t},$$

provided that this limit exists.

4.2 Maximum likelihood estimation

We will estimate the parameters in the ETAS model by maximum likelihood. The idea of maximum likelihood estimation is to estimate the parameters in such a way that the estimated model has the highest probability to have generated the observed data. This can be done by maximizing the joint probability density function, which is called the likelihood function (so $L(\boldsymbol{\theta}) = f(X_1, ..., X_n; \boldsymbol{\theta})$). As mentioned in Heij et al. (2004), it is equivalent (and often more convenient computationally) to maximize the log-likelihood function (i.e., log $L(\boldsymbol{\theta})$). If this log-likelihood function is differentiable with respect to its parameters, we can analytically derive the gradient of the log-likelihood and check for which values of the parameters this vector of first derivatives is equal to a vector of zeros. This gives us critical points, which are possible maxima. To verify that the obtained parameters are indeed corresponding to a maximum, one can calculate the Hessian matrix of the log-likelihood function. If this matrix of second derivatives is negative definite, then the critical points are indeed maxima.

To illustrate the maximum likelihood estimation of parameters in the ETAS model, we use the findings in Ogata (1988). In this paper, it is argued that the parameter vector $\boldsymbol{\theta}$ appearing in the conditional intensity of the Hawkes process can be estimated by maximum likelihood estimation. The relevant log-likelihood is given by

$$\log\, L(\boldsymbol{\theta}) = \sum_{i=1}^{N} \log\, \lambda(t_i|\boldsymbol{\theta};\mathcal{H}_t) - \int_0^T \lambda(t|\boldsymbol{\theta};\mathcal{H}_t)\; dt, \tag{1}$$

where $\{t_1, ..., t_N\}$ are the arrival times in the time interval $[0, T]$. These arrival times are determined by our definition of a crash; if we define, for example, crashes as the 95% quantile of negative returns, then $\{t_1, ..., t_N\}$ are the times at which the return was smaller than 95% of the returns in the estimation period. A (sketch of the) derivation of the log-likelihood in Equation (1) can be found in Rizoiu et al. (2017). We will illustrate how the maximum likelihood estimators of the parameters in this model can be obtained. To this end, we first describe how the conditional intensity $\lambda(t|\boldsymbol{\theta};\mathcal{H}_t)$ is defined, for which we use several equations from Ogata (1988). First of all, we use that

$$\lambda(t|\boldsymbol{\theta};\mathcal{H}_t) = \mu + \sum_{t_i<t} g(t-t_i) = \mu + \int_0^t g(t-s)\; dN(s),$$

where μ represents the constant term of the conditional intensity and $g(t - t_i)$ represents a self-exciting function (also known as a triggering function). For the self-exciting function, we choose $g(t) = ae^{-\alpha t}$, where a controls the maximum intensity of event triggering and α determines how fast the possibility that an event triggers another event decays. The relevant conditional intensity is therefore given by

$$\lambda(t|\boldsymbol{\theta};\mathcal{H}_t) = \mu + \sum_{t_i<t} ae^{-\alpha(t-t_i)} = \mu + \int_0^t ae^{-\alpha(t-s)}\; dN(s). \tag{2}$$

Plugging the conditional intensity in Equation (2) into Equation (1) and working out the integrals allows us to find an expression for the log-likelihood, which now becomes

$$\log\, L(\boldsymbol{\theta}) = \sum_{i=1}^{N} \log\Bigg(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\Bigg) - \mu T + \frac{a}{\alpha}\sum_{t_i<T}\Bigg(e^{-\alpha(t_i-t_j)} - 1\Bigg). \tag{3}$$

A full derivation of how this is done can be found in Appendix A. Now that we know an expression for the log-likelihood, we can derive the gradient and the Hessian. The expressions for the gradient and Hessian and their derivations can be found in Appendix B. As the gradient is non-linear with respect to its parameters, we cannot find an analytical expression for the maximum likelihood estimators of the parameters. We will therefore have to estimate the maximum likelihood estimates numerically.

Maximizing the log-likelihood can also be done entirely numerically, for example with the Nelder-Mead simplex direct search algorithm, which has been used in Gresnigt et al. (2015). Is is mentioned in their paper that they have tested several estimation methods and that this approach turned out to be most satisfactory. We also estimated the parameters in different ways; by (i) numerically solving for which values the analytically derived gradient is zero, (ii) using the Nelder-Mead simplex direct search algorithm (with parameter bounds), and (iii) using an interior-point algorithm in which we added first order information by supplying the analytical expression of the gradient. Using initial parameter estimates that we would expect (based on findings in Gresnigt et al. (2015)) to be close to the maximum likelihood estimates (i.e., the parameter estimates that maximize the log-likelihood), we found that all those approaches give the same maximum likelihood estimates (that indeed have a gradient that is (very close to) zero). Furthermore, the computation times were roughly equal (the slowest and fastest time were not more than a factor 2 apart). Changing the initial parameter estimates, we indeed found that the Nelder-Mead simplex direct search algorithm (with parameter bounds) remained giving the same maximum likelihood estimates. However, changing the initial values caused the other two methods to give very different (and counterintuitive) results, whereas computation times of the three methods remained roughly equal. We will therefore use the Nelder-Mead simplex direct search algorithm with parameter bounds to estimate the model parameters.

4.3 Predicting crashes

The parameter estimates obtained by maximum likelihood can then be used to estimate the conditional intensity of arrivals that follow a Hawkes process. We can derive the probability that a crash will happen in a certain time interval if the occurrence of events follows a Hawkes process. In particular, the probability of at least one arrival in the time interval $[t_{n-1}, t_n]$ (when the events follow a Hawkes process with conditonal intensity $\lambda(t|\boldsymbol{\theta}; \mathcal{H}_t)$) can be calculated as

$$P(N(t_n) - N(t_{n-1}) > 0) = F(t \leq t_n - t_{n-1}) = 1 - \exp\bigg(- \int_{t_{n-1}}^{t_n} \lambda(t|\boldsymbol{\theta}; \mathcal{H}_t) dt \bigg), \tag{4}$$

where we used that the interarrival times of a Poisson process are exponentially distributed. A full derivation of what this probability looks like when we use the conditional intensity of Equation (2) can be found in Appendix C. We can then plug the estimated conditional intensity $\lambda(t|\hat{\boldsymbol{\theta}}; \mathcal{H}_t)$ into Equation (4) to estimate the probability that a crash will happen in a certain time interval. With these estimated probabilities, we can construct an Early Warning System as in Gresnigt et al. (2015), by predicting a crash in the next five days if the estimated probability of a crash in the next five days exceeds a certain threshold, to be determined later.

Once we have obtained the predictions of when a crash will happen, we will evaluate the the predictive power of the EWS. The idea of this evaluation is to compare $\hat{p}_t$ or $\hat{y}_t$ with y_t. Here $\hat{p}_t$ denotes the estimated probability of a crash in the next five days starting at day t. Further, $\hat{y}_t$ denotes a dummy variable that takes the value 1 if the estimated probability of a crash in the next five days (starting at day t) exceeds a certain threshold and takes the value 0 otherwise. Lastly, y_t denotes a dummy variable that takes the value 1 if a crash happens in the next five days (starting at day t) and takes the value 0 otherwise. There exist several methods to compute the predictive power. For example, one can compute the Quadratic Probability Score (QPS), the Log Probability Score (LPS), the adjusted Mean Squared Prediction Error (adjusted MSPE), or the Hanssen-Kuiper Skill Score (KSS). Those measures of predictive power are discussed in Gresnigt et al. (2015). The KSS uses the estimated dummy $\hat{y}_t$, whereas the other measures do not. As we are interested in whether or not a crash will happen and as the KSS is discussed most extensively in Gresnigt et al. (2015), we will use the KSS to evaluate the predictive power of the EWS. The KSS is defined as the hit rate minus the false alarm rate. The hit rate is a fraction that measures how often a crash is predicted when it should be predicted, whereas the false alarm rate is a fraction that measures how often a crash is predicted when it should not be predicted. We will calculate those Hanssen-Kuiper Skill Scores for the S&P 500 index over the period from 2 September 2008 until 1 January 2013.

4.4 Optimizing and using the Early Warning System

To optimize the EWS of Gresnigt et al. (2015), we will try to find an optimal threshold value (say τ_{opt}) for determining whether we should predict a crash. One possibility is to find the threshold value that maximizes the Hanssen-Kuiper Skill Score, so that the fraction of correct crash predictions is large and the fraction of false crash predictions is low. (Other possibilities for finding an optimal threshold value related to portfolio construction will be discussed at the end of this section.) Since setting the threshold value very close to 1 will result in very few warning signals (and hence few correct crash predictions) and setting the threshold value close to 0 will result in a lot of warning signals (and hence much false crash predictions), there should be an optimal value in between those numbers. A value of 0.5 is chosen in Gresnigt et al. (2015), but we will explore whether some other threshold might result in better skill scores. We will first do this on the same data as in Gresnigt et al. (2015), by letting the threshold value τ run between 0 and 1 in small steps and computing the skill scores for the different thresholds.

4.4.1 Simulating asset returns

To draw more general conclusions about an optimal threshold value in our EWS, we will simulate financial data rather than using some particular observed dataset. We will consider several different models to simulate asset returns, and we will pick the most suitable one(s). As a first example, we might consider the possibility that the asset prices are generated by Geometric Brownian Motion, as is relatively often assumed in the financial world (for example in the Black-Scholes model for option prices). In particular, this implies that asset returns are normally distributed, with constant volatility.

However, in Longin (2005) an empirical analysis is used to make a choice about the distribution of asset returns. In that paper, it is concluded that the assumption of normality does not seem realistic. Further, it is concluded that a Student-t distribution seems more appropriate for the unconditional modeling of returns and that an ARCH process could be used in a conditional modeling of returns. Additionally, some stylized facts about asset returns are reported in Cont (2001). Some of the stylized facts include the heavy tails/non-normality of the (unconditional) distribution and the presence of volatility clustering. This suggests that GARCH-type models might be appropriate to model the asset returns, and that Geometric Brownian Motion does not seem appropriate because it has constant volatility. Another stylized fact reported in Cont (2001) is the conditional heavy tails of asset returns; the residual time series still exhibits heavy tails after returns are corrected for volatility clustering, which can be done by using GARCH-type models. We will therefore use a Student-t distribution for the error terms in GARCH-type models, in order to capture the conditional heavy tails.

We will now explain which GARCH-type models we use to simulate asset returns. As it is indicated in Hansen and Lunde (2005) that it is difficult to outperform the GARCH(1,1) model in forecasting volatility (among others compared with other combinations of the orders p and q in the GARCH(p, q) model), we start with a GARCH(1,1) model, with a Student-t distribution for the error terms. We thus model the returns as

$$
\begin{aligned}
r_t &= \mu + z_t \sigma_t, \\
\sigma_t^2 &= \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2,
\end{aligned}
$$

where $z_t \sim t(\nu)$. Another stylized fact reported in Cont (2001) is that most measures of volatility are negatively correlated with the returns of that asset. This stylized fact is also known as the 'leverage effect', and we will consider a GJR(p, q) model that incorporates this effect, again with a Student-t distribution for the error terms. For simplicity purposes, we take the orders p and q again equal to 1 as in our GARCH model. We thus model the returns as

$$
\begin{aligned}
r_t &= \mu + z_t \sigma_t, \\
\sigma_t^2 &= \omega + \alpha \epsilon_{t-1}^2 + \gamma \mathbb{I}\{\epsilon_{t-1} < 0\} \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2,
\end{aligned}
$$

where $z_t \sim t(\nu)$ and where $\mathbb{I}\{A\}$ denotes an indicator function that takes the value 1 if A occurs and 0 otherwise.

To simulate asset returns with these two models, we have to plug in values for the unknown parameters $\{\mu, \omega, \alpha, \gamma, \beta, \nu\}$. In order to use parameter estimates that are capable of generating a series of asset returns that could have been observed in the market, we fit the two models to our series of S&P 500 returns from 1 January 1957 until 1 September 2008. The obtained parameter estimates will be reported and interpreted at the start of Subsection 5.2.

We then use the estimated GARCH(1,1) and GJR(1,1) models (both with a Student-t distribution for the error terms) to simulate data. We simulate returns for the same number of trading days as in our S&P 500 dataset. The first part of the simulated data (again 13005 trading days just as for the S&P 500 dataset) in each separate replication will be converted to a series of crashes in the same way as in Gresnigt et al. (2015). We will only model arrivals of extreme negative returns (and not separately extreme returns consisting of both positive and negative returns) when we optimize and use the EWS, as we want to predict when a crash is going to happen, and not necessarily when an extreme return is going to happen.

4.4.2 Optimizing thresholds and using them for investing

We then use these simulated series of crashes to estimate the parameters in the ETAS model. With these estimated parameters we can estimate the probability of a crash in the next time interval with

Equation (4), as already described before. If we fix the threshold value τ between 0 and 1, we can calculate the average KSS over the remaining (1087) observations of all different replications [1]. We will then vary the threshold value τ between 0 and 1 and calculate for each τ the average skill score over all replications. The threshold value τ_{opt} that yields the highest average KSS will be preferred in the EWS using thresholds based on the average KSS.

Once we have obtained the threshold value that yields the highest average KSS, we also want to explore whether an investor can benefit from selling the asset if the estimated probability of a crash in the next days exceeds this threshold value, and buying the asset back when the estimated crash probability becomes again lower than the threshold value. If the asset is sold, the proceeds are temporarily put into a bank account yielding a negligible yearly interest rate of 0.1%. In order to use the optimal threshold based on the average KSS for investing, we make use of the same simulated dataset. As we should not use the same subsample to determine the optimal threshold and to evaluate portfolios (because this results in a forward looking bias), we make use of different subsamples of the simulated dataset. We use the first 10000 observations of all simulated returns to estimate the parameters in the ETAS model, so our estimation sample consists of observations 1 until 10000. We then use observations 10001 until 11000 of all simulated returns to determine the optimal threshold, so our determination sample consists of observations 10001 until 11000. We finally use observations 11001 until 12000 of all simulated returns to evaluate the performance of investment strategies that use the previously obtained threshold, so our evaluation sample consists of observations 11001 until 12000 [2]. In addition to an EWS using thresholds based on the average skill score, we will also explore other methods (than maximizing the average KSS) to find optimal threshold values for withdrawing money. In particular, we can also take for each simulated return series the threshold value that maximizes the KSS of that particular series (so that we obtain different thresholds for each simulated return series). Another possibility is to take for each return series a threshold that maximizes the value of the asset over previous observations.

