Cover page

(This page is intentionally left blank)

Identifying and Predicting Financial Earthquakes using Hawkes Processes

Francine Gresnigt

Identifying and Predicting Financial Earthquakes using Hawkes Processes

Financiële aardbevingen identificeren en voorspellen met behulp van Hawkes processen

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof. dr. R.C.M.E. Engels

and in accordance with the decision of the Doctorate Board.

The public defense shall be held on

Thursday, February 20, 2020 at 15:30 hours

by

FRANCINE GRESNIGT

born in Utrecht, The Netherlands

Erasmus University Rotterdam

Doctorate Committee

Promotor:	Prof. dr. P.H.B.F. Franses
Co-promotor:	Dr. H.J.W.G Kole
Other members:	Prof. dr. H.P. Boswijk
	Prof. dr. D.J.C. van Dijk
	Prof. dr. C.G. de Vries

ISBN: ... .

This book is no. ... of the Tinbergen Institute Research Series, established through cooperation between Rozenberg Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Acknowledgements

Tijdens de zomer van 2012 wil ik graag als PhD-er aan de slag. Meer leren en mij verder verdiepen in modellen, wiskunde, economie en de combinatie econometrie, leek mij van alle keuzemogelijkheden na het afronden van mijn master econometrie het meest uitdagend en interessant. Niet wetende wat mij te wachten stond, begon ik vol goede moed aan het zoeken van een onderwerp voor mijn eerste artikel. Na een maand besprak een artikel over Hawkes modellen van Aït-Sahalia met mijn professoren. Echter de gedachte om een model uit de aardbevingsliteratuur te nemen om crashes op de aandelingenmarkt te voorspellen, liet ons niet los. Na ruim een half jaar bestudeerde ik het ETAS model (Epidemic Type Aftershock Model) dat het verloop van naschokken van aardbevingen tracht na te bootsen. Wat bleek? Het eerder genoemde Hawkes model is een directe afgeleide van het ETAS model. Dit is zeker niet de eerste keer dat ik ben verdwaald. Onderzoek doen is een zoektocht voor mij geweest, niet zozeer naar het juiste model als wel naar mezelf, naar wat ik wil en welke eisen ik aan mezelf stel. De laatste dwaling heeft ruim 2 jaar gekost. Maar hier is het boekje dan toch. Ik heb veel geleerd van mijn PhD: ik kan zeggen dat ik daar trots op ben. Graag bedank hier iedereen die dit mogelijk heeft gemaakt. Zonder jullie was het zeker niet gelukt!

Erik, bedankt voor al je aanmoedigingen, het pragmatisch meedenken wanneer ik het even niet meer zag, het delen van je eigen ervaringen en het tonen van interesse ook al had ik de universiteit al geruime tijd verlaten. Veel dank ook voor de samenwerking tijdens het werken aan de papers die dit proefschrift omvat. Philip Hans, bedankt voor je vertrouwen in mij vanaf het begin, je optimisme, de ruimte die je me gaf om mijn eigen weg te vinden, maar ook de tijd die je maakte om mij te coachen en gezamenlijke projecten van feedback te voorzien. Door je functie als decaan van de Erasmus School of Economics was je altijd druk (en had je altijd een verhaal klaar liggen), toch kon ik altijd bij je aankloppen. Veel dank ook voor het mogelijk maken van mijn research visit naar Princeton University, en dat

ik deel kon uitmaken van het docententeam van de inmiddels immens populaire MOOC voor Econometrie.

Oud-mede-PhD's, bedankt: Voor alle interessante gesprekken over onderzoek en carrière plannen, maar vooral voor alle gesprekken over al wat in het nieuws was, politiek, vegetarisch dan wel veganistisch zijn, het milieu, vriendjes en vriendinnetjes, trouwplannen etc. Voor het delen van jullie ervaringen met dwalingen tijdens periodes waarin ik mijn laptop uit het raam wilde gooien ;). Sander, ik had me geen fijnere kamergenoot kunnen wensen. Wat hebben we gelachen vanaf het allereerste moment! Bijpraten over alles, en weer door (voor jou vooral dat laatste op het laatst). Bart, ook met jou heb ik kamers gedeeld, in het Tinbergen gebouw, in Toulouse en in Milaan. Je artikel bracht me op ideën, en je was altijd bereid met me mee te denken. Met jou kon ik het hebben over onderzoek, maar ook over kleine en grote life events buiten de universiteit. Tom, met jou ben ik het PhD avontuur begonnen. Samen de wereld van een onderzoeker op de universiteit verkennen, het studentenleven langzaam achter ons latend. Je positviteit, scherpte en doelgerichtheid (bij het oplossen van een wiskundige puzzel, het behalen van een nieuwe hardlooptijd etc.) maakt je een inspirerende collega. Myrthe, heerlijk om af en toe de koffie op de B8 af te wisselen met een Starbucks en dat moment te gebruiken om bij te praten over lesgeven, onderzoek en nog veel meer over alles daarbuiten. Wat heb ik daarnaast fijn met je samengewerkt bij de ontwikkeling van de MOOC, en wat ben ik blij dat wij in jullie oude huis hebben mogen wonen. Gertjan, Carolien, Bruno, Koen, Victor, Didier, Dennis en Matthijs bedankt voor alle leuke koffiepauzes, lunches, vrimibo's en conferenties. Veel van jullie heb ik al mogen zien staan achter de katheder in de promotiezaal. Veel van jullie zijn ook verder gegaan als post-doc of met een tenure track. Trots ben ik daarop. Ik wens jullie het beste toe.

Dick, naast Philip Hans en Erik wil ik jou bedanken voor je interesse, feedback op mijn werk en het meedenken met onderzoeksplannen. Ik zal ons uitstapje naar New York nooit vergeten. De opwinding van je vrouw Leontien over grote M&M stores werkt aanstekelijk. Christiaan, bedankt voor de samenwerking tijdens de MOOC. Ik heb mogen ervaren hoe betrokken je bent bij studenten, de afdeling en mij als collega. Twan, Willem en Remy, het was fijn dat de deur bij jullie altijd open stond.

Vrienden en vriendinnen door dik en dun, bedankt: Liz, voor alle vermakelijke gesprekken met af en toe een steunend woord, onder het genot van een goed glas wijn op de vele terassen in het mooie Rotterdam, bij mij of bij jou thuis. Door jou voelde ik me thuis in Rotterdam.

Leonie, nu collega maar vooral vriendin en straks bijna-buur, voor alle keren dat je mij hebt aangehoord als ik ergens mee zat, je wijze woorden (ja, je hebt vaaaak gelijk), je heerlijke kookkunsten, en het delen van de passie voor goede wijn. Jullie partners, Jelle en Jan, voor jullie humor en goede gesprekken tijdens onze dubbeldates. Marloes, Sabine, Annelies en Malou, voor de vele gezellige etentjes en de voorbereiding op mijn aankomend moederschap (met jullie voel ik me doorlopend zwanger ;)). Rogier en Karel, voor alle dansjes 's nachts en goede gesprekken met hoofdpijn of nog niet geheel nuchter 's ochtends tijdens mijn PhD. Jullie zijn goede vrienden van Vic, maar zeker ook van mij. Alle andere stapmaten en plussen, voor de fantastische kerstdiners, skireizen, borrels, bruiloften en andere feestjes. Ik weet zeker dat we daar nog veel meer fantastisch aan toe gaan voegen. Carlijn, voor het meedenken wanneer ik vast zat in mijn onderzoek, èn het doen van spelletjes met je mooie kids Annemijn en Lauren. Maaike, voor de vele koffie'tjes en biertjes de eerste maanden tijdens mijn PhD als collega en de vele maanden erna als vriendin. Pascal, voor de lachsalvo's die je me hebt bezorgd tijdens mijn laatste maanden op de universiteit. Mijn jaarclub, Charlotte, Wenda, Marieke, Floortje, Nicole, Cathy, Kim, Floor, Lisette, Esther en Laurence, voor alle gezellige borrels en etentjes en het delen van mooie momenten op bruiloften etc. Anne en Lies, ook jullie bedank ik graag. Jullie waren er dan wel niet bij tijdens de jaren op de universiteit, maar ik ben blij dat ik jullie nu tot mijn goede vriendinnen mag rekenen.