We will compare all different investment strategies ((i) always investing, and investing based on an EWS (ii) using threshold 0.5 or thresholds based on (iii) the mean KSS, (iv) all KSSs, or (v) past portfolio returns) by computing two well-known performance measures: the Value at Risk and the Sharpe ratio. The Value at Risk (VaR(95)) is defined as the return that is exceeded in 95% of the cases. It therefore gives an indication of the riskiness of a particular investment; a higher Value at Risk indicates that an investment is less risky, which is preferable. The Sharpe ratio measures the sample average of the excess return of an asset $(\bar{r} - r_f)$, with $\bar{r} = \sum_{t=1}^{T} r_t$ the sample average of the returns and r_f the risk-free interest rate, divided by the sample standard deviation, so

$$\text{Sharpe Ratio} = \frac{(\bar{r} - r_f)}{\sqrt{\frac{\sum_{t=1}^{T}(r_t - \bar{r})^2}{T-1}}}.$$

[1] We do not compute skill scores for the last 14096 - 13005 - 1087 = 4 trading days, because our prediction interval has a length of five days, so that it does not make sense to predict a crash in the next five days if less than five days are remaining.

[2] As mentioned in the previous paragraph, we use the estimated probability of a crash in the next five days, so we actually compute the KSS and portfolio values for 996 rather than for 1000 trading days.

A higher Sharpe ratio therefore indicates that an investment has a high (excess) return relative to how risky that investment is, which is preferable.

After comparing investment strategies for single assets, we will explore strategies involving more assets. For this we want to look at a portfolio that consists of N assets, where we take $N = 10$. The benchmark model is the $1/N$ portfolio that takes a position of $1/N$ of the total wealth in all assets and is rebalanced at a certain frequency (monthly in this research). The strategy that we propose to use also takes a position of $1/N$ of the total wealth in all assets and also rebalances the portfolio monthly. However, in addition to this, the portfolio will be adjusted daily. If the estimated probability that a certain asset will crash exceeds the threshold value (that we will determine in the same way as for the strategies for single assets), then the total wealth will be equally distributed over the other assets. If the estimated probability that a particular asset will crash decreases and becomes again smaller than the threshold value, then this asset will be incorporated in the portfolio again and an equal part of the total wealth will be distributed over the assets in the portfolio. Basically, we distribute every day an equal part of our wealth over the assets that are not predicted to crash. It is possible however that all assets are predicted to crash. In that case we will temporarily invest in none of the assets (and put our money into a savings account with a negligible interest rate of 0.1% annually), until some assets will become less likely to crash. These investment strategies for portfolios will also be evaluated by looking at their Values at Risk and Sharpe ratios. To check whether possible differences in performance between the $1/N$ portfolio and the other strategies might be caused by the fact that the other strategies rebalance daily (instead of monthly), we will also compute the Value at Risk and Sharpe ratio of a $1/N$ portfolio that rebalances daily.

5 Results

This section describes the results that we obtained from estimating and evaluating an EWS based on an ETAS model with a conditional intensity for the arrivals of events as given in Equation (2). We will first report the results obtained by applying our methods to S&P 500 data (in Subsection 5.1) and then we will describe the results obtained by applying our methods to simulated data (in Subsection 5.2).

5.1 Results obtained from S&P 500 data

When we use the ETAS model with a conditional intensity for the arrivals of events as given in Equation (2) on the S&P 500 data between 2 January 1957 and 1 September 2008, we obtain parameter estimates that can be found in Table 1. Comparing these estimation results with the results in Gresnigt et al. (2015), we see that our estimates are roughly equal, including the standard errors of the constant intensity parameter μ. However, the standard errors for the parameters a and α are smaller in our model (that does not model the magnitudes of past arrivals).

Table 1: Estimation results of model with exponential triggering function

Arrival choice	Neg 95	Neg 97	Neg 99	Extr 95	Extr 97	Extr 99
$\hat{\mu}$	0.0120	0.0065	0.0033	0.0083	0.0056	0.0031
	(0.0014)	(0.0010)	(0.0006)	(0.0011)	(0.0009)	(0.0006)
$\hat{a}$	0.0303	0.0276	0.0218	0.0366	0.0387	0.0340
	(0.0015)	(0.0017)	(0.0025)	(0.0017)	(0.0023)	(0.0038)
$\hat{\alpha}$	0.0397	0.0349	0.0324	0.0436	0.0473	0.0489
	(0.0019)	(0.0021)	(0.0037)	(0.0019)	(0.0017)	(0.0051)
log $L(\boldsymbol{\theta})$	-2355.69	-1550.50	-647.79	-2226.65	-1473.35	-619.92
AIC	4717.37	3107.01	1301.58	4459.30	2952.71	1245.85

The estimation results are obtained by applying an ETAS model with the conditional intensity in Equation (2) to different quantiles of extreme (negative) returns of the S&P 500 index over a period from 2 January 1957 until 1 September 2008. The entries in the row 'Arrival choice' indicate to which quantile of extreme returns or extreme negative returns the model is applied. Standard deviations of the estimated parameters are shown in parentheses.

We also determined the Hanssen-Kuiper Skill Scores of the predictions obtained by using the ETAS model (with a conditional intensity for the arrivals of events as given in Equation (2)) on S&P 500 data from 2 September 2008 until 1 January 2013. The optimal thresholds are reported in Table 2, together with the corresponding KSS and the KSS if we simply take the threshold equal to 0.5. It can be noted from these results that the KSS is higher if we use extreme returns (so both positive and negative returns) to model arrivals than when we only use negative returns to model arrivals. However, as already mentioned in Section 4, we will proceed to only use negative returns. This is because we want to predict when a crash is going to happen, and not necessarily when an extreme return is going to happen. Figures of the value of the KSS for thresholds between 0 and 1 can be found in Figure 3 of Appendix D.

Table 2: Prediction results of model with exponential triggering function

Arrival choice	Neg 95	Neg 97	Neg 99	Extr 95	Extr 97	Extr 99
maximum KSS	0.433	0.471	0.536	0.586	0.626	0.6808
optimal threshold	0.59	0.48	0.31	0.57	0.40	0.27
KSS for threshold 0.5	0.386	0.467	0.499	0.561	0.619	0.550

The Hanssen-Kuiper Skill Scores are obtained by predicting a crash in the next five days from 2 September 2008 until 1 January 2013, based on an ETAS model with the conditional intensity in Equation (2). Events are predicted when the estimated probability of an event in the next five days exceeds the threshold.

We can also apply our investment strategy (of investing in the S&P 500 index if and only if its estimated probability of a crash in the next five days is smaller than the threshold) on the S&P 500 data. For this purpose, we can only use the EWS with the threshold set equal to 0.5 as in Gresnigt et al. (2015). This is because we determined the optimal thresholds over a period from 2 September

2008 until 1 January 2013, so evaluating investment strategies (using these thresholds) over the same time period would result in a forward looking bias. The wealth that one would obtain by remaining invested in the S&P 500 index and the wealth that could be obtained by applying our investment strategy are both plotted in Figure 1. From this, it can be seen that it would be very beneficial to apply the investment strategy, and that this strategy would have worked best if one defines crashes as the 95% quantile of extreme negative returns.

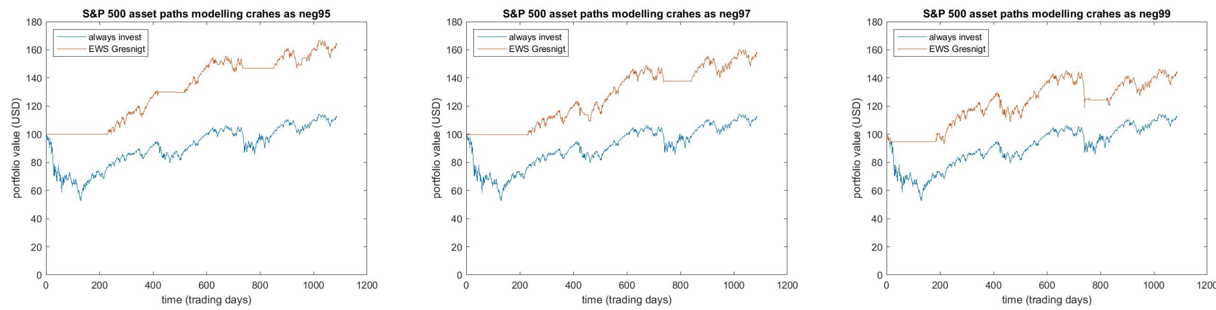

Figure 1: Wealth of the S&P 500 index from 2 September 2008 until 1 January 2013. The blue line just keeps all money in the S&P 500, whereas the orange line withdraws money if the EWS based on an ETAS model with a conditional intensity of the arrivals as specified in Equation (2) predicts that the probability of a crash in the next five days is bigger than 0.5.

5.2 Results obtained from simulated data

We will now describe the results that we obtained by applying our methods to simulated returns. As mentioned in Subsection 4.4.1, we fit a GARCH(1,1) and a GJR(1,1) model to our series of S&P 500 returns from 1 January 1957 until 1 September 2008 to obtain parameter estimates that we use to simulate returns. We get the following parameter estimates for the GARCH(1,1) model: $\hat{\mu} \approx 0.051$ (0.006), $\hat{\omega} \approx 0.005$ (0.001), $\hat{\alpha} \approx 0.069$ (0.004), $\hat{\beta} \approx 0.927$ (0.004), and $\hat{\nu} \approx 7.5$ (0.4). These parameter estimates confirm the stylized facts in Cont (2001); the highly significant positive value of $\hat{\beta}$ indicates that the volatility depends to a large extent on its past (unobserved) value. Moreover, $\hat{\alpha} + \hat{\beta} \approx 0.996$, which indicates that the conditional variance reverts very slowly to its unconditional mean $\bar{\sigma}^2$, resulting in a clustering of the volatility. This can be seen from rewriting the conditional variance as

$$\sigma_t^2 = \bar{\sigma}^2 + \alpha(\epsilon_{t-1}^2 - \sigma_{t-1}^2) + (\alpha + \beta)(\sigma_{t-1}^2 - \bar{\sigma}^2), \tag{5}$$

from which we see that $\hat{\alpha} + \hat{\beta} \approx 0.996$ leads to slow mean reversion. A derivation of Equation (5) can be found in Appendix E. For the GJR(1,1) model, we obtain the following parameter estimates: $\hat{\mu} \approx 0.040$ (0.006), $\hat{\omega} \approx 0.006$ (0.001), $\hat{\alpha} \approx 0.025$ (0.004), $\hat{\gamma} \approx 0.086$ (0.007), $\hat{\beta} \approx 0.926$ (0.004), and $\hat{\nu} \approx 8.1$ (0.4). These parameter estimates also confirm the stylized facts in Cont (2001); the highly significant positive value of $\hat{\beta}$ indicates that the volatility depends to a large extent on its past (unobserved) value. Also, $\hat{\alpha} + \hat{\beta} + \frac{\hat{\gamma}}{2} \approx 0.994$, which indicates that the conditional variance reverts very slowly to its unconditional mean $\bar{\sigma}^2$, resulting in a clustering of the volatility. This can be seen from rewriting the conditional variance as

$$\sigma_t^2 = \bar{\sigma}^2 + \alpha(\epsilon_{t-1}^2 - \sigma_{t-1}^2) + \gamma\left(\mathbb{I}\{\epsilon_{t-1} < 0\}\epsilon_{t-1}^2 - \frac{1}{2}\sigma_{t-1}^2\right) + \left(\alpha + \beta + \frac{\gamma}{2}\right)(\sigma_{t-1}^2 - \bar{\sigma}^2), \qquad (6)$$

from which we see that $\hat{\alpha} + \hat{\beta} + \frac{\hat{\gamma}}{2} \approx 0.994$ leads to slow mean reversion. A derivation of Equation (6) can be found in Appendix F. Moreover, we see that $\hat{\gamma}$, the partial effect of negative abnormal returns on the volatility, is significantly positive, confirming the stylized fact that volatility is negatively correlated with the asset return.

For each separate return series that we simulated using the parameter estimates above, we used the first 13005 observations of each return series to estimate the ETAS model with a conditional intensity for the arrivals of events as given in Equation (2). We computed (for both the simulation models) the parameter estimates in all 1000 simulations. The average and standard deviation of those different estimates are reported in Table 3. From these results, it can be seen that the parameter estimates obtained from the S&P 500 data agree with these simulated results (as their 95% confidence intervals are not disjoint).

Table 3: Estimation results of model with exponential triggering function obtained by simulation

Simulation model	Arrival choice	Neg 95	Neg 97	Neg 99
	$\hat{\mu}$	0.0098	0.0058	0.0021
		(0.0025)	(0.0017)	(0.0009)
GARCH-t	$\hat{a}$	0.0286	0.0289	0.0274
		(0.0031)	(0.0037)	(0.0058)
	$\hat{\alpha}$	0.0356	0.0359	0.0348
		(0.0042)	(0.0050)	(0.0069)
	$\hat{\mu}$	0.0105	0.0063	0.0022
		(0.0023)	(0.0017)	(0.0008)
GJR-t	$\hat{a}$	0.0341	0.0346	0.0335
		(0.0034)	(0.0040)	(0.0063)
	$\hat{\alpha}$	0.0432	0.0438	0.0431
		(0.0045)	(0.0054)	(0.0078)

The estimation results are obtained by applying an ETAS model with the conditional intensity in Equation (2) to different quantiles of extreme negative returns of simulated returns over 13005 trading days. The first column indicates which model is used to simulate returns and the first row indicates to which quantile of negative returns the ETAS model is applied. The reported parameter estimates are the averages over 1000 simulations. In between parentheses are the standard deviations of the estimates obtained in the 1000 simulations.

We also determined for each separate return series that we simulated the Hanssen-Kuiper Skill Score of the predictions obtained by using the ETAS model with a conditional intensity for the arrivals of events as given in Equation (2) on the last 1087 simulated trading days. Plots of the

KSS against the threshold for all 1000 simulations of the two simulation models can be found in Figure 4 of Appendix G. The KSS can be quite different for different return series, but to determine one optimal threshold, we look at the threshold value that maximizes the average KSS. These optimal thresholds (for different ways to define crashes) are reported in Table 4, together with the corresponding maximum average KSS and the average KSS if we simply take the threshold equal to 0.5. From these results, it can be seen that the optimal threshold is smaller than 0.5, and the optimal threshold becomes smaller as we use less observations to model the crashes (i.e., if we use only the 99% quantile of negative returns instead of the 97% or 95% quantile of negative returns). Using these thresholds will therefore result in more warning signals than using the threshold 0.5. The optimal thresholds based on the average KSS over all simulations are lower than the optimal thresholds based on the S&P 500 data. This can be explained by the fact that the hold-out sample for the S&P 500 data contained the financial crisis, so that the estimated probabilities of having crashes in the next time interval are also larger than in the simulated return series; setting the threshold very low with these high estimated crash probabilities would have resulted in more false crash predictions. It can further be observed that the average KSS is somewhat higher if we use the GJR(1,1) model instead of the GARCH(1,1) model to simulate return series. This small difference is significant, as for each arrival choice the 95% confidence intervals of the maximum average KSS are disjoint; however, if we use another subsample of the simulated data, the GJR(1,1) model still yields a higher average KSS, but this difference is not significant anymore (see Table 8 in Appendix I). We therefore see that our results are quite robust for the choice of the asset returns model, which is good if we want to apply our EWS in practice.

Table 4: Prediction results of model with exponential triggering function obtained by simulation

Simulation model	Arrival choice	Neg 95	Neg 97	Neg 99
	maximum average KSS	0.305	0.329	0.308
		(0.005)	(0.006)	(0.006)
GARCH-t	optimal threshold	0.24	0.12	0.05
	average KSS for threshold 0.5	0.199	0.144	0.079
		(0.006)	(0.006)	(0.006)
	maximum average KSS	0.337	0.366	0.354
		(0.005)	(0.006)	(0.009)
GJR-t	optimal threshold	0.23	0.16	0.05
	average KSS for threshold 0.5	0.241	0.187	0.106
		(0.006)	(0.006)	(0.007)

The Hanssen-Kuiper Skill Scores are obtained by predicting a crash in the next five days over the last 1087 simulated trading days, based on an ETAS model with the conditional intensity in Equation (2). Crashes are predicted when the estimated probability of a crash in the next five days exceeds the threshold. The first column indicates which model is used to simulate returns and the first row indicates to which quantile of negative returns the ETAS model is applied. The reported average KSSs are averages over 1000 simulations. In between parentheses are the standard deviations of average KSSs over 1000 simulations.

Figures of the value of the average KSS for different thresholds (when crashes are defined as the 95% quantile of negative returns) can be found in Figure 2. Based on the narrow confidence bounds, we see that indeed certain thresholds are preferred. We see that the plots for both ways of simulating asset returns are very similar; the average KSS remains zero for a short time, to sharply increase for increasing thresholds and thereafter gradually decline when the threshold keeps growing. Figures of the value of the average KSS for different thresholds when crashes are defined as other quantiles of negative returns have a similar shape, and can be found in Figure 5 of Appendix J. The confidence bounds are constructed by adding (respectively subtracting) two times the estimated standard deviation of the average KSS. These standard deviations of the average KSS are computed by dividing the estimated standard deviation of the KSS by the square root of the number of times that we computed the KSS. This number is sometimes (slightly) smaller than the number of simulations, as we discarded the observations in which no crash actually happened in the hold-out sample, as this would make the KSS undefined.

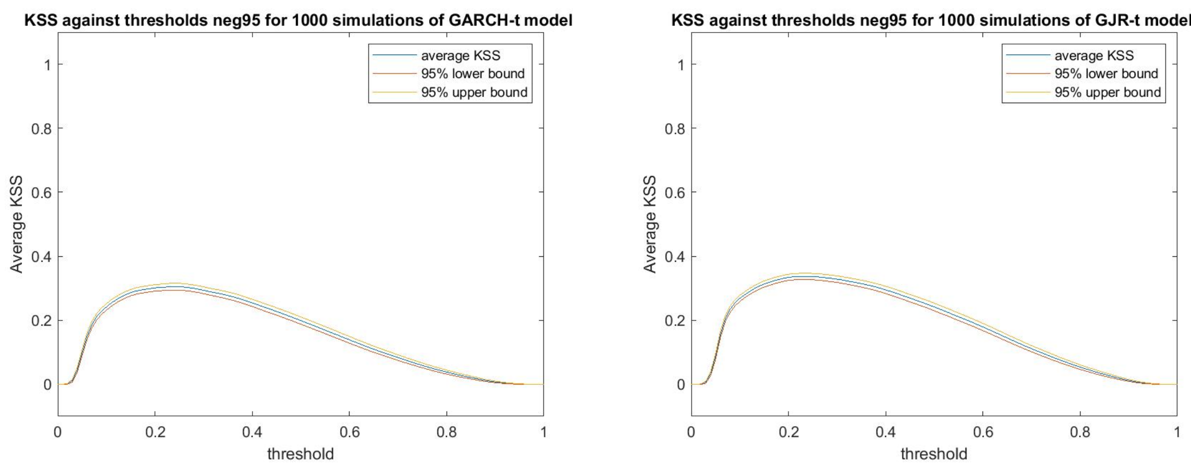

Figure 2: Average Hanssen-Kuiper Skill Scores (and 95% confidence bounds) based on 1000 simulations for different thresholds of the EWS based on an ETAS model with a conditional intensity of the arrivals as specified in Equation (2) when predicting a crash in the next five days, over 1087 trading days. A crash is predicted when the estimated probability of a crash in the next five days exceeds the threshold.

We will now evaluate the different investment strategies using the simulated asset returns. We used the estimation sample (for each simulated return series) to obtain parameter estimates, which can be found in Table 7 of Appendix H and are clearly not significantly different from the parameter estimates over the first 13005 observations. We then used these parameter estimates (and the determination sample of each simulated return series) to compute the optimal thresholds based on (i) the average KSS over all simulations, (ii) the KSS in each individual simulation (so 1000 thresholds in total), and (iii) the highest portfolio value (again 1000 thresholds in total). The optimal thresholds based on the average KSS (and values of the maximum average KSS) can be found in Table 8 of Appendix I. Subsequently, we used the evaluation sample of all simulated returns to evaluate the investment strategies. Figures of all individual asset paths can be found in Appendix K, and figures with the average wealth of the individual assets (Figure 11) and portfolios (Figure 12) can be found in Appendix L.

The average Sharpe ratios and average Values at Risk of the different strategies applied to each of the 1000 simulated return series can be found in Table 5. We first see that all investment strategies that make use of an EWS have significantly higher average Sharpe ratios (up to twice as high) and average Values at Risk than an investment strategy that does not adjust investments. Based on the average Sharpe ratios, we see that it is preferred to define crashes as the 95% quantile of negative returns, as the average Sharpe ratio is higher (and this difference is significant for three out of four investment strategies that model crashes for determining the investment strategy) when we define crashes that way. We further see that the Sharpe ratio is significantly higher using an investment strategy that adjusts investments based on an EWS that gives warning signals if the estimated crash probability exceeds a threshold based on (i) the average KSS over the simulations or (ii) the KSS of each individual asset. From these two investment strategies, (i) has a higher average Sharpe ratio than (ii), but this difference is (just) not significant as the 95% confidence intervals are (just) not disjoint. Based on the average Values at Risk, we also see that it is preferred to define crashes as the 95% quantile of negative returns. Also, the same two investment strategies are preferred, where an EWS based on the average KSS is now significantly better than an EWS based on the KSS of each individual asset. In particular, we thus see that both performance measures indicate that an EWS that uses optimized thresholds based on the (average) KSS over all simulations performs better than taking the threshold equal to 0.5, due to the fact that those strategies take less risk. We further note that all results above hold for both choices for simulating asset returns, and are thus robust in that sense.

We also determined the average Sharpe ratios and average Values at Risk of the different strategies applied to 100 simulations of portfolios consisting of (a weighted combination of) 10 asset return series each. These can be found in Table 6. We see again that all investment strategies that make use of an EWS have significantly higher average Sharpe ratios than the $1/N$ portfolio (up to twice as high again). This difference is not due to the fact that the investment strategies that make use of an EWS rebalance daily, as they also have significantly higher average Sharpe ratios than the $1/N$ portfolio that rebalances daily. Also the Values at Risk of all the investment strategies that make use of an EWS are higher than the Value at Risk of the $1/N$ portfolio (for both rebalancing frequencies), but this difference is not significant for the EWS with a threshold that maximized the portfolio value in the determination sample. Based on the average Sharpe ratio, we find the same results as for the individual assets. That is, we first find that it is preferred to define crashes as the 95% quantile of negative returns (with again for all but one strategy significantly higher average Sharpe ratios). Second, we find that investment strategies implementing an EWS with a treshold based on the (average) KSS over all simulations are preferred. In particular, those two investment strategies perform better than an investment strategy that uses a threshold of 0.5 for predicting crashes. The Value at Risk is less informative for comparing the portfolio strategies than for comparing investment strategies of individual assets, as the Values at Risk indicate the same results, but the differences are not significant. This is in part due to the fact that we only used 100 portfolios (whereas we used 1000 individual assets), causing the standard deviations of the mean Value at Risk to be a factor $\sqrt{10}$ bigger. We note again that all results above hold for both choices for simulating asset returns, and are thus robust in that sense.

Table 5: Performance measures of different investment strategies obtained by simulation

Average Measure	Strategy	GARCH			GJR		
		Neg 95	Neg 97	Neg 99	Neg 95	Neg 97	Neg 99
Sharpe ratio	always invest	0.056	0.056	0.056	0.051	0.051	0.051
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	EWS mean KSS	0.094	0.087	0.071	0.092	0.080	0.063
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	EWS individual KSS	0.091	0.085	0.071	0.088	0.080	0.064
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	EWS Gresnigt	0.083	0.074	0.063	0.080	0.070	0.058
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	EWS portfolio history	0.075	0.073	0.066	0.069	0.066	0.059
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
VaR(95)	always invest	-1.510	-1.510	-1.510	-1.377	-1.377	-1.377
		(0.021)	(0.021)	(0.021)	(0.020)	(0.020)	(0.020)
	EWS mean KSS	-0.738	-0.826	-1.018	-0.683	-0.776	-0.972
		(0.007)	(0.007)	(0.007)	(0.005)	(0.005)	(0.006)
	EWS individual KSS	-0.789	-0.873	-1.016	-0.726	-0.812	-0.953
		(0.006)	(0.006)	(0.010)	(0.004)	(0.005)	(0.009)
	EWS Gresnigt	-1.039	-1.183	-1.351	-0.938	-1.070	-1.223
		(0.006)	(0.007)	(0.011)	(0.005)	(0.005)	(0.008)
	EWS portfolio history	-1.201	-1.213	-1.267	-1.047	-1.0813	-1.127
		(0.021)	(0.021)	(0.020)	(0.018)	(0.018)	(0.018)

The average performance measures are obtained by evaluating investment strategies, based on an ETAS model with the conditional intensity in Equation (2), for 996 simulated trading days. The entries in the column 'Strategy' (with EWS in their name) indicate how the thresholds for giving a warning are determined. The reported average performance measures are averages over 1000 simulations. In between parentheses are the standard deviations of the average performance measures over 1000 simulations.

Table 6: Performance measures of different portfolio strategies obtained by simulation

Average Measure	Strategy	GARCH			GJR		
		Neg 95	Neg 97	Neg 99	Neg 95	Neg 97	Neg 99
Sharpe ratio	$1/N$	0.153	0.153	0.153	0.134	0.134	0.134
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
	$1/N$ daily rebalance	0.154	0.154	0.154	0.133	0.133	0.133
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
	EWS mean KSS	0.284	0.265	0.216	0.282	0.244	0.194
		(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)
	EWS individual KSS	0.274	0.257	0.214	0.267	0.243	0.193
		(0.003)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
	EWS Gresnigt	0.256	0.228	0.189	0.247	0.213	0.172
		(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
	EWS portfolio history	0.201	0.197	0.184	0.180	0.1754	0.158
		(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.004)
VaR(95)	$1/N$	-0.493	-0.493	-0.493	-0.447	-0.447	-0.447
		(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
	$1/N$ daily rebalance	-0.492	-0.492	-0.492	-0.454	-0.454	-0.454
		(0.010)	(0.010)	(0.010)	(0.012)	(0.012)	(0.012)
	EWS mean KSS	-0.351	-0.356	-0.384	-0.316	-0.328	-0.349
		(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
	EWS individual KSS	-0.363	-0.365	-0.396	-0.323	-0.330	-0.361
		(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
	EWS Gresnigt	-0.363	-0.387	-0.427	-0.327	-0.348	-0.385
		(0.003)	(0.004)	(0.005)	(0.003)	(0.003)	(0.004)
	EWS portfolio history	-0.473	-0.472	-0.469	-0.429	-0.426	-0.433
		(0.011)	(0.011)	(0.011)	(0.012)	(0.012)	(0.011)

The average performance measures are obtained by evaluating investment strategies, based on an ETAS model with the conditional intensity in Equation (2), for 996 simulated trading days. The entries in the column 'Strategy' (with EWS in their name) indicate how the thresholds for giving a warning are determined. The reported average performance measures are averages over 100 simulations, in each of which portfolios consisting of ten assets are evaluated. In between parentheses are the standard deviations of the average performance measures over 100 simulations.