Lieve Pap en Mam, naast jullie genen heb ik ook veel levenslessen van jullie meegekregen. Dit maakt me tot wie ik ben. Ik weet dat jullie altijd voor mij klaar staan (ook op stel en sprong zoals bij het afronden van dit boekje en het vernielen van de koppeling van de auto), en dat de lat die ik mezelf opleg niet van jullie komt (cassière bij de Dirk van de Broek is meer dan goed genoeg). Lieve Roos, je bent het beste, liefste en leukste zusje van de wereld zonder dat je daar iets voor hoeft te doen. Zorgzaam, sociaal maar ook eerlijk. Van jou heb ik er maar èèn. Ik kijk uit naar de toekomst, met jou, Bart en Vic, én onze dochters! Ik houd van jullie. Bart, je bent een topgozer voor mijn zusje, én voor Vic en mij. Huub, Hanneke, Philip, Toyah en Miguel, jullie zijn mijn tweede familie. En dat voelt ook zo. Geïnteresseerd, betrokken en vol verhalen, is het altijd heel fijn om bij jullie te zijn.

Liefste Vic, echtgenoot sinds 27-09-2019 en toekomstig vader van onze frummel, mijn partner in crime. Lachen, ruzie maken, huilen, ik doe het het liefste met jou. Jouw enthousiasme, spontaniteit en positiviteit maken mijn leven leuker. Je hebt me geleerd tevreden met mezelf te zijn, en je helpt me herinneren aan hoe dat moet als ik het weer eens vergeet.

Zorgzaam en behulpzaam als je bent heb ik nooit het gevoel dat ik er alleen voor sta. Ik prijs mezelf gelukkig dat ik samen met jou mijn levensreis mag maken.

Contents

Chapter 1

Introduction

1.1 Introduction

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

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

practitioners), earthquakes and stock returns therefore share characteristics observable as the clustering of extremes and serial dependence.

This thesis focuses on the identification and prediction of crashes using Hawkes processes (Chapter 2 en 4), on testing these Hawkes processes for correct specification (Chapter 3), and on the estimation of Hawkes processes using option prices in a non-affine continuous-time setting (Chapter 5). Hawkes processes, first proposed by Hawkes (1971), match the self-exciting behavior of stock returns around a financial market crash, which is similar to the seismic activity around earthquakes. The jump rate of the Hawkes process increases when a jump (or shock) arrives after which the rate decays as a function of the time passed since the jump. As the probability of jumps increases after a jump has occurred, the Hawkes process is thus called self-exciting. Hence, while events can be triggered by an exogenous factor, for a Hawkes process the risk of events grows endogenously. Characteristics typically observed in data that fit Hawkes models, are the clustering of events and serial dependence.

The Hawkes process was first applied in the so called Epidemic Type After Sequence (ETAS) model, to model the occurrence rate of earthquakes above a certain threshold. This model has been developed by Ogata (1988) and its use for earthquakes is widely investigated by geophysicists.[1] Thereafter the ETAS model has been exploited for crime rates (Mohler et al., 2011) and the spread of red banana plants (Balderama et al., 2012). More interesting is that the ETAS model (in the financial literature often referred to as Hawkes model) is applied to financial data, for example to model arrival data of buy and sell trades (Hewlett, 2006), the duration between trades (Bauwens and Hautsch, 2009) and the returns on one of more indices.[2] This thesis focuses on the latter application of the Hawkes process. The Hawkes modeling framework differs from Extreme Value models as the framework allows for dependencies across arrival times and magnitudes of shocks. At the same time, the framework differs from well known and commonly used volatility models, as it is capable of generating highly insightful forecasts without stringent assumptions on the tail behavior of error distributions. This makes the modelling framework rather easy to implement and understand in practice.

[1] See amongst others: Ogata (1988), Helmstetter and Sornette (2002), Zhuang et al. (2002), Zhuang et al. (2004), Saichev et al. (2005), Hardebeck et al. (2008), and Veen and Schoenberg (2008).

[2] See amongst others: Chavez-Demoulin et al. (2005), Herrera and Schipp (2009), Embrechts et al. (2011), Grothe et al. (2014) and Aït-Sahalia et al. (2015).

Earthquakes exhibit clustering behaviour in space as well as in time. Like earthquake sequences, financial shocks seem to cluster in a dimension other than the time dimension.[3] Extreme stock returns across markets are found to be more correlated than small returns (Bae et al., 2003). They occur more frequently at the same time than expected under the assumption of a normal dependence structure (Mashal and Zeevi, 2002; Hartmann et al., 2004; Sun et al., 2009). This suggests that different financial markets experience stress at the same time. For example, volatility spillover effects between stock markets have been detected in numerous studies.[4] Interpreting volatility as a measure for the tension, these findings indicate that stress from financial markets pours over to other financial markets.

The consequences of this cross-dependence between markets became more apparent during the financial crisis of 2008, also mentioned in the first paragraph of this introduction. This crisis demonstrated the overlap of periods in which financial markets are subject to tension with extreme price movements as a result. For example, on September 29, October 15 and December 1 in 2008 the S&P 500, the Dow Jones Industrial Average (DJI) and the NASDAQ, all suffered top 20 percentage losses. Furthermore, on September 29 the euro/dollar rate and the pound/dollar rate also dropped by a large amount, while the US bond market boomed. On the 16th of October, just one day after the major US stock markets crashed, and on the 1th of December both currencies fell again sharply. Moreover, 4 days after these dates US bond prices shifted significantly upward.

In Chapter 3 and 4, we aim to model the dependence between financial markets. That is, we extend the univariate Hawkes modelling framework to allow extreme events in one financial market to trigger the occurrence and/or the magnitude of extreme events in other markets. In these chapters, we assess whether incorporating cross-sectional dependence improves in- and out-of-sample performance of Hawkes models. This way we confirm that financial shocks exhibit clustering behaviour in the cross section on top of the clustering behaviour in the time dimension.

As option prices reflect expected future stock returns, exploiting the information in option prices can be used to estimate Hawkes models for stock returns more accurately. Even though affine model specifications are far more popular as they provide closed-form derivative prices

[3] See amongst others: Eun and Shim (1989), Fischer and Palasvirta (1990), King and Wadhwani (1990), Lin et al. (1994) and Connolly and Wang (2003).

[4] See amongst others: Hamao et al. (1990), Bae and Karolyi (1994), Koutmos and Booth (1995), Booth et al. (1997) and Kanas (1998).

which facilitates model calibration using option prices, non-affine specifications, seems to fit and predict asset prices considerably better.[5] However, it is very computationally demanding to estimate such models when non-affine dynamics are assumed. In Chapter 5, a framework is developed to estimate non-affine Hawkes models using MCMC and particle methods in a learning setting with latent volatility and jump states as a by-product. Utilizing information from option prices, the compensation investors receive for diffusive and jump risk can be derived using this framework which can not be identified from stock prices alone (Andersen et al., 2015b). Santa-Clara and Yan (2010), Bollerslev and Todorov (2011), Bollerslev et al. (2015), Andersen et al. (2015b) and Boswijk et al. (2015) show that the compensation for the risk of jumps, not attributable to volatility, explains to a large extent the equity and variance risk premia, of which the last one can be seen an indication of the fear of investors. Therefore, disentangling of volatility and jump components in risk premia using option prices provides one with important information regarding the state and development of the financial market with far-reaching implications for asset allocation, hedging, and risk management.

1.2 Outline

Chapter 2 is based on Gresnigt et al. (2015), in which we use the ETAS model as a tool to create probability predictions for an upcoming crash (read: earthquake) in a financial market on the medium term, like sometime in the next five days. A large literature in finance has focused on predicting the risk of downward price movements one-step ahead with measures like Value-at-Risk and Expected Shortfall. Our approach differs as we interpret financial crashes as earthquakes in the financial market, which allows us to develop an Early Warning System (EWS) for crash days within a given period. Testing our EWS on S&P 500 data during the recent financial crisis, we find positive Hanssen-Kuiper Skill Scores. Furthermore, our modeling framework is capable of exploiting information in the returns series not captured by well known and commonly used volatility models. EWS based on our models outperform EWS based on the volatility models forecasting extreme price movements, while forecasting is much less time-consuming.

[5] Chernov et al. (2003), Jones (2003), Christoffersen et al. (2010), Kaeck and Alexander (2012), Durham (2013) and Ignatieva et al. (2015) find non-affine models should be preferred above affine models as they are more flexible and better capable of modeling the tails of the heavy-tailed asset return distribution, while remaining equally parsimonious.

Of course, to accurately identify and predict the occurrence of extreme price movements in financial markets using ETAS models, these models should be properly specified. Hence, specification tests for Hawkes processes are essential. Chapter 3 is based on Gresnigt et al. (2016a), in which we propose various specification tests for Hawkes models based on the Lagrange Multiplier (LM) principle. Our testing focus is on extending a univariate model to a multivariate model, that is, we examine whether there is a conditional dependence between series of extreme events in (different) markets. Thereby we fill the gap in the financial literature which, despite efforts to detect dependence between series (Hartmann et al., 2004; Gonzalo and Olmo, 2005; Hu, 2006), insufficiently describes how to adequately assess the contribution of cross-sectional dependence in a point process framework with serial dependence. LM based specification tests can also be used to test for omitted explanatory variables, breaks in the model parameters, omitted impact of the sizes of events on the triggering of new events and omitted predictability of event sizes. Simulations show that the test has good size and power, in particular for sample sizes that are typically encountered in practice. Moreover, in contrast to de Likelihood Ratio test, the LM test does not require estimation under the alternative hypothesis. As the LM test performs comparable to the LR test and is a lot less time consuming, this test is to be preferred in our opinion. Applying the specification test for dependence to US stocks, bonds and exchange rate data, we find strong evidence for cross-excitation within segments as well as between segments, which cannot simply be explained by volatility spillovers. Therefore, we recommend that univariate Hawkes models be extended to account for the cross-triggering phenomenon.

Nowadays, a large literature focuses to the modeling of extremal dependence between financial markets, though with an in-sample focus.[6] Chapter 4 is based on Gresnigt et al. (2016b), in which we extend these studies on contagion, as we examine whether incorporating this dependence improves forecasts. We follow the recommendation of Chapter 3 Gresnigt et al. (2016a), and utilize Hawkes models in which events are triggered through self-excitation as well as cross-excitation to create our forecasts. The models are applied to US stocks, bonds and dollar exchange rates. We predict the probability of crashes in the series and the Value-at-Risk over a period that includes the financial crisis of 2008 using a moving window. Out-of-sample, we find that the models that include cross-triggering effects

[6] See amongst others: Longin and Solnik (1995), Poon et al. (2003), Poon et al. (2004), Bekaert et al. (2010), Grothe et al. (2014), and Aït-Sahalia et al. (2015).

forecast crashes and the Value-at-Risk significantly more accurately than the models without these effects.

Chapter 5 contains a research proposol. In this Chapter a framework is proposed in which option prices are used to estimate continous-time Hawkes models more accurately. The framework is very general and allows for models to be of the non-affine type, in which asset prices do not have an analytical characteristic function. Using learning methods, models are efficiently estimated and assessed sequentially such that models can be updated quickly when new information arrives. Within the framework, MCMC techniques (Lindsten et al., 2014) and particle filtering methods (Pitt and Shephard, 1999; Johannes et al., 2009) are used to derive the distribution of the model parameters and the latent volatility and jump process. The estimation framework is very flexible and can be made fit to tailor the application at hand. For example the technique can be extended to the multivariate case as it does not require direct optimization of a multidimensional integral which is a problem in several classic estimation frameworks that consider option prices. This makes the estimation technique very attractive for further investigation as Chapter 3 and 4 show jump intensities are mutually exciting. Furthermore, the framework allows information from options to be utilized at a lower frequency than the information of asset prices to estimate models. Including option prices in the estimation of models not only increases accuracy, also it allows one to derive to derive the different compensations investors require for taking on diffusive and jump risk. This provides insight in the state and development of the financial market with important guidance for risk management.

Chapter 2

Interpreting financial market crashes as earthquakes: A new early warning system for medium term crashes

2.1 Introduction

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

[1] This Chapter is based on Gresnigt et al. (2015)

assumptions on the tail behavior of error distributions. This makes our approach rather easy to implement and understand in practice.

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

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

Potential similarities across the behavior of stock returns around crashes and the dynamics of earthquake sequences have been noted in the so-called econophysics literature, in which physics models are applied to economics.[2] In contrast to the studies in the econophysics literature and also to related studies like Bowsher (2007) and Clements et al. (2013), in our framework we do not model the (cumulative) returns but only the extreme returns. As such, we most effectively exploit the information contained in the returns about the crash behavior. As Aït-Sahalia et al. (2015) already show, only taking the jump dynamics of returns

[2]See amongst others: Sornette (2003), Weber et al. (2007), Petersen et al. (2010), Baldovin et al. (2011), Baldovin et al. (2013), Baldovin et al. (2015), and Bormetti et al. (2015)

into account to approximate the timing of crashes gives more accurate results than using the full distribution of the returns. As is well known, the distribution of stock returns is more heavy-tailed than the Gaussian distribution as extreme returns occur more often than can be expected under normality. Furthermore, the distribution of stock returns is usually negatively skewed. As risk in financial markets is predominantly related to extreme price movements, we propose to model only extreme (negative) returns in order to improve predictions.

To model the extreme (negative) returns we use a particular model that is often used for earthquake sequences, and which is the so-called Epidemic-type Aftershock Sequence model (ETAS). This model has been developed by Ogata (1988) and its use for earthquakes is widely investigated by geophysicists.[3] In the ETAS model a Hawkes process, an inhomogeneous Poisson process, is used to model the occurrence rate of earthquakes above a certain threshold. The jump rate of the Hawkes process increases when a jump (or shock) arrives after which the rate decays as a function of the time passed since the jump. As the probability of jumps increases after a jump has occurred, the Hawkes process is thus called self-exciting. The ETAS model has been exploited for crime rates (Mohler et al., 2011) and for the spread of red banana plants (Balderama et al., 2012). Interestingly, the ETAS model has also been applied to financial data, for example to model arrival data of buy and sell trades (Hewlett, 2006), the duration between trades (Bauwens and Hautsch, 2009) or the returns on multiple indices (Embrechts et al., 2011; Grothe et al., 2014; Aït-Sahalia et al., 2015).

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

In somewhat more detail, we consider several specifications of the key triggering functions. The parameters of the ETAS models are estimated by maximum likelihood. And, to

[3] See amongst others: Ogata (1988), Helmstetter and Sornette (2002), Zhuang et al. (2002), Zhuang et al. (2004), Saichev et al. (2005), Hardebeck et al. (2008), and Veen and Schoenberg (2008)

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

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

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

2.2 Models

The Epidemic-Type Aftershock Sequence (ETAS) model is a branching model, in which each event can trigger subsequent events, which in turn can trigger subsequent events of their own. The ETAS model is based on the mutually self-exciting Hawkes point process (Hawkes, 1971), which is an inhomogeneous Poisson process. For the Hawkes process, the intensity at which events arrive at time t depends on the history of events prior to time t.

Consider an event process (t_1, m_1),...,(t_n, m_n) where t_i defines the time and m_i the mark of event i. Let $\mathcal{H}_t = \{(t_i, m_i) : t_i < t_g\}$ represent the entire history of events up to time t.

The conditional intensity of jump arrivals following a Hawkes process is given by

$$\lambda(t|\boldsymbol{\theta};\mathcal{H}_t) = \mu + \sum_{i:t_i<t} g(t-t_i, m_i) \tag{2.1}$$

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

We consider the following specifications of event triggering functions

$$g_{pow}(t-t_i, m_i) = \frac{K_0}{(\gamma(t-t_i)+1)^{1+\omega}} c(m_i) \tag{2.2}$$

$$g_{exp}(t-t_i, m_i) = K_0 e^{-\beta(t-t_i)} c(m_i) \tag{2.3}$$

where K_0 controls the maximum intensity of event triggering. Furthermore in (2.3) K_0 covers the expected number of events directly triggered by an event in (2.3). In (2.2) the expected number of direct descendants is covered by the parameter γ. The influence of the sizes of past events on the intensity with which events are triggered in the future is given by $c(m_i)$.

The possibility of an event triggering a subsequent event decays according to a power law distribution for (2.2), while it decays according an exponential distribution for (2.3). The parameters ω and β determine how fast the possibility of triggering events decays with respectively the time passed since an event. When ω and β are larger, the possibility that an event triggers another event dies out more quickly.