6 Conclusion

In this paper, we have investigated whether we can optimize an Early Warning System based on earthquakes to improve investments. To do so, we have first implemented the framework described in Gresnigt et al. (2015) to estimate the probability of a crash in the next five days, for which we first had to estimate the parameters in an ETAS model. A crash is predicted by our EWS if the estimated crash probability exceeds a certain threshold value; this threshold is set equal to 0.5 in

Gresnigt et al. (2015), but we have explored several methods to choose this threshold value in a way that leads to better investments. In order to draw general conclusions about the threshold choice, we have simulated 1000 asset return series with both a GARCH(1,1) and a GJR(1,1) model (both with a Student-t distribution for the error terms). Using these simulated return series, we determined thresholds that (i) maximize the average Hanssen-Kuiper Skill Score (KSS) over the evaluation period for all simulations, (ii) maximize the KSS of each individual return series over the evaluation period, and (iii) maximize the value of an investment in each return series over the evaluation period. Based on these different thresholds, we have investigated investment strategies for individual assets and for portfolios of assets. The investment strategies for individual assets withdraw money (and put it in a bank account) when the estimated crash probability of that asset exceeds the threshold value, and investment strategies for portfolios only invest in assets whose estimated crash probability is smaller than the threshold.

We evaluated the investment strategies based on different ways to determine the threshold using two well-known performance measures: the Sharpe ratio and the Value at Risk. For both ways to simulate return series, we found the same results; the performance measures both indicate that investment strategies that implement an EWS significantly outperform a strategy that remains invested in the asset. When implementing different EWSs for individual assets, we found that all EWSs work best if we define crashes as the 95% quantile of negative returns. We further found that the EWSs that use thresholds that maximize the average KSS over all simulations are superior, and that EWSs that use thresholds that maximize the KSS of the separate return series in each simulation are a good number two. When looking at portfolios of assets, we find that the portfolios that make use of an EWS perform significantly better than the $1/N$ portfolio that invests an equal amount in all assets. We also find for portfolios that the EWSs work better if we define crashes as the 95% quantile of negative returns, and that EWSs that use thresholds that maximize (average) skill scores over the evaluation period lead to higher average Sharpe ratios. Moreover, we find for both individual assets and portfolios that the average Sharpe ratio and the average Value at Risk are higher if we use optimimized thresholds than if we simply take the threshold equal to 0.5.

We thus find the following answer to our research question: it is possible to optimize an EWS based on earthquakes to improve investments. In particular, we have found two main things. First, it is preferred to define crashes as the 95% quantile of negative returns. Second, the optimal threshold for giving a warning signal is preferably chosen as the threshold that maximizes the average KSS of all assets that an investor wants to invest in.

We also have some suggestions for further research. In this research, the conditional intensity that we used for the arrivals of crashes depends (only) on the previous arrival times. However, it might be interesting to see whether the sizes of previous crashes are also explanatory for the intensity at which crashes arrive. In that case we would allow crashes to follow a marked Hawkes process, where the marks are the return magnitudes. If we want to evaluate how well an EWS based on a marked Hawkes process predicts crash dates, we can still use the KSS. However, if we also want to evaluate how well such an EWS predicts the return magnitudes, we should use evaluation measures that also take return sizes into account. Further research might also focus on the application of our investment strategies in practice; for example, we rebalanced our portfolio at a daily frequency, and made the idealized assumption of no transaction costs for rebalancing. Hence, it would be interesting to see how beneficial our strategy remains if we incorporate transaction costs, and whether we should rebalance at lower frequencies in the presence of transaction costs. Finally, we focussed in this research on predicting crashes (so negative returns) and adjusting investments accordingly; it might also be of interest to investigate and test whether we can use an ETAS model to estimate the probability that a large positive return will happen, and (if that is indeed possible) adjust investments based on that information as well.

References

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues.

Gerlach, J.-C., Demos, G., and Sornette, D. (2018). Dissection of bitcoin's multiscale bubble history from january 2012 to february 2018. *arXiv preprint arXiv:1804.06261*.

Gresnigt, F., Kole, E., and Franses, P. H. (2015). Interpreting financial market crashes as earthquakes: A new early warning system for medium term crashes. *Journal of Banking & Finance*, 56:123–139.

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(7):873–889.

Heij, C., de Boer, P., Franses, P. H., Kloek, T., van Dijk, H. K., et al. (2004). *Econometric methods with applications in business and economics*. Oxford University Press.

Longin, F. (2005). The choice of the distribution of asset returns: How extreme value theory can help? *Journal of Banking & Finance*, 29(4):1017–1035.

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

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.

Rizoiu, M.-A., Lee, Y., Mishra, S., and Xie, L. (2017). A tutorial on hawkes processes for events in social media. *arXiv preprint arXiv:1708.06401*.

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

Appendix

A Derivation of log-likelihood with exponential triggering function

Below is a full derivation of the log-likelihood of a Hawkes process when we use the triggering function $g(t) = ae^{-\alpha t}$.

$$
\begin{aligned}
\log\, L(\boldsymbol{\theta}) &= \sum_{i=1}^{N} \log\, \lambda(t_i|\boldsymbol{\theta};\mathcal{H}_t) - \int_0^T \lambda(t|\boldsymbol{\theta};\mathcal{H}_t)\, dt \\
&= \sum_{i=1}^{N} \log\Big(\mu + \sum_{t_j<t_i} ae^{-\alpha(t_i-t_j)}\Big) - \int_0^T \Big(\mu + \sum_{t_i<t} ae^{-\alpha(t-t_i)}\Big)\, dt \\
&= \sum_{i=1}^{N} \log\Big(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\Big) - \int_0^T \Big(\mu + \int_0^t ae^{-\alpha(t-s)}\, dN(s)\Big)\, dt \\
&= \sum_{i=1}^{N} \log\Big(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\Big) - \int_0^T \mu\, dt - \int_0^T\int_0^t ae^{-\alpha(t-s)}\, dN(s)\, dt \\
&= \sum_{i=1}^{N} \log\Big(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\Big) - \big[\mu t\big]_{t=0}^{t=T} - \int_0^T\int_s^T ae^{-\alpha(t-s)}\, dt\, dN(s) \\
&= \sum_{i=1}^{N} \log\Big(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\Big) - (\mu T - 0) - \int_0^T \Big[-\frac{a}{\alpha}e^{-\alpha(t-s)}\Big]_{t=s}^{t=T}\, dN(s) \\
&= \sum_{i=1}^{N} \log\Big(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\Big) - \mu T + \int_0^T \frac{a}{\alpha}\Big(e^{-\alpha(T-s)} - e^{-\alpha(s-s)}\Big)\, dN(s) \\
&= \sum_{i=1}^{N} \log\Big(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\Big) - \mu T + \frac{a}{\alpha}\int_0^T \Big(e^{-\alpha(T-s)} - 1\Big)\, dN(s) \\
&= \sum_{i=1}^{N} \log\Big(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\Big) - \mu T + \frac{a}{\alpha}\sum_{t_i<T} \Big(e^{-\alpha(T-t_i)} - 1\Big)
\end{aligned}
$$

In the fifth line we have changed the order of integration, which can be done by noting that the region over which we integrate can both be written as a type I region (i.e. $\mathrm{D} = \{(t,s) \in \mathbb{R}^2 : 0 \leq t \leq T, 0 \leq s \leq t\}$) and as a type II region (i.e. $\mathrm{D} = \{(t,s) \in \mathbb{R}^2 : s \leq t \leq T, 0 \leq s \leq T\}$), and subsequently applying Fubini's theorem.

B Derivation of the gradient and Hessian of the log-likelihood

Using Equation (3) for the log-likelihood, we can derive the first and second derivatives of the log-likelihood with respect to its parameters. We note that $\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial a \partial \mu}$, $\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \alpha \partial \mu}$, and $\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \alpha \partial a}$ do not need to be derived separately, because of the symmetry of second derivatives.

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial \mu} = \sum_{i=1}^{N} \left(\frac{1}{\mu + a \sum_{t_j < t_i} e^{-\alpha(t_i - t_j)}} \right) - T$$

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial a} = \sum_{i=1}^{N} \left(\frac{\sum_{t_j < t_i} e^{-\alpha(t_i - t_j)}}{\mu + a \sum_{t_j < t_i} e^{-\alpha(t_i - t_j)}} \right) + \frac{1}{\alpha} \sum_{t_i < T} \left(e^{-\alpha(T - t_i)} - 1 \right)$$

$$\begin{aligned}
\frac{\partial \log L(\boldsymbol{\theta})}{\partial \alpha} &= \sum_{i=1}^{N} \left(\frac{\frac{\partial}{\partial \alpha}\left(\mu + a \sum_{t_j < t_i} e^{-\alpha(t_i - t_j)}\right)}{\mu + a \sum_{t_j < t_i} e^{-\alpha(t_i - t_j)}} \right) + \\
&\quad \left(- \frac{a}{\alpha^2} \sum_{t_i < T} \left(e^{-\alpha(T - t_i)} - 1 \right) - \frac{a}{\alpha} \sum_{t_i < T} (T - t_i) e^{-\alpha(T - t_i)} \right) \\
&= - \sum_{i=1}^{N} \left(\frac{a \sum_{t_j < t_i} (t_i - t_j) e^{-\alpha(t_i - t_j)}}{\mu + a \sum_{t_j < t_i} e^{-\alpha(t_i - t_j)}} \right) - \frac{a}{\alpha} \sum_{t_i < T} \left(\frac{1}{\alpha} \left(e^{-\alpha(T - t_i)} - 1 \right) + (T - t_i) e^{-\alpha(T - t_i)} \right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \mu^2} &= \frac{\partial}{\partial \mu} \left(\frac{\partial \log L(\boldsymbol{\theta})}{\partial \mu} \right) \\
&= \frac{\partial}{\partial \mu} \left(\sum_{i=1}^{N} \left(\frac{1}{\mu + a \sum_{t_j < t_i} e^{-\alpha(t_i - t_j)}} \right) - T \right) \\
&= - \sum_{i=1}^{N} \left(\frac{1}{\mu + a \sum_{t_j < t_i} e^{-\alpha(t_i - t_j)}} \right)^2
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \mu \partial a} &= \frac{\partial}{\partial a}\left(\frac{\partial \log L(\boldsymbol{\theta})}{\partial \mu}\right) \\
&= \frac{\partial}{\partial a}\left(\sum_{i=1}^{N}\left(\frac{1}{\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}}\right) - T\right) \\
&= -\sum_{i=1}^{N} \frac{\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}}{\left(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\right)^2}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \mu \partial \alpha} &= \frac{\partial}{\partial \alpha}\left(\frac{\partial \log L(\boldsymbol{\theta})}{\partial \mu}\right) \\
&= \frac{\partial}{\partial \alpha}\left(\sum_{i=1}^{N}\left(\frac{1}{\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}}\right) - T\right) \\
&= \sum_{i=1}^{N}\left(-\frac{\frac{\partial}{\partial \alpha}\left(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\right)}{\left(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\right)^2}\right) \\
&= \sum_{i=1}^{N}\left(-\frac{-a\sum_{t_j<t_i}(t_i-t_j)e^{-\alpha(t_i-t_j)}}{\left(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\right)^2}\right) \\
&= \sum_{i=1}^{N} \frac{a\sum_{t_j<t_i}(t_i-t_j)e^{-\alpha(t_i-t_j)}}{\left(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\right)^2}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial a^2} &= \frac{\partial}{\partial a}\left(\frac{\partial \log L(\boldsymbol{\theta})}{\partial a}\right) \\
&= \frac{\partial}{\partial a}\left(\sum_{i=1}^{N}\left(\frac{\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}}{\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}}\right) + \frac{1}{\alpha}\sum_{t_i<T}\left(e^{-\alpha(T-t_i)} - 1\right)\right) \\
&= \sum_{i=1}^{N}\left(-\frac{\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}}{\left(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\right)^2}\right) \\
&= -\sum_{i=1}^{N}\left(\frac{\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}}{\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}}\right)^2
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial a \partial \alpha} &= \frac{\partial}{\partial \alpha}\left(\frac{\partial \log L(\boldsymbol{\theta})}{\partial a}\right) \\
&= \frac{\partial}{\partial \alpha}\left(\sum_{i=1}^{N}\left(\frac{\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}}{\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}}\right) + \frac{1}{\alpha}\sum_{t_i<T}\left(e^{-\alpha(T-t_i)} - 1\right)\right) \\
&= \sum_{i=1}^{N}\left(\frac{\left(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\right)\frac{\partial}{\partial \alpha}\left(\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\right)}{\left(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\right)^2} - \right. \\
&\qquad \left. \frac{\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\frac{\partial}{\partial \alpha}\left(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\right)}{\left(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\right)^2}\right) + \\
&\quad \left(-\frac{1}{\alpha^2}\left(\sum_{t_i<T} e^{-\alpha(T-t_i)} - 1\right) - \frac{1}{\alpha}\sum_{t_i<T}(T-t_i)e^{-\alpha(T-t_i)}\right) \\
&= \sum_{i=1}^{N}\left(\frac{\left(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\right)\left(-\sum_{t_j<t_i}(t_i-t_j)e^{-\alpha(t_i-t_j)}\right)}{\left(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\right)^2} - \right. \\
&\qquad \left. \frac{\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\left(-a\sum_{t_j<t_i}(t_i-t_j)e^{-\alpha(t_i-t_j)}\right)}{\left(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\right)^2}\right) + \\
&\quad -\frac{1}{\alpha}\sum_{t_i<T}\left(\frac{1}{\alpha}\left(e^{-\alpha(T-t_i)} - 1\right) + (T-t_i)e^{-\alpha(T-t_i)}\right) \\
&= \sum_{i=1}^{N}\left(\frac{-\sum_{t_j<t_i}(t_i-t_j)e^{-\alpha(t_i-t_j)}}{\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}} + \right. \\
&\qquad \left. \frac{a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\sum_{t_j<t_i}(t_i-t_j)e^{-\alpha(t_i-t_j)}}{\left(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\right)^2}\right) + \\
&\quad -\frac{1}{\alpha}\sum_{t_i<T}\left(\frac{1}{\alpha}\left(e^{-\alpha(T-t_i)} - 1\right) + (T-t_i)e^{-\alpha(T-t_i)}\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \alpha^2} &= \frac{\partial}{\partial \alpha}\left(\frac{\partial \log L(\boldsymbol{\theta})}{\partial \alpha}\right) \\
&= \frac{\partial}{\partial \alpha}\Bigg(-\sum_{i=1}^{N}\left(\frac{a\sum_{t_j<t_i}(t_i-t_j)e^{-\alpha(t_i-t_j)}}{\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}}\right) - \\
&\qquad \frac{a}{\alpha}\sum_{t_i<T}\left(\frac{1}{\alpha}\left(e^{-\alpha(T-t_i)}-1\right) + (T-t_i)e^{-\alpha(T-t_i)}\right)\Bigg) \\
&= -\sum_{i=1}^{N}\Bigg(\frac{\left(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\right)\left(-a\sum_{t_j<t_i}(t_i-t_j)^2 e^{-\alpha(t_i-t_j)}\right)}{\left(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\right)^2} - \\
&\qquad \frac{a\sum_{t_j<t_i}(t_i-t_j)e^{-\alpha(t_i-t_j)}\left(-a\sum_{t_j<t_i}(t_i-t_j)e^{-\alpha(t_i-t_j)}\right)}{\left(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}\right)^2}\Bigg) - \\
&\quad \Bigg(-\frac{a}{\alpha^2}\sum_{t_i<T}\left(\frac{1}{\alpha}\left(e^{-\alpha(T-t_i)}-1\right) + (T-t_i)e^{-\alpha(T-t_i)}\right) + \\
&\quad \frac{1}{\alpha}\sum_{t_i<T}\left(-\frac{1}{\alpha^2}\left(e^{-\alpha(T-t_i)}-1\right) - \frac{1}{\alpha}(T-t_i)e^{-\alpha(T-t_i)} - (T-t_i)^2 e^{-\alpha(T-t_i)}\right)\Bigg) \\
&= \sum_{i=1}^{N}\left(\frac{a\sum_{t_j<t_i}(t_i-t_j)^2 e^{-\alpha(t_i-t_j)}}{\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}} - \left(\frac{a\sum_{t_j<t_i}(t_i-t_j)e^{-\alpha(t_i-t_j)}}{\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}}\right)^2\right) + \\
&\quad \frac{a}{\alpha}\sum_{t_i<T}\left(\frac{2}{\alpha^2}\left(e^{-\alpha(T-t_i)}-1\right) + \frac{2}{\alpha}(T-t_i)e^{-\alpha(T-t_i)} + (T-t_i)^2 e^{-\alpha(T-t_i)}\right)
\end{aligned}
$$