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

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

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

In the literature on Hawkes processes, the exponential function is frequently used to capture the influence of the size of past events on the arrival rate of new events, also when applied to financial data. Using the exponential form, referred to with the subscript 'e', the impact of the magnitude of an event on the triggering intensity becomes

$$c_e(m_i) = e^{\alpha(m_i - M_0)} \tag{2.4}$$

There are theoretical reasons to use this functional form for earthquakes; here also other choices can be made. Therefore we examine two other impact functions. The first function, referred to with a subscript 'p', is the power law function

$$c_p(m_i) = (m_i/M_0)^{\alpha} \tag{2.5}$$

The second function is the impact function preferred by Grothe et al. (2014). They argue that to accurately extract information from the magnitudes of events, the quantile of magnitude of the event in the conditional distribution from which it is drawn should be considered. Their function, referred to with a subscript 'd', has the following form

$$c_d(m_i) = 1 - \alpha \log\left(1 - G_{\xi,\sigma(t)}(m_i)\right) \tag{2.6}$$

where $G_{\xi,\sigma(t)}$ is the Generalized Pareto Distribution of the sizes of excess magnitude events. Using this impact function, the probability of an event i having a magnitude between M_0 and m_i, determines i's influence on the triggering intensity. This influence depends on the history of the event process, whenever the scale parameter of the GPD distribution of the sizes of the excess magnitude events is not constant ($\eta \neq 0$).

Table 2.1: Specification ETAS models

Triggering function	Power law				Exponential			
Model	A		B		C		D	
Influence event history	$\eta = 0$		$\eta \neq 0$		$\eta = 0$		$\eta \neq 0$	
Influence magnitude events	n	e p d	n	e p d	n	e p d	n	e p d

In the models indicated by the subscript 'n', the influence of the magnitude of events on the triggering subsequent events is restricted to zero. In the models referred to with the subscripts 'e', 'p' and 'd', the impact of the sizes of events on the triggering of subsequent events is given by the impact functions (2.4), (2.5) and (2.6) respectively. The influence of the event history on the magnitude of events is zero when η is restricted to 0.

When $\alpha \neq 0$ the intensity at which subsequent events are triggered by a past event is influenced by the size of this past event. The minimum magnitude of an event is represented by M_0. How the size of an event affects the probability of triggering other events is determined by α. Assuming that larger events trigger more events than smaller events, so that $\alpha > 0$, the probability of triggering events increases with the size past events (m_i). The larger α, the more pronounced is the influence of the size of events. When $\eta > 0$ the magnitude of events is expected to be more extreme when the tension in the financial market is high. Using (2.6) when $\eta > 0$ the impact of extreme events in turbulent periods is therefore smaller than in tranquil periods, when the probability of having these events is lower.

We proceed to investigate several specifications of the ETAS model. We consider both the power law triggering function (2.2) and the exponential triggering function (2.3) in combination with different functions for the impact of the magnitude of previous events on the triggering of events in the future as given in (2.4), (2.5) and (2.6). Furthermore in some models the history of event process can affect the magnitude of events in the future while in other models there is no such influence. In Table 2.1 we present the configurations of the different models.

The process is stationary when the expected number of off springs of an event, that is the branching ratio n, is smaller than 1. When $n \geq 1$ the number of events arriving will grow to infinity over time. The condition for stationarity of the Hawkes process with triggering

function (2.2) and (2.3) can be stated as respectively

$$\int_0^\infty g_{pow}(t-t_i,m_i)dt = \frac{K_0}{\gamma\omega} < 1 \tag{2.7}$$

$$\int_0^\infty g_{exp}(t-t_i,m_i)dt = \frac{K_0}{\beta} < 1 \tag{2.8}$$

While Bacry et al. (2012) use a non-parametric kernel estimation technique for a symmetric Hawkes process on high frequency data, we prefer parametric kernel estimation to make the model more interpretable. We can advocate this technique as the literature is not consistent in which triggering function to use for financial data. A well known stylized fact of the absolute returns is that they decay roughly according to a power law (Cont, 2001). Selçuk and Gençay (2006), Weber et al. (2007) and Petersen et al. (2010) conclude that the intraday volatility of stock returns above a certain threshold decays roughly according a power-law, approximating the intraday volatility by the absolute returns. However while for example Hardiman et al. (2013) find power law functions fit the S&P 500 data, Filimonov and Sornette (2015) among others report the superior performance of exponential functions. We consider both functions.

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

$$\begin{aligned}\log L(\boldsymbol{\theta}) = &\sum_{i=1}^{N} \log \lambda(t_i|\boldsymbol{\theta};\mathcal{H}_t) - \log\sigma(t) + \left(1+\frac{1}{\xi}\right)\log\left(1+\xi\frac{m_i-M_0}{\sigma(t)}\right) \\ &- \int_0^T \lambda(t_i|\boldsymbol{\theta};\mathcal{H}_t)dt\end{aligned} \tag{2.9}$$

where $\lambda(t_i|\boldsymbol{\theta};\mathcal{H}_t)$ is the conditional intensity and t_i are the event arrival times in the interval $[0,T]$. We optimize the log-likelihood numerically using the Nelder-Mead simplex direct search algorithm. The difficulty of accurately estimating the parameters of a Hawkes process has been well recognized in the literature on Hawkes processes.[4] After exploiting several estimation methods and optimization algorithms and testing our procedure on simulated data series, we found this approach most satisfactory. To check whether the obtained optima are

[4]See amongst others: Veen and Schoenberg (2008), Chavez-Demoulin and McGill (2012), Hardiman et al. (2013), Rasmussen (2013), Aït-Sahalia et al. (2015), Filimonov and Sornette (2015) and Bacry et al. (2012)

not of a local nature, we estimate the models using different starting values. Furthermore we use models to generate data and estimate the parameters of the models using this data.

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

$$\begin{aligned} \Pr\left(N(t_n) - N(t_{n-1}) > 0\right) &= 1 - \Pr\left(N(t_n) - N(t_{n-1}) = 0\right) \\ &= 1 - F\left(t^* > t_n - t_{n-1}\right) \\ &= 1 - \exp\left(-\int_{t_{n-1}}^{t_n} \lambda(t|\boldsymbol{\theta}; \mathcal{H}_t) dt\right) \end{aligned} \tag{2.10}$$

Thus, using the conditional intensity (2.1) specified by the estimated parameters of the ETAS models and the history of the stock returns, we are able to predict the probability of the occurrence of an event during a given time period. These probability predictions form the basis of our Early Warning system.

2.3 Application to Financial Data

We consider data of the S&P 500 index over a period from 2 January, 1957, to 1 September, 2008 to calibrate our models and 5 years thereafter for an out-of-sample evaluation of the models. The dataset consists of daily returns $R_t = \frac{p_t - p_{t-1}}{p_{t-1}} \times 100$, where p_t denotes the value of the index at t. Figure 2.1 shows the evolution of the S&P 500 index and also the returns on this index. Severe drops in the price index and large negative returns corresponding to these drops, are observed around famous crash periods, "Black Monday" (1987) and the stock market downturn of 2002 after the "dot-com bubble" (1997–2000). Furthermore the Figure illustrates the clustering of extreme returns, that is tranquil periods with small price changes alternate with turbulent periods with large price changes. This clustering feature can be related to the positive herding behavior of investors and the endogenous growth of instability in the financial market.

We apply the ETAS models to the 95% quantile of extreme returns and the 95% quantile of extreme negative returns referred to as extremes and crashes, respectively. The minimum magnitude M_0 of the events under consideration corresponding to the 95% quantile of extreme (negative) returns is calculated over the estimation period, that is over a period from 2 January, 1957, to 1 September, 2008. Each quantile includes 687 events from the $13,738$

trading days. The estimation of various model parameters are presented in Table 2.2 and Table 2.3.

To give an interpretation to the parameter μ consider the following. Returns above the 95% threshold not triggered by previous extremes occur on average at a daily rate that ranges from 0.0059 (model A_n) to 0.0082 (model D_p). Over the considered time period approximately 81–113 of the total of 687 events arrived spontaneously according to the models. This means that about 84–88% of the events were triggered by prior events. For the crashes, the mean background intensity of events ranges from 0.0077 (model A_n) to 0.0119 (model D_p), so that about 76–85% of the events are triggered by other events according to the models. Also the branching ratio (n), that is the expected number of direct descendants of an event, lies in the interval $[0.86, 0.89]$, $[0.79, 0.86]$, for extremes and crashes respectively, in the models where the magnitude of an event has no influence on the triggering of descendants ($\alpha = 0$). In the models where α is not restricted to zero, the branching ratio differs across events as it depends on the magnitude of events. However as $\alpha > 0$ and other parameter estimates are similar, the expected number of descendants of an event tends to be at least as high in these models as in the models with $\alpha = 0$.

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

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

The estimates for η in the models B and D, are positive and significant for both sets of returns. The models score better in both log-likelihood and AIC value than the models A and C. This suggests a model which incorporates the history of the event process to prospect the sizes of subsequent events, matches the extreme (negative) returns closer than a model which assumes the sizes of events are independent of the past. When $\eta > 0$, the mean and

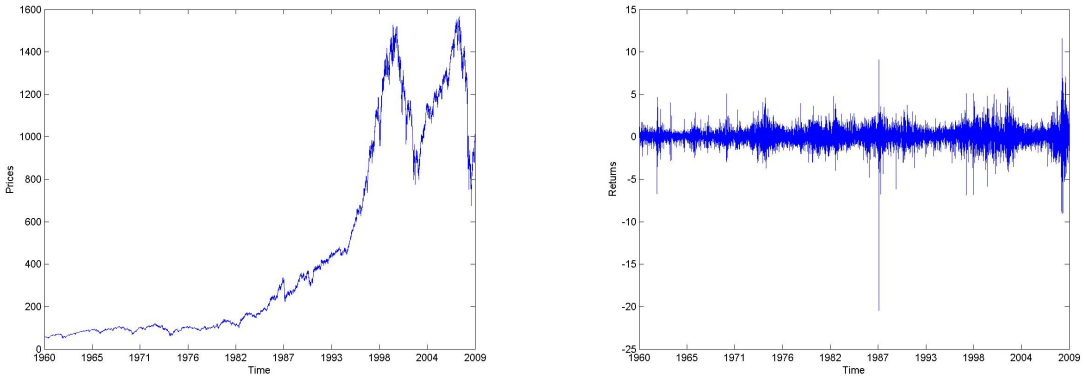
variance of the distribution of the excess magnitudes of the events scale with the value of the cumulative triggering function, and thus the probability of the arrival of an event triggered by another event. This means that on average larger extremes are observed after the occurrence of more and/or big events than after a tranquil period.

Comparing the ETAS models in which the intensity does not depend on the sizes of prior events, i.e. with the parameter restriction $\alpha = 0$, to the ETAS models without this restriction, the magnitude of an extreme has a significant positive influence on the probability of triggering another extreme for both sets of returns. This means that on average larger events trigger more events than smaller events. The models A, B, C and D with either the subscript 'e', 'p' or 'd', have a higher ranking both in terms of log-likelihood as in AIC value than their counterparts with $\alpha = 0$, that is model A_n, B_n, C_n and D_n respectively. Incorporating the size of the events into an ETAS model for the extreme (negative) returns thus improves the model. Amongst the models with $\alpha \neq 0$, the models with the exponential function (2.4) perform the worst for the extreme returns as well as for the extreme negative returns. Therefore, we can indeed conclude that there are no solid reasons to use this function to describe the influence of the magnitude of events on the triggering intensity. For the crashes, and the extremes whenever η is restricted to zero, the power law function (2.5) is preferred over the other two impact specifications. For the extremes the impact function of Grothe et al. (2014) (2.6) performs best when $\eta \neq 0$. In this model the impact of the sizes of events is smaller in turbulent periods than in tranquil periods.

A likelihood ratio test shows that all the estimated parameters of the models are significant at a 5% level. All together model B with a power law triggering function, and non-zero influence of the size of the events on the triggering of subsequent events and predictable event sizes, fits best according to the log-likelihoods and AIC values for both the extremes and crashes. However for the extremes the impact function as specified in (2.6) is preferred, while for the crashes the impact function (2.5) gives slightly better results.

Figure 2.2 presents the intensity with which extremes and crashes occur estimated with respectively model B_d and B_p, over the estimation period, that is from 2 January, 1957, to 1 September, 2008. The estimated intensity shows large spikes around the famous crash periods, "Black Monday" (1987) and the "stock market downturn of 2002" (2002) after the "dot-com bubble" (1997-2000). As expected, the rate at which events arrive is high around crashes, reflecting the increase in the triggering probability after the occurrence of events.

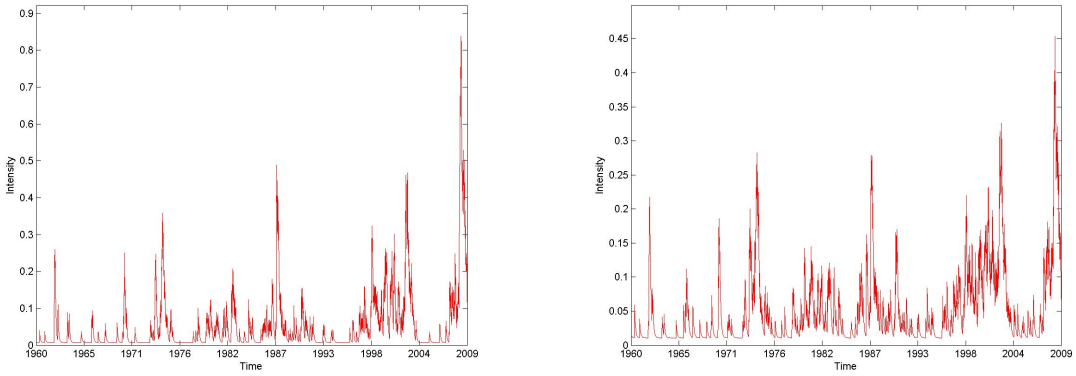
Figure 2.1: S&P 500 index



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

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

Figure 2.2: Conditional intensity



(a) Extremes model B_d **(b)** Crashes model B_p

Estimated conditional intensity for the 95% quantile of daily extreme returns, extreme negative returns, over the period January 2, 1957, until September 1, 2008, using respectively model B_d, B_p.

Table 2.2: Estimation results extremes

	A_n	B_n	C_n	D_n	A_e	B_e	C_e	D_e	A_p	B_p	C_p	D_p	A_d	B_d	C_d	D_d
μ	0.0059	0.0062	0.0068	0.0074	0.0061	0.0064	0.0071	0.0078	0.0063	0.0067	0.0073	0.0082	0.0063	0.0060	0.0073	0.0074
	(0.0012)	(0.0012)	(0.0011)	(0.0011)	(0.0012)	(0.0013)	(0.0011)	(0.0012)	(0.0012)	(0.0013)	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)
K_0	0.0441	0.0522	0.0382	0.0442	0.0381	0.0446	0.0325	0.0369	0.0341	0.0382	0.0295	0.0320	0.0325	0.0222	0.0284	0.0193
	(0.0057)	(0.0064)	(0.0041)	(0.0045)	(0.0052)	(0.0058)	(0.0039)	(0.0041)	(0.0051)	(0.0054)	(0.0040)	(0.0040)	(0.0056)	(0.0060)	(0.0044)	(0.0048)
γ	0.0180	0.0262			0.0209	0.0322			0.0221	0.0353			0.0221	0.0284		
	(0.0102)	(0.0127)			(0.0110)	(0.0142)			(0.0114)	(0.0152)			(0.0114)	(0.0129)		
ω	2.7584	2.2532			2.3756	1.8570			2.2792	1.7422			2.2941	1.9115		
	(1.3899)	(0.9531)			(1.0938)	(0.7080)			(1.0252)	(0.6458)			(1.0333)	(0.7512)		
β			0.0440	0.0517			0.0434	0.0511			0.0440	0.0524			0.0441	0.0467
			(0.0047)	(0.0053)			(0.0047)	(0.0053)			(0.0048)	(0.0054)			(0.0048)	(0.0050)
α					0.1293	0.1435	0.1241	0.1390	0.7017	0.8691	0.6497	0.8133	0.3736	1.1660	0.3305	1.0694
					(0.0248)	(0.0181)	(0.0256)	(0.0185)	(0.1686)	(0.1375)	(0.1689)	(0.1382)	(0.1524)	(0.4656)	(0.1390)	(0.4163)
ξ	0.2602	0.1325	0.2602	0.1361	0.2602	0.1156	0.2602	0.1200	0.2602	0.1125	0.2602	0.1180	0.2591	0.1011	0.2594	0.1071
	(0.0470)	(0.0383)	(0.0470)	(0.0385)	(0.0470)	(0.0366)	(0.0470)	(0.0369)	(0.0470)	(0.0367)	(0.0470)	(0.0370)	(0.0468)	(0.0359)	(0.0469)	(0.0363)
ϕ	0.6956	0.3654	0.6957	0.3758	0.6956	0.3874	0.6956	0.3974	0.6956	0.3956	0.6956	0.4039	0.6964	0.3836	0.6961	0.3887
	(0.0415)	(0.0362)	(0.0415)	(0.0360)	(0.0415)	(0.0364)	(0.0415)	(0.0361)	(0.0415)	(0.0365)	(0.0415)	(0.0362)	(0.0415)	(0.0367)	(0.0415)	(0.0363)
η		0.1480		0.1215		0.1143		0.0917		0.0943		0.0771		0.0592		0.0505
		(0.0248)		(0.0182)		(0.0205)		(0.0150)		(0.0185)		(0.0141)		(0.0178)		(0.0141)
$\log L(\boldsymbol{\theta})$	-2860.80	-2806.85	-2863.19	-2810.35	-2854.57	-2793.17	-2857.63	-2797.98	-2854.17	-2792.21	-2857.42	-2797.41	-2854.28	-2785.86	-2857.50	-2790.39
AIC	5733.61	5625.69	5740.39	5632.71	5723.15	5600.34	5731.26	5609.96	5722.33	5600.41	5726.84	5608.83	5722.57	5587.73	5727.01	5594.77