We can obtain the gradient and Hessian by filling in the partial derivatives. The gradient is defined by

$$
\frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial \log L(\boldsymbol{\theta})}{\partial \mu} \\ \frac{\partial \log L(\boldsymbol{\theta})}{\partial a} \\ \frac{\partial \log L(\boldsymbol{\theta})}{\partial \alpha} \end{bmatrix},
$$

and the Hessian is defined by

$$
\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \begin{bmatrix} \frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \mu^2} & \frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \mu \partial a} & \frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \mu \partial \alpha} \\ \frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial a \partial \mu} & \frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial a^2} & \frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial a \partial \alpha} \\ \frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \alpha \partial \mu} & \frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \alpha \partial a} & \frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \alpha^2} \end{bmatrix}.
$$

C Derivation of the probability of the occurrence of an event

Below is a full derivation of the probability of the occurrence of an event between times t_{n-1} and t_n, when the arrivals of those events follow a Hawkes process with the conditional intensity of Equation (2).

$$
\begin{aligned}
\mathrm{P}(N(t_n) - N(t_{n-1}) > 0) = F(t \leq t_n - t_{n-1}) &= 1 - \exp\Big(- \int_{t_{n-1}}^{t_n} \lambda(t|\boldsymbol{\theta}; \mathcal{H}_t)\; dt \Big) \\
&= 1 - \exp\Big(- \int_{t_{n-1}}^{t_n} \Big(\mu + \sum_{t_i < t} a e^{-\alpha(t-t_i)}\Big)\; dt \Big) \\
&= 1 - \exp\Big(- \int_{t_{n-1}}^{t_n} \Big(\mu + \int_0^t a e^{-\alpha(t-s)}\; dN(s)\Big)\; dt \Big) \\
&= 1 - \exp\Big(- \int_{t_{n-1}}^{t_n} \mu\; dt \; - \int_{t_{n-1}}^{t_n} \int_0^t a e^{-\alpha(t-s)}\; dN(s)\; dt \Big) \\
&= 1 - \exp\Big(- \big[\mu t\big]_{t=t_{n-1}}^{t=t_n} \; - \Big(\int_0^{t_n} \int_0^t a e^{-\alpha(t-s)}\; dN(s)\; dt \; - \\
&\qquad \int_0^{t_{n-1}} \int_0^t a e^{-\alpha(t-s)}\; dN(s)\; dt \Big)\Big) \\
&= 1 - \exp\Big(- \mu(t_n - t_{n-1}) \; - \int_0^{t_n} \int_0^t a e^{-\alpha(t-s)}\; dN(s)\; dt \; + \\
&\qquad \int_0^{t_{n-1}} \int_0^t a e^{-\alpha(t-s)}\; dN(s)\; dt \Big) \\
&= 1 - \exp\Big(- \mu(t_n - t_{n-1}) \; - \int_0^{t_n} \int_s^{t_n} a e^{-\alpha(t-s)}\; dt\; dN(s) \; + \\
&\qquad \int_0^{t_{n-1}} \int_s^{t_{n-1}} a e^{-\alpha(t-s)}\; dt\; dN(s) \Big) \\
&= 1 - \exp\Big(- \mu(t_n - t_{n-1}) \; - \int_0^{t_n} \Big[- \frac{a}{\alpha} e^{-\alpha(t-s)} \Big]_{t=s}^{t=t_n} \; dN(s) \; + \\
&\qquad \int_0^{t_{n-1}} \Big[- \frac{a}{\alpha} e^{-\alpha(t-s)} \Big]_{t=s}^{t=t_{n-1}} \; dN(s) \Big)
\end{aligned}
$$

$$
\begin{aligned}
&= 1 - \exp\Bigg(- \mu(t_n - t_{n-1}) \; + \; \int_0^{t_n} \frac{a}{\alpha}\Big(e^{-\alpha(t_n - s)} - e^{-\alpha(s-s)}\Big)\, dN(s) - \\
&\qquad \int_0^{t_{n-1}} \frac{a}{\alpha}\Big(e^{-\alpha(t_{n-1} - s)} - e^{-\alpha(s-s)}\Big)\, dN(s)\Bigg) \\
&= 1 - \exp\Bigg(- \mu(t_n - t_{n-1}) \; + \; \frac{a}{\alpha}\int_0^{t_n} \Big(e^{-\alpha(t_n - s)} - 1\Big)\, dN(s) - \\
&\qquad \frac{a}{\alpha}\int_0^{t_{n-1}} \Big(e^{-\alpha(t_{n-1} - s)} - 1\Big)\, dN(s)\Bigg) \\
&= 1 - \exp\Bigg(- \mu(t_n - t_{n-1}) \; + \; \frac{a}{\alpha}\sum_{t_i < t_n} \Big(e^{-\alpha(t_n - t_i)} - 1\Big) - \\
&\qquad \frac{a}{\alpha}\sum_{t_i < t_{n-1}} \Big(e^{-\alpha(t_{n-1} - t_i)} - 1\Big)\Bigg)
\end{aligned}
$$

In the seventh line we have changed the order of integration, which can (just as in Appendix A) be done by noting that the region over which we integrate can both be written as a type I region and as a type II region, and subsequently applying Fubini's theorem.

D Figures of KSS against thresholds for S&P 500 data

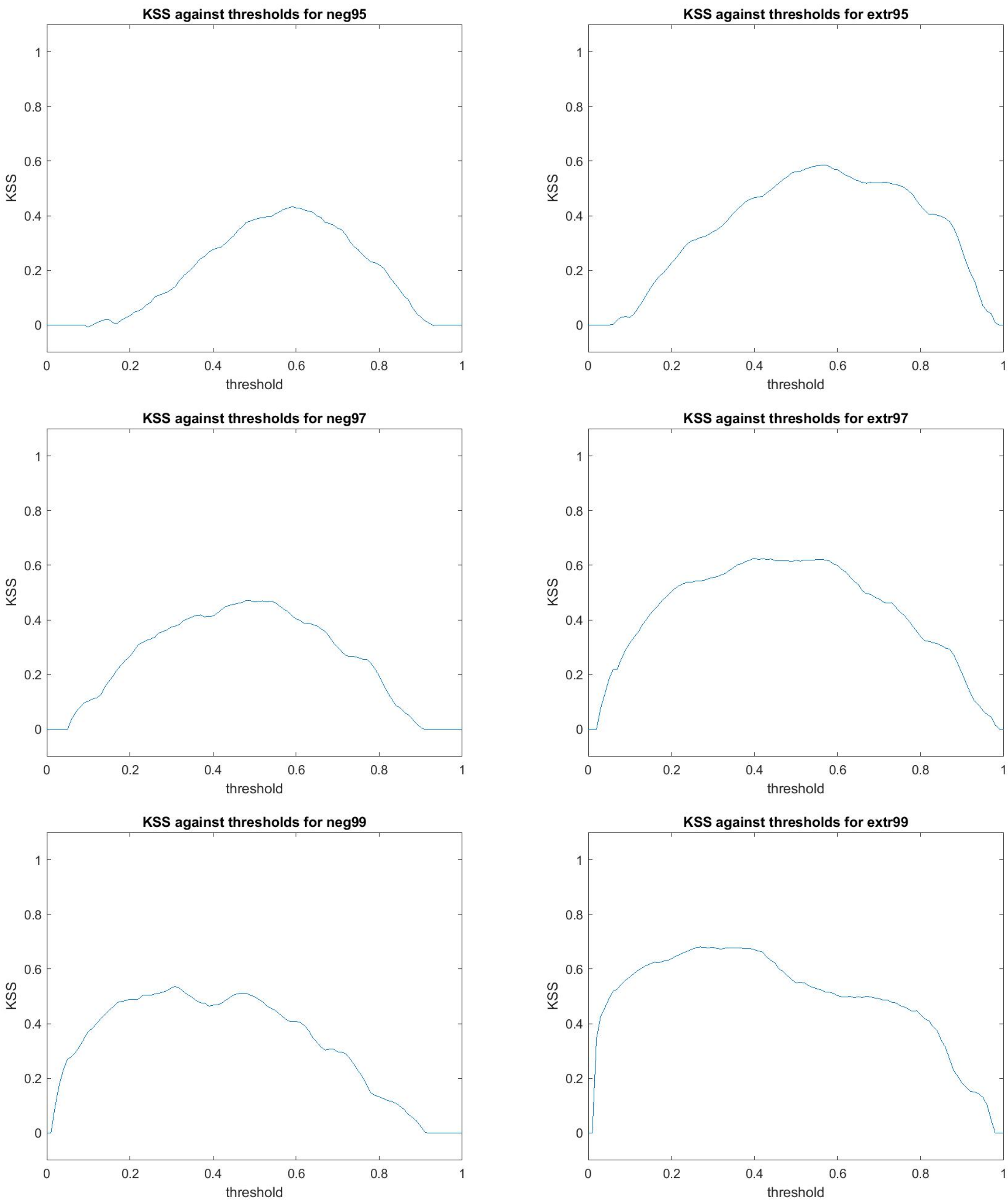

Figure 3: Hanssen-Kuiper Skill Scores for different thresholds of the EWS based on an ETAS model with a conditional intensity of the arrivals as specified in Equation (2) when predicting a crash in the next five days, from 2 September 2008 until 1 January 2013. An event is predicted when the estimated probability of an event in the next five days exceeds the threshold.

E Derivation of the alternative form of a GARCH(1,1) model

Below is a derivation of the alternative form in Equation (5) to write the conditional variance in a GARCH(1,1) model. We can take expectations in the GARCH(1,1) model to obtain that

$$\begin{aligned}\mathbb{E}[\sigma_t^2] &= \mathbb{E}[\omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2] \\ &= \omega + \alpha\mathbb{E}[\epsilon_{t-1}^2] + \beta\mathbb{E}[\sigma_{t-1}^2].\end{aligned}$$

If we then use that $\mathbb{E}[\epsilon_t^2]$ and $\mathbb{E}[\sigma_t^2]$ are both equal to $\bar{\sigma}^2$ for all times t, we obtain that

$$\bar{\sigma}^2 = \omega + \alpha\bar{\sigma}^2 + \beta\bar{\sigma}^2,$$

from which we see that

$$\bar{\sigma}^2 = \frac{\omega}{1-\alpha-\beta},$$

or, equivalently,

$$\omega = \bar{\sigma}^2(1-\alpha-\beta).$$

Plugging this expression into the equation of the conditional variance in the GARCH(1,1) model gives us

$$\begin{aligned}\sigma_t^2 &= \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \\ &= \bar{\sigma}^2(1-\alpha-\beta) + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \\ &= \bar{\sigma}^2 + \alpha(\epsilon_{t-1}^2 - \bar{\sigma}^2) + \beta(\sigma_{t-1}^2 - \bar{\sigma}^2) \\ &= \bar{\sigma}^2 + \alpha(\epsilon_{t-1}^2 - \sigma_{t-1}^2) + (\alpha+\beta)(\sigma_{t-1}^2 - \bar{\sigma}^2),\end{aligned}$$

as required.