The models are applied to the 95% quantile of extreme returns on the S&P 500 index over the period January 2, 1957, until September 1, 2008. Model A and B correspond to an ETAS model with a power law triggering function. Model C and D correspond to an ETAS model with an exponential triggering function. In model A and C the history of the events has no influence on the magnitude of subsequent events, that is the parameter restriction $\eta = 0$ is imposed. In the models indicated by the subscript 'n', the influence of the magnitude of events on the triggering subsequent events is restricted to zero. In the models referred to with the subscripts 'e', 'p' and 'd', the impact of the sizes of events on the triggering of subsequent events is given by the impact functions (2.4), (2.5) and (2.6) respectively. Standard deviations are shown in between parentheses.

Table 2.3: Estimation results crashes

	A_n	B_n	C_n	D_n	A_e	B_e	C_e	D_e	A_p	B_p	C_p	D_p	A_d	B_d	C_d	D_d
μ	0.0077	0.0084	0.0102	0.0105	0.0080	0.0087	0.0106	0.0112	0.0082	0.0093	0.0108	0.0119	0.0082	0.0077	0.0108	0.0107
	(0.0020)	(0.0019)	(0.0015)	(0.0015)	(0.0020)	(0.0020)	(0.0016)	(0.0015)	(0.0020)	(0.0020)	(0.0016)	(0.0016)	(0.0020)	(0.0020)	(0.0016)	(0.0015)
K_0	0.0337	0.0358	0.0280	0.0298	0.0309	0.0325	0.0258	0.0265	0.0289	0.0281	0.0242	0.0231	0.0284	0.0211	0.0240	0.0181
	(0.0052)	(0.0051)	(0.0036)	(0.0034)	(0.0049)	(0.0048)	(0.0035)	(0.0032)	(0.0050)	(0.0044)	(0.0037)	(0.0031)	(0.0052)	(0.0049)	(0.0039)	(0.0038)
γ	0.0257	0.0244			0.0270	0.0300			0.0282	0.0344			0.0284	0.0313		
	(0.0142)	(0.0141)			(0.0146)	(0.0160)			(0.0150)	(0.0173)			(0.0150)	(0.0155)		
ω	1.5290	1.7428			1.4670	1.4645			1.4256	1.3355			1.4204	1.2934		
	(0.7537)	(0.8974)			(0.7061)	(0.6949)			(0.6752)	(0.5985)			(0.6686)	(0.5748)		
β			0.0351	0.0377			0.0354	0.0382			0.0359	0.0398			0.0360	0.0357
			(0.0047)	(0.0044)			(0.0047)	(0.0044)			(0.0048)	(0.0047)			(0.0048)	(0.0044)
α					0.0983	0.1302	0.0939	0.1249	0.4322	0.7351	0.3943	0.6828	0.2035	0.6499	0.1766	0.5544
					(0.0341)	(0.0205)	(0.0348)	(0.0211)	(0.1840)	(0.1324)	(0.1837)	(0.1329)	(0.1209)	(0.2671)	(0.1114)	(0.2284)
ξ	0.2885	0.1842	0.2885	0.1858	0.2885	0.1641	0.2885	0.1679	0.2885	0.1559	0.2885	0.1613	0.2884	0.1582	0.2881	0.1634
	(0.0479)	(0.0414)	(0.0479)	(0.0412)	(0.0479)	(0.0402)	(0.0479)	(0.0401)	(0.0479)	(0.0400)	(0.0479)	(0.0398)	(0.0479)	(0.0400)	(0.0479)	(0.0399)
ϕ	0.5550	0.2374	0.5550	0.2463	0.5550	0.2483	0.5550	0.2585	0.5550	0.2555	0.5550	0.2656	0.5550	0.2487	0.5552	0.2596
	(0.0334)	(0.0289)	(0.0334)	(0.0283)	(0.0334)	(0.0293)	(0.0334)	(0.0286)	(0.0334)	(0.0295)	(0.0334)	(0.0288)	(0.0334)	(0.0308)	(0.0334)	(0.0297)
η		0.1459		0.1216		0.1289		0.1044		0.1095		0.0894		0.0830		0.0712
		(0.0245)		(0.0171)		(0.0224)		(0.0154)		(0.0202)		(0.0145)		(0.0211)		(0.0164)
$\log L(\boldsymbol{\theta})$	-2928.44	-2867.73	-2931.76	-2870.74	-2926.29	-2858.83	-2929.83	-2862.72	-2926.14	-2856.21	-2929.81	-2860.64	-2926.15	-2856.42	-2929.85	-2860.87
AIC	5868.88	5747.46	5877.53	5753.49	5866.59	5731.66	5875.65	5739.44	5866.27	5728.41	5871.62	5735.29	5866.31	5728.85	5871.69	5735.73

The models are applied to the 95% quantile of extreme negative returns on the S&P 500 index over the period January 2, 1957, until September 1, 2008. Model A and B correspond to an ETAS model with a power law triggering function. Model C and D correspond to an ETAS model with an exponential triggering function. In model A and C the history of the events has no influence on the magnitude of subsequent events, that is the parameter restriction $\eta = 0$ is imposed. In the models indicated by the subscript 'n', the influence of the magnitude of events on the triggering subsequent events is restricted to zero. In the models referred to with the subscripts 'e', 'p' and 'd', the impact of the sizes of events on the triggering of subsequent events is given by the impact functions (2.4), (2.5) and (2.6) respectively. Standard deviations are shown in between parentheses.

2.4 Goodness-of-fit

2.4.1 Simulation

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

We generate 1000 data series from the models using the parameters estimates derived from the extreme negative returns on the S&P 500 index (Table 2.3). We set the sample period equal to the number of trading days over which we estimated the models for the S&P 500 crashes. Estimation results for these series are shown in Table 2.4. One thing that stands out is the estimation results of the ETAS models with a power law triggering function (models A and B) are not so satisfactory. The maximum likelihood estimation does not converge in a number of simulations. Furthermore the estimated $\widehat{\omega}$ of the triggering functions deviate much from the ω used to simulate the data and the standard deviations of the $\widehat{\omega}$ are much larger than the standard deviation of $\widehat{\omega}$ derived from the crashes. The estimates for ω derived from data series generated with a continuous time procedure are much closer to values used to simulate the series. Also the standard deviations of these $\widehat{\omega}$ are much smaller.

We have examined several methods to simulate and estimate the ETAS model with the power law triggering function. When estimating the models, the Expectation-Maximization procedure of Veen and Schoenberg (2008), the Bayesian procedure of Rasmussen (2013) and gradient-based optimization algorithms give inferior results in terms of speed and robustness for our kind of data. The estimated ETAS models with the exponential triggering function (models C and D) appear more reliable.

In Figure 2.3 the S&P 500 crashes are compared to a series simulated with the discrete time procedure from model B_p (power law triggering function) and D_p (exponential trigger-

ing function). In these models the influence of the magnitude on the triggering of subsequent events and the influence of the history of the event process on the sizes of subsequent events, are both non-zero. For crashes, Model B_p has the highest log-likelihood and lowest AIC value amongst the models. The simulated series share the major features characteristic to the models and similar to the crashes like the clustering of events, heavy-tailed distributed event sizes, and large events are especially observed after the occurrence of more and/or other big events.

When looking at the figures the S&P 500 crashes are more similar to the events simulated from model B_p. Histograms show that the data simulated with model D_p differ from the S&P 500 data because many fewer event pairs are observed with a shorter inter event time. Examining graphs of the logarithm of the cumulative number of events against the logarithm of time, the events from model D_p seem to deviate more from the S&P 500 crashes than the events from model B_p. Also the clustering feature in the magnitude-time plots, being more pronounced for model B_p than for model D_p, indicates model B_p should be preferred above model D_p to match the S&P data.

2.4.2 Residual analysis

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

$$\tau_i = \int_0^{t_i} \lambda(t)dt \tag{2.11}$$

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

$$\tau_i - \tau_{i-1} = \int_{t_{i-1}}^{t_i} \lambda(t)dt \tag{2.12}$$

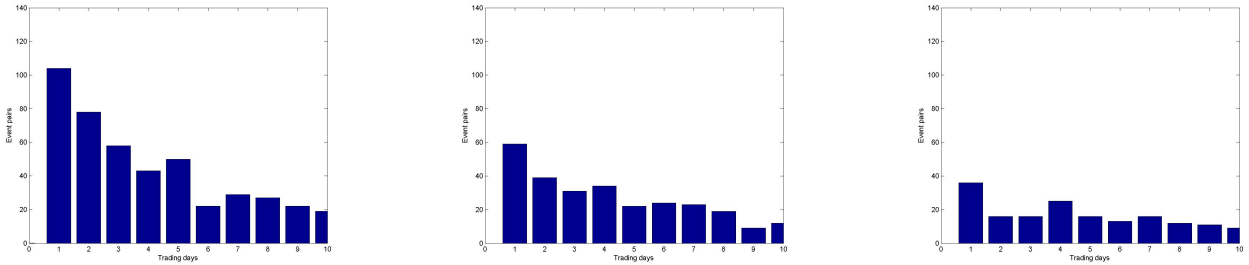
are independent exponential random variables with mean 1. If the models are correctly specified, $\lambda(t)$ can be approximated by $\lambda(t|\widehat{\boldsymbol{\theta}}; \mathcal{H}_t)$. The sequence $\{\tau_i\}$ is called the residual process. In order to verify whether the residual process derived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

Table 2.4: Estimation results simulated data

	A_n	B_n	C_n	D_n	A_e	B_e	C_e	D_e	A_p	B_p	C_p	D_p	A_d	B_d	C_d	D_d
μ	0.0071	0.0073	0.0105	0.0108	0.0075	0.0076	0.0109	0.0113	0.0073	0.0075	0.0109	0.0120	0.0076	0.0079	0.0112	0.0111
	(0.0017)	(0.0016)	(0.0016)	(0.0016)	(0.0018)	(0.0016)	(0.0016)	(0.0016)	(0.0017)	(0.0015)	(0.0016)	(0.0016)	(0.0018)	(0.0019)	(0.0017)	(0.0016)
K_0	0.0205	0.0218	0.0237	0.0256	0.0190	0.0204	0.0216	0.0231	0.0174	0.0174	0.0206	0.0204	0.0165	0.0136	0.0200	0.0151
	(0.0036)	(0.0035)	(0.0030)	(0.0030)	(0.0036)	(0.0038)	(0.0032)	(0.0033)	(0.0040)	(0.0042)	(0.0037)	(0.0037)	(0.0041)	(0.0040)	(0.0040)	(0.0037)
γ	0.0107	0.0110			0.0115	0.0134			0.0125	0.0167			0.0117	0.0153		
	(0.0091)	(0.0089)			(0.0095)	(0.0106)			(0.0110)	(0.0153)			(0.0097)	(0.0108)		
ω	3.3283	3.2184			3.3227	2.8662			3.2587	2.4611			2.9422	2.5049		
	(6.9441)	(5.3769)			(6.0221)	(4.5929)			(8.4964)	(5.3455)			(5.5820)	(3.4493)		
β			0.0319	0.0345			0.0319	0.0346			0.0326	0.0367			0.0324	0.0325
			(0.0041)	(0.0040)			(0.0042)	(0.0044)			(0.0045)	(0.0056)			(0.0042)	(0.0040)
α					0.0827	0.0692	0.0927	0.0829	0.3714	0.4740	0.4284	0.5623	0.2191	0.7110	0.2104	0.6606
					(0.0596)	(0.1017)	(0.0686)	(0.1126)	(0.3535)	(0.5519)	(0.4216)	(0.6243)	(0.1782)	(0.4325)	(0.1861)	(0.3714)
ξ	0.2316	0.1417	0.2853	0.1793	0.2354	0.1252	0.2855	0.1577	0.2262	0.1147	0.2848	0.1541	0.2286	0.1351	0.2864	0.1565
	(0.0450)	(0.0401)	(0.0545)	(0.0500)	(0.0455)	(0.0422)	(0.0543)	(0.0509)	(0.0479)	(0.0468)	(0.0577)	(0.0566)	(0.0440)	(0.0436)	(0.0540)	(0.0485)
ϕ	0.4560	0.1885	0.5586	0.2467	0.4646	0.2002	0.5582	0.2565	0.4474	0.1979	0.5578	0.2667	0.4535	0.2167	0.5569	0.2570
	(0.0311)	(0.0253)	(0.0379)	(0.0303)	(0.0315)	(0.0259)	(0.0377)	(0.0304)	(0.0334)	(0.0259)	(0.0400)	(0.0314)	(0.0306)	(0.0295)	(0.0374)	(0.0314)
η		0.0982		0.1138		0.0892		0.1002		0.0735		0.0854		0.0599		0.0660
		(0.0205)		(0.0195)		(0.0222)		(0.0206)		(0.0251)		(0.0236)		(0.0197)		(0.0187)

The discrete time procedure is used to generate 1000 data series from each model. Model A and B correspond to an ETAS model with a power law triggering function. Model C and D correspond to an ETAS model with an exponential triggering function. In model A and C the history of the events has no influence on the magnitude of subsequent events, that is the parameter restriction $\eta = 0$ is imposed. In the models indicated by the subscript 'n', the influence of the magnitude of events on the triggering subsequent events is restricted to zero. In the models referred to with the subscripts 'e', 'p' and 'd', the impact of the sizes of events on the triggering of subsequent events is given by the impact functions (2.4), (2.5) and (2.6) respectively. Standard deviations are shown in between parentheses.

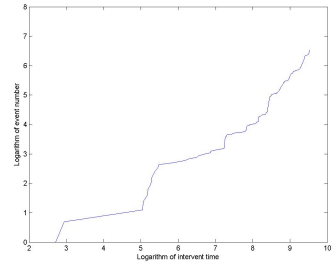
Figure 2.3: S&P 500 crashes and series simulated in discrete time