F Derivation of the alternative form of a GJR(1,1) model

Below is a derivation of the alternative form in Equation (6) to write the conditional variance in a GJR(1,1) model. We can take expectations in the GJR(1,1) model to obtain that

$$\begin{aligned}\mathbb{E}[\sigma_t^2] &= \mathbb{E}[\omega + \alpha\epsilon_{t-1}^2 + \gamma\mathbb{I}\{\epsilon_{t-1} < 0\}\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2] \\ &= \omega + \alpha\mathbb{E}[\epsilon_{t-1}^2] + \gamma\mathbb{E}[\mathbb{I}\{\epsilon_{t-1} < 0\}\epsilon_{t-1}^2] + \beta\mathbb{E}[\sigma_{t-1}^2] \\ &= \omega + \alpha\mathbb{E}[\epsilon_{t-1}^2] + \frac{\gamma}{2}\mathbb{E}[\epsilon_{t-1}^2] + \beta\mathbb{E}[\sigma_{t-1}^2],\end{aligned}$$

where we used in the last line that ϵ_{t-1} $(= z_{t-1}\sigma_{t-1})$ is symmetrically distributed, as it has a (scaled) Student-t distribution (by assumption), which is symmetric around 0. If we then use that $\mathbb{E}[\epsilon_t^2]$ and $\mathbb{E}[\sigma_t^2]$ are both equal to $\bar{\sigma}^2$ for all times t, we obtain that

$$\bar{\sigma}^2 = \omega + \alpha\bar{\sigma}^2 + \frac{\gamma}{2}\bar{\sigma}^2 + \beta\bar{\sigma}^2,$$

from which we see that

$$\bar{\sigma}^2 = \frac{\omega}{1 - \alpha - \beta - \frac{\gamma}{2}},$$

or, equivalently,

$$\omega = \bar{\sigma}^2\left(1 - \alpha - \beta - \frac{\gamma}{2}\right).$$

Plugging this expression into the equation of the conditional variance in the GJR(1,1) model gives us

$$\begin{aligned}\sigma_t^2 &= \omega + \alpha\epsilon_{t-1}^2 + \gamma\mathbb{I}\{\epsilon_{t-1} < 0\}\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \\ &= \bar{\sigma}^2\left(1 - \alpha - \beta - \frac{\gamma}{2}\right) + \alpha\epsilon_{t-1}^2 + \gamma\mathbb{I}\{\epsilon_{t-1} < 0\}\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \\ &= \bar{\sigma}^2 + \alpha(\epsilon_{t-1}^2 - \bar{\sigma}^2) + \gamma\left(\mathbb{I}\{\epsilon_{t-1} < 0\}\epsilon_{t-1}^2 - \frac{1}{2}\bar{\sigma}^2\right) + \beta(\sigma_{t-1}^2 - \bar{\sigma}^2) \\ &= \bar{\sigma}^2 + \alpha(\epsilon_{t-1}^2 - \sigma_{t-1}^2) + \gamma\left(\mathbb{I}\{\epsilon_{t-1} < 0\}\epsilon_{t-1}^2 - \frac{1}{2}\sigma_{t-1}^2\right) + \left(\alpha + \beta + \frac{\gamma}{2}\right)(\sigma_{t-1}^2 - \bar{\sigma}^2),\end{aligned}$$

as required.

G Figures of KSS against thresholds obtained from simulation

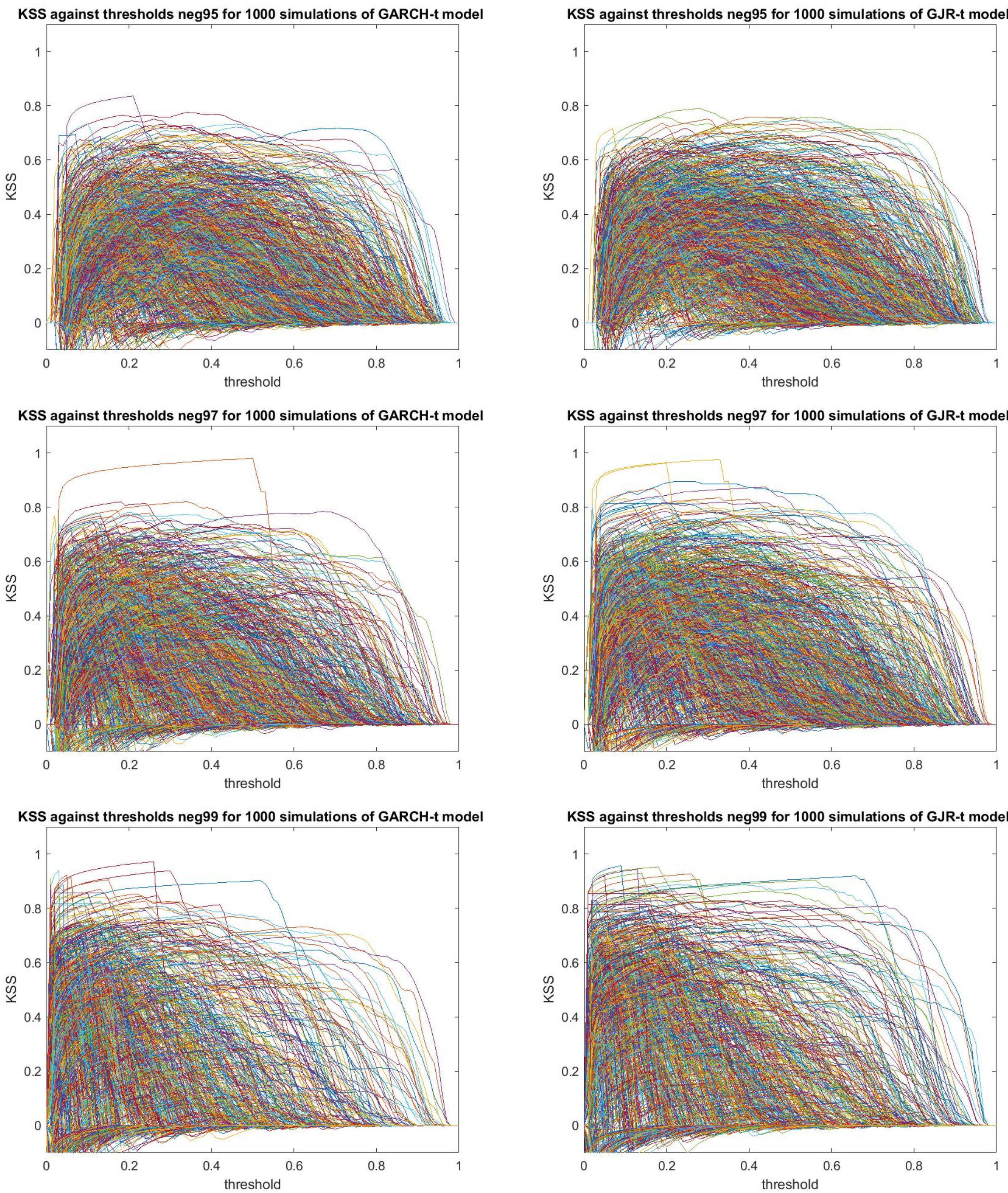

Figure 4: Hanssen-Kuiper Skill Scores of 1000 simulations for different thresholds of the EWS based on an ETAS model with a conditional intensity of the arrivals as specified in Equation (2) when predicting a crash in the next five days, over 1087 trading days. A crash is predicted when the estimated probability of a crash in the next five days exceeds the threshold.

H Estimation results obtained by simulation with other sample sizes

Table 7: Estimation results of model with exponential triggering function obtained by simulation

Simulation model	Arrival choice	Neg 95	Neg 97	Neg 99
	$\hat{\mu}$	0.0101	0.0060	0.0022
		(0.0027)	(0.0019)	(0.0010)
GARCH-t	$\hat{a}$	0.0286	0.0287	0.0271
		(0.0036)	(0.0042)	(0.0063)
	$\hat{\alpha}$	0.0358	0.0360	0.0349
		(0.0048)	(0.0056)	(0.0077)
	$\hat{\mu}$	0.0107	0.0064	0.0023
		(0.0027)	(0.0019)	(0.0009)
GJR-t	$\hat{a}$	0.0339	0.0345	0.0332
		(0.0039)	(0.0045)	(0.0071)
	$\hat{\alpha}$	0.0432	0.0439	0.0432
		(0.0053)	(0.0060)	(0.0088)

The estimation results are obtained by applying an ETAS model with the conditional intensity in Equation (2) to different quantiles of extreme negative returns of simulated returns over 10000 trading days. The first column indicates which model is used to simulate returns and the first row indicates to which quantile of negative returns the ETAS model is applied. The reported parameter estimates are the averages over 1000 simulations. In between parentheses are the standard deviations of the estimates obtained in the 1000 simulations.

I Prediction results obtained by simulation with other sample sizes

Table 8: Prediction results of model with exponential triggering function obtained by simulation

Simulation model	Arrival choice	Neg 95	Neg 97	Neg 99
	maximum average KSS	0.364	0.395	0.423
		(0.004)	(0.005)	(0.007)
GARCH-t	optimal threshold	0.25	0.16	0.06
	average KSS for threshold 0.5	0.260	0.207	0.124
		(0.005)	(0.006)	(0.007)
	maximum average KSS	0.377	0.416	0.437
		(0.004)	(0.005)	(0.007)
GJR-t	optimal threshold	0.25	0.16	0.07
	average KSS for threshold 0.5	0.282	0.231	0.144
		(0.005)	(0.006)	(0.007)

The Hanssen-Kuiper Skill Scores are obtained by predicting a crash in the next five days for 996 simulated trading days (starting at day 10001), based on an ETAS model with the conditional intensity in Equation (2). Crashes are predicted when the estimated probability of a crash in the next five days exceeds the threshold. The first column indicates which model is used to simulate returns and the first row indicates to which quantile of negative returns the ETAS model is applied. The reported average KSSs are averages over 1000 simulations. In between parentheses are the standard deviations of average KSSs over 1000 simulations.

J Figures of average KSS against thresholds obtained by simulation

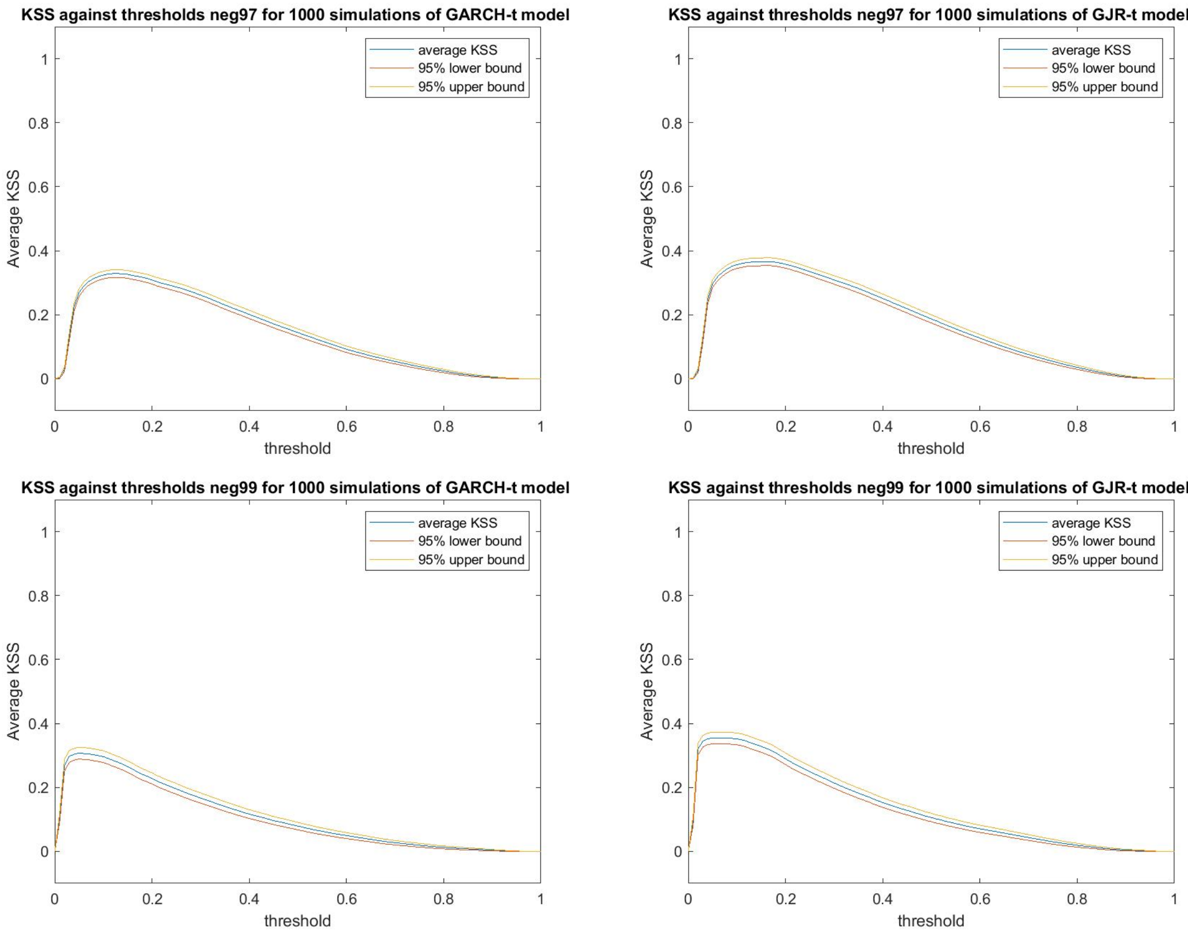

Figure 5: Average Hanssen-Kuiper Skill Scores (and 95% confidence bounds) based on 1000 simulations for different thresholds of the EWS based on an ETAS model with a conditional intensity of the arrivals as specified in Equation (2) when predicting a crash in the next five days, over 1087 trading days. A crash is predicted when the estimated probability of a crash in the next five days exceeds the threshold.

K Figures of individual asset paths obtained by simulation

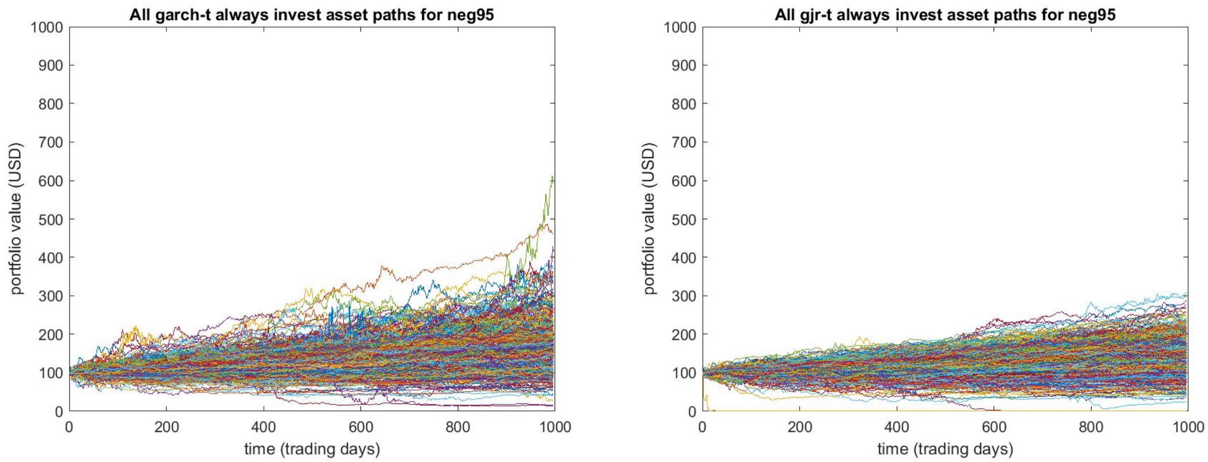

Figure 6: All individual asset paths over 996 trading days based on 1000 simulations, when we just keep our money invested all the time. Investment value at the start is set to 100.

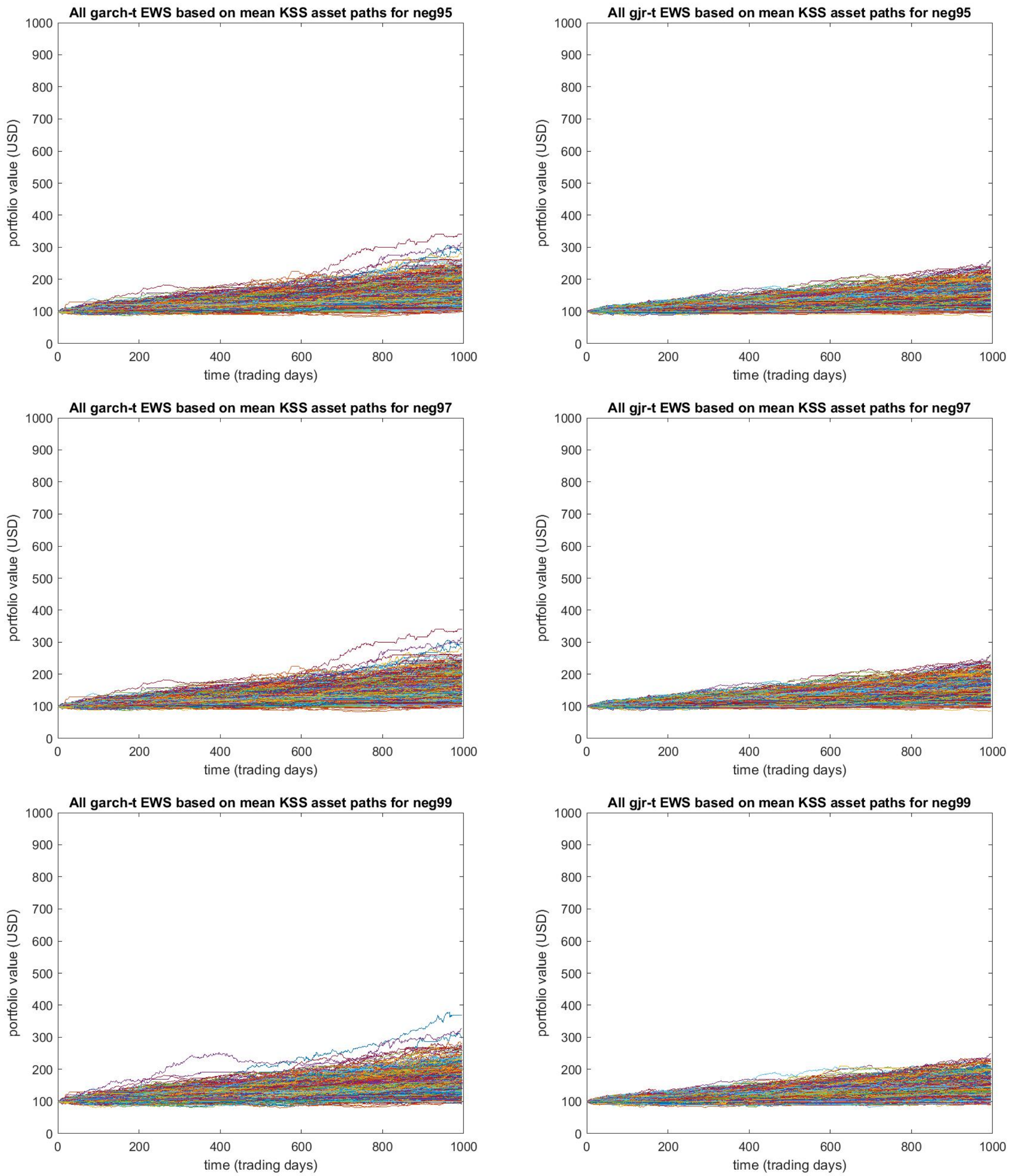

Figure 7: All individual asset paths over 996 trading days based on 1000 simulations. The investment strategy applied here makes use of an EWS based on an ETAS model with a conditional intensity of the arrivals as specified in Equation (2). In particular, money is withdrawn if the estimated probability of a crash in the next five days exceeds the optimal threshold based on the average KSS over all simulations. Investment value at the start is set to 100.

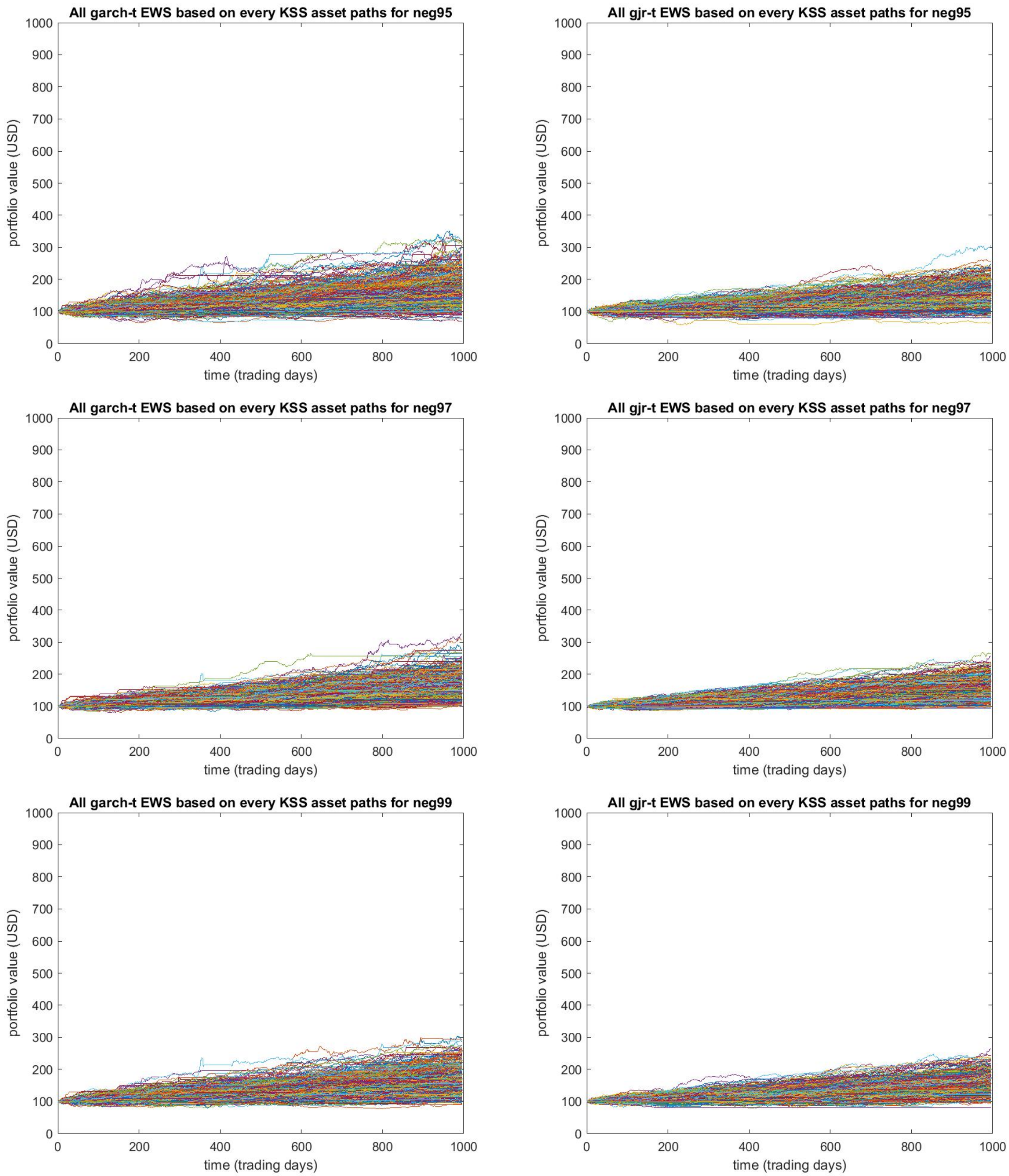

Figure 8: All individual asset paths over 996 trading days based on 1000 simulations. The investment strategy applied here makes use of an EWS based on an ETAS model with a conditional intensity of the arrivals as specified in Equation (2). In particular, money is withdrawn if the estimated probability of a crash in the next five days exceeds the optimal threshold based on the KSS of that asset over all simulations. Investment value at the start is set to 100.

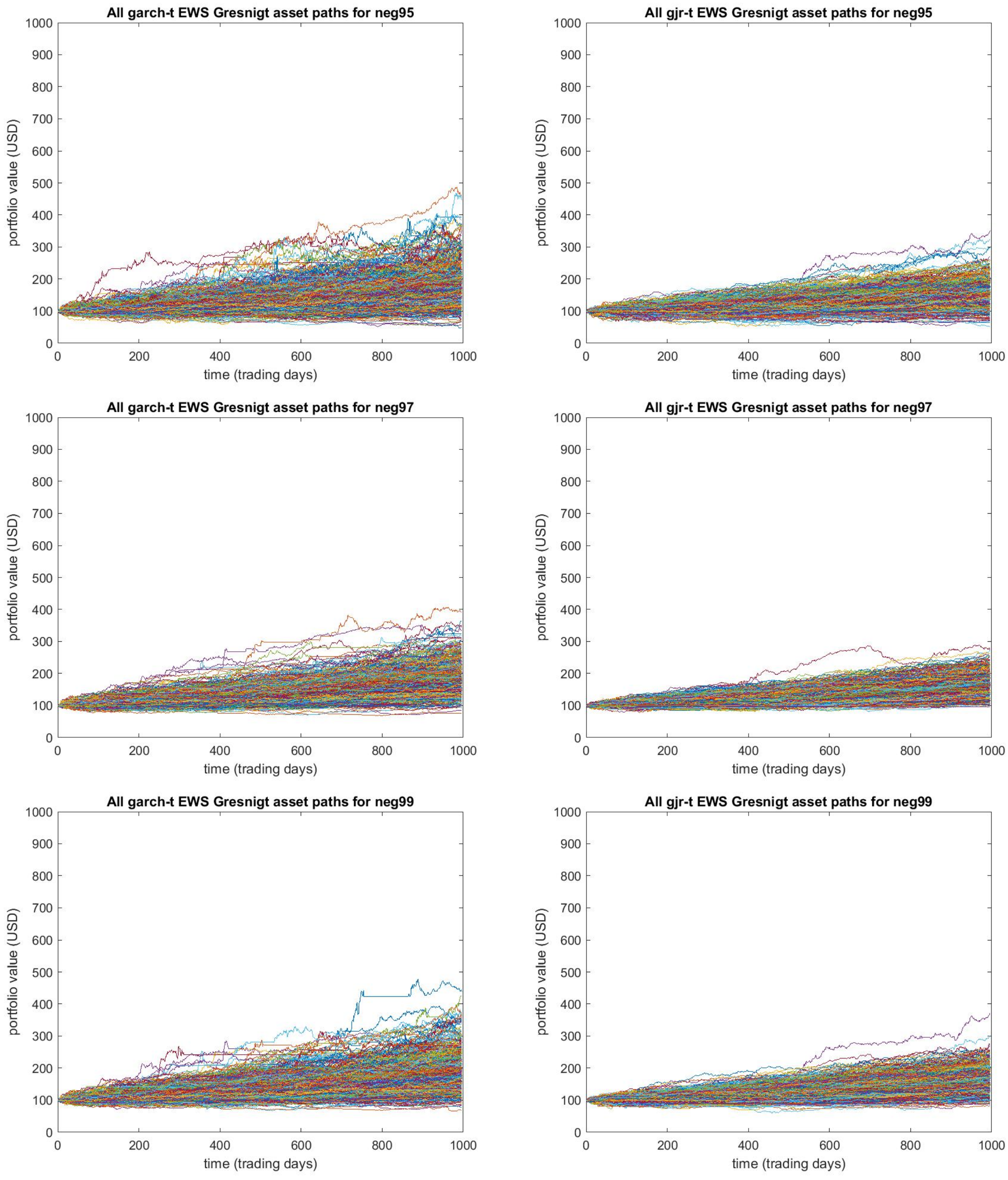

Figure 9: All individual asset paths over 996 trading days based on 1000 simulations. The investment strategy applied here makes use of an EWS based on an ETAS model with a conditional intensity of the arrivals as specified in Equation (2). In particular, money is withdrawn if the estimated probability of a crash in the next five days exceeds the value 0.5. Investment value at the start is set to 100.

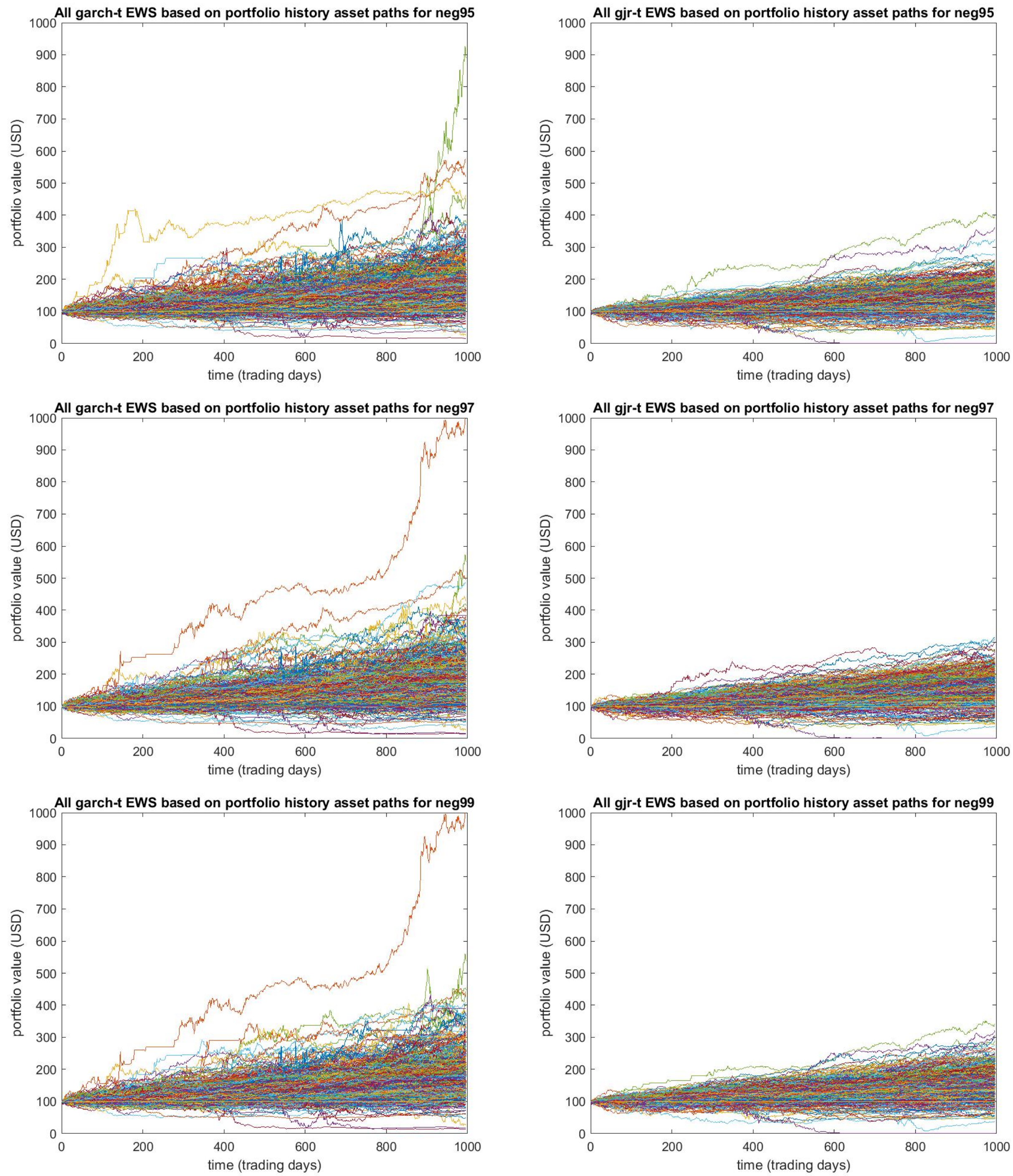

Figure 10: All individual asset paths over 996 trading days based on 1000 simulations. The investment strategy applied here makes use of an EWS based on an ETAS model with a conditional intensity of the arrivals as specified in Equation (2). In particular, money is withdrawn if the estimated probability of a crash in the next five days exceeds the threshold that maximized the portfolio value for previous observations of each individual asset over all simulations. Investment value at the start is set to 100.

L Figures of mean asset paths obtained by simulation

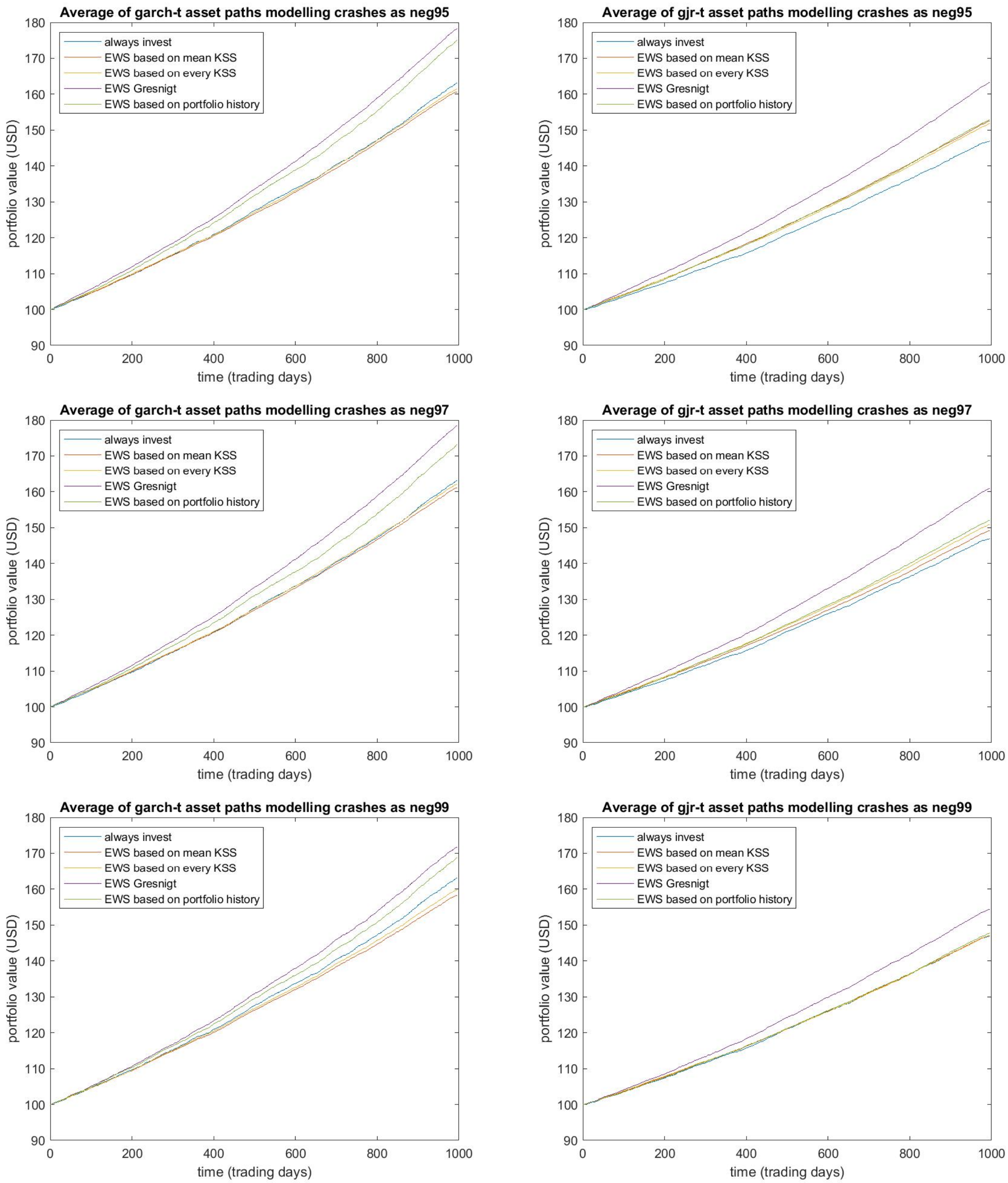

Figure 11: Average asset paths over 996 trading days based on 1000 simulations. Investment strategies make use of different EWSs based on an ETAS model with a conditional intensity of the arrivals as specified in Equation (2). Investment value at the start is set to 100.

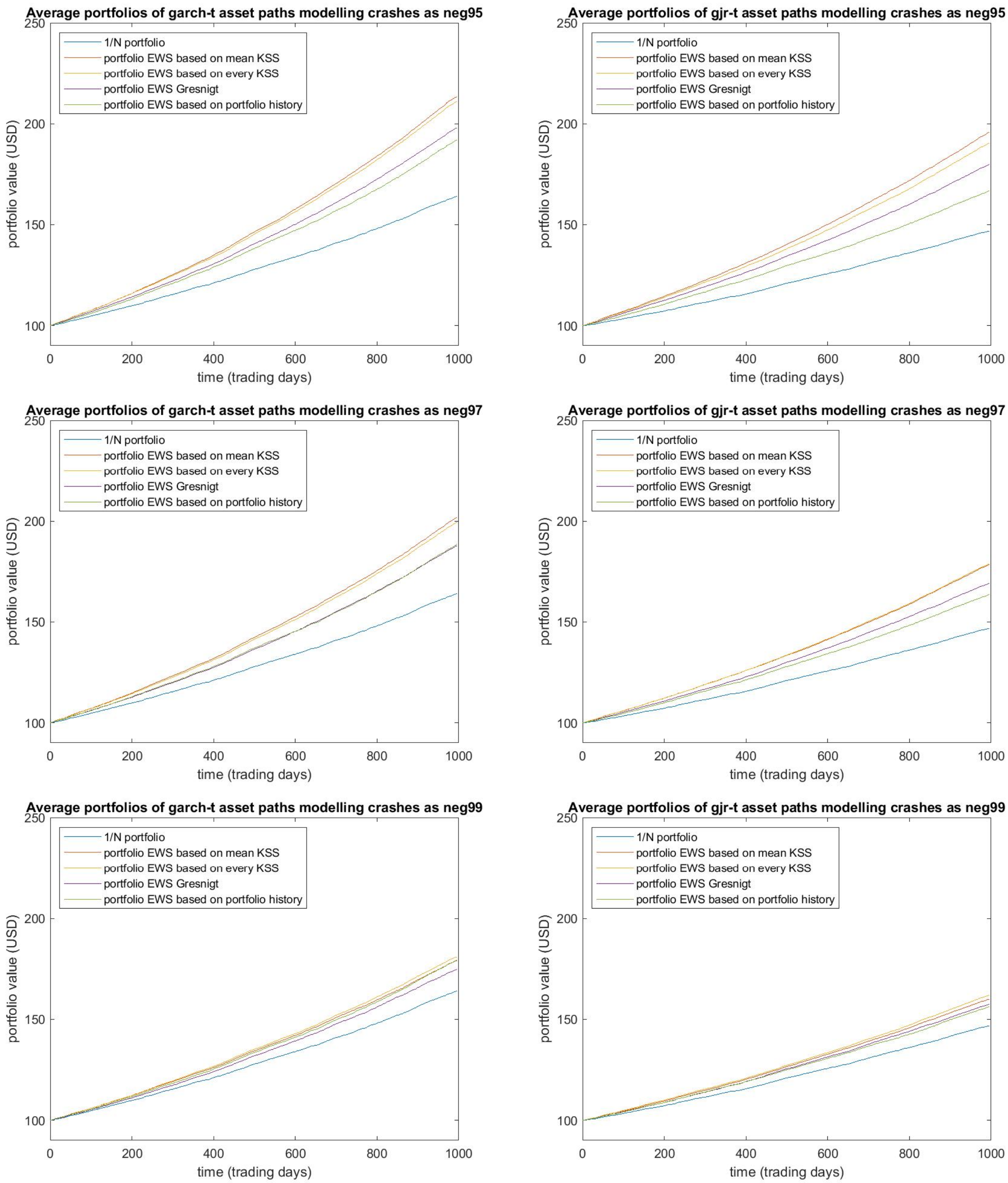

Figure 12: Average portfolio value of simulated assets paths over 996 trading days based on 100 simulated portfolios. Investment strategies make use of different EWSs based on an ETAS model with a conditional intensity of the arrivals as specified in Equation (2). Portfolio value at the start is set to 100.

M Programs used in this thesis

Below is a list of all MATLAB programs that we have used during our research.

- *arrival_maker*: generates vectors with the times and magnitudes of (negative) extreme returns
- *compute_KSS*: computes the Hanssen-Kuiper Skill Score (KSS) (and additionally the hit rate and false alarm rate) for certain estimated crash probabilities, crashdata and a threshold
- *estimate_crash_probabilities*: estimates the probabilities of a crash in a certain time interval for each time point and returns a logical vector that takes the value 1 if a crash occurs in the next time interval and the value 0 otherwise
- *estimate_parameters*: estimates the parameters in an Epidemic-Type Aftershock Sequence (ETAS) model with maximum likelihood estimation
- *fminsearchbnd*: uses the Nelder-Mead simplex direct search algorithm to find a minimum in a problem with parameter restrictions
- *gradient_ogata*: calculates the gradient of the log likelihood function in Equation (3)
- *Hessian_ogata*: calculates the Hessian of the log likelihood function in Equation (3)
- *main_program*: determines the optimal threshold value for predicting crashes and computes wealth matrices of different investment strategies
- *main_program_sp500*: determines the optimal threshold value for predicting crashes and computes the wealth of different investment strategies based on S&P 500 data
- *NegativeLogLikelihood_Ogata*: calculates the negative log-likelihood function (given certain parameters and event times), by taking the negative value of the log-likelihood in Equation (3)
- *one_over_N_portfolio*: loads the simulated returns, constructs $1/N$ portfolios out of them and then calculates the average Sharpe ratio and average Value at Risk for the simulated portfolios
- *performance_measures*: calculates the average Sharpe ratio and average Value at Risk over the different simulations for different investment strategies
- *portfolio_plots*: loads for different ways to model crashes and for different ways to simulate asset returns the relevant wealth matrices, and plots them against time
- *portfolio_program*: determines the optimal threshold value for predicting crashes and computes wealth matrices of different investment strategies that consist of multiple assets
- *simulate_garch_returns*: uses the S&P 500 returns to estimate a GARCH-type model and uses the obtained estimated model to simulate return series