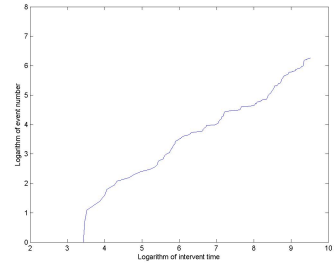
(a) S&P 500 crashes

(b) Model B_p

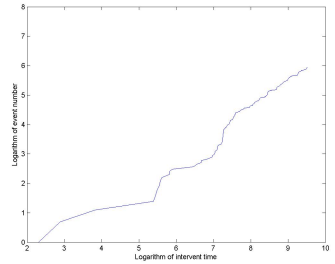
(c) Model D_p



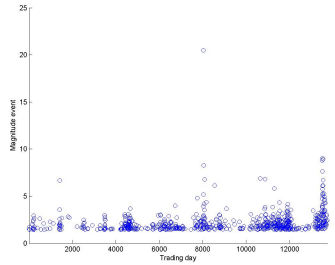
(d) S&P 500 crashes



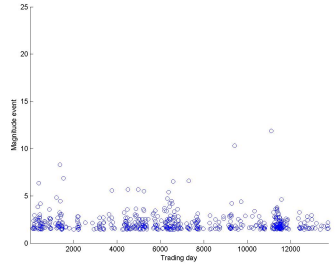
(e) Model B_p



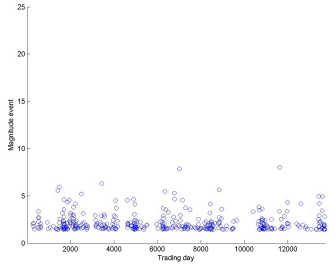
(f) Model D_p



(g) S&P 500 crashes



(h) Model B_p



(i) Model D_p

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

Table 2.5: Kolmogorov-Smirnov tests

		A_n	B_n	C_n	D_n	A_e	B_e	C_e	D_e	A_p	B_p	C_p	D_p	A_d	B_d	C_d	D_d
Crash	95 %	0.152	0.064	0.052	0.204	0.256	0.107	0.086	0.323	0.604	0.198	0.227	0.570	0.333	0.206	0.104	0.600
	97 %	0.122	0.071	0.056	0.134	0.114	0.079	0.050	0.148	0.292	0.134	0.127	0.252	0.083	0.147	0.036	0.284
	99 %	0.096	0.045	0.036	0.101	0.051	0.024	0.019	0.058	0.092	0.041	0.033	0.097	0.055	0.065	0.020	0.152
Extreme	95 %	0.200	0.154	0.104	0.235	0.343	0.274	0.177	0.410	0.621	0.551	0.432	0.748	0.174	0.539	0.085	0.732
	97 %	0.185	0.110	0.076	0.222	0.178	0.119	0.075	0.234	0.493	0.268	0.211	0.522	0.064	0.313	0.022	0.607
	99 %	0.142	0.077	0.060	0.159	0.036	0.022	0.014	0.046	0.053	0.031	0.020	0.066	0.018	0.039	0.006	0.088

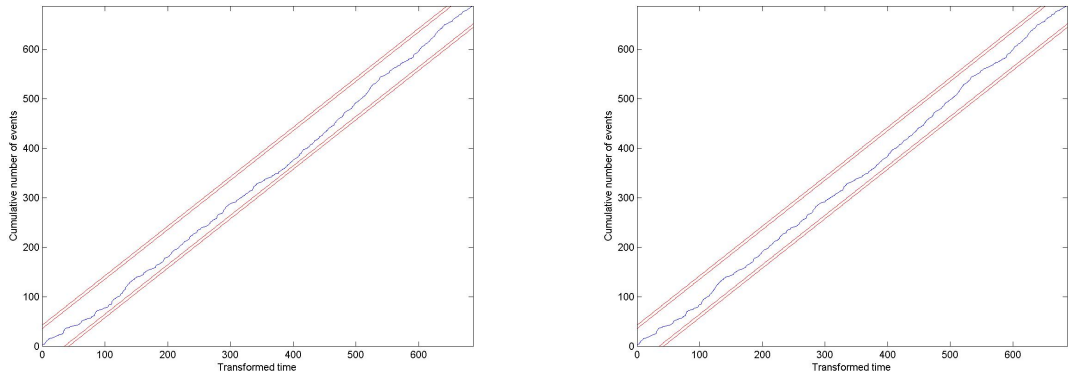
The tests are performed on the transformed times $\{\tau_i\}$ specified by the models. The models are applied to the 95% to 99% quantile of the extreme (negative) returns on the S&P 500 index over the period January 2, 1957, until September 1, 2008. The null hypothesis of the test is transformed times $\{\tau_i\}$ are distributed according to a homogeneous Poisson process with intensity 1. In the Table the p-values of the Kolmogorov-Smirnov tests for the 95%, 97% and 99% quantile are reported.

our test is that the distribution of the residual process and the unit Poisson distribution are equal.

The KS tests are performed on the transformed times derived by applying the ETAS models to the 95% to 99% quantile of the extreme (negative) returns. The p-values of the tests for the 95%, 97% and 99% quantile are reported in Table 2.5. Figure 2.4 shows the cumulative number of S&P 500 crashes for the 95% quantile against the transformed times derived from models B_p and D_d. The 95% and 99% error bounds of the KS statistic are also displayed in the Figure. The first model fits the data best according to the log-likelihood and AIC scores, while the second model seems most appropriate when looking at the results of the residual analysis.

The p-values and the Figure indicate that for all models extreme (negative) returns above the 95% quantile do not deviate from an event process specified by the model at a 5% level. At a 5% level the extreme (negative) returns above the 99% quantile are not correctly specified by many models. Furthermore, model C, the model with the exponential triggering function and unpredictable event sizes, gives low(er) p-values, such that it seems less appropriate to model both extremes and crashes than the other models, especially in combination with the impact function specified by Grothe et al. (2014) (2.6). The models without influence of the magnitude of events on the triggering intensity and the models with an exponential impact function have lower p-values than their counterparts with a power law impact function for all sets of returns. Overall, model D_d, the model with the exponential triggering function and predictable event sizes, in combination with (2.6) for the influence of the sizes of events on the triggering intensity, seems to fit the data best.

Figure 2.4: Residual analysis for the S&P 500 crashes 95% quantile



(a) Model B_p **(b)** Model D_d

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

2.5 Forecasting

2.5.1 Early Warning System

The identification of financial market crashes is of great importance to traders, regulators of financial markets and risk management. They can benefit from an Early Warning System that sets an alarm when the probability of a crash becomes too high, urging the traders, regulators and risk managers to take action. We develop an Early Warning System for extremes and crashes in the financial market within a certain time period using the conditional intensity specified by the estimated parameters of the ETAS models and the history of the stock returns. The probability of an extreme or a crash occurring between t_{n-1} and t_n is given by (4.7). The minimum magnitude M_0 of the events under consideration corresponding to the 95–99% quantile of extreme (negative) returns is calculated over the estimation period, that is over a period from 2 January, 1957, to 1 September, 2008. As we do not calculate the threshold value for events over the out-of-sample period, there is no look-ahead bias.

To evaluate the performance of the EWS, we use measures reported in Candelon et al. (2012). We do not compute the optimal threshold value for giving an alarm. Instead we set the threshold at 0.5. Therefore an alarm is given when the models predict that it is more likely that at least one event occurs than that no event occurs within a certain time period. Here we consider the occurrence of events within a time period of 5 days during the last few

years, that is from 1 September, 2008, to 1 January, 2013, and during the recent financial crisis, that is from 1 September, 2008, to 1 January, 2010.

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

$$QPS = \frac{2}{T}\sum_{t=1}^{T}(\hat{p}_t - y_t)^2 \tag{2.13}$$

$$LPS = -\frac{1}{T}\sum_{t=1}^{T}\left[(1-y_t)\log(1-\hat{p}_t) + y_t \log(\hat{p}_t)\right] \tag{2.14}$$

where t is a day and T is the total number of days. Here p_t represents the predicted probability of crash in the 5 day period commencing day t and y_t is an indicator function taking the value one when a crash occurs within the 5 day period and the value zero otherwise, i.e y_t will be 1 on five days. The QPS and LPS range respectively from 0 to 1 and from 0 to ∞, with 0 indicating perfect accuracy.

When the QPS or the LPS are higher, the probability predictions deviate more from a binary variable indicating the occurrence of events. The LPS punishes large deviations heavier than small deviations. The QPS and the LPS for (negative) extremes above the 95%, 97% and 99% in sample quantile, are displayed in Table 2.6. Overall, the QPS and LPS of the models with an event triggering probability dependent on the size of previous events ($\alpha \neq 0$) are slightly lower. For almost all sets of returns the QPS and LPS of the model with an exponential triggering function and unpredictable event sizes in combination with the impact function of Grothe et al. (2014) (2.6) (model C_d), is the lowest. Comparing different versions of models with the same function for the impact of sizes of past events on the intensity (e.g. A_e, B_e, C_e and D_e), overall the model with an exponential triggering function and predictable event sizes (model D), has a higher QPS and LPS than the other models (A, B and C). Thus, while model C_d seems to be the worst choice for making probability predictions when taking into account the results of the residual analysis, this model actually delivers the most accurate probability predictions. In contrast, for model D_d it is the other way around. The probability predictions for crashes seem somewhat less accurate than the probability predictions for the extremes as the QPS and LPS of the models for crashes are higher than the QPS and LPS of the models for extremes.

Furthermore we check whether the probability predictions for the extreme (negative) returns above the 95% in sample quantile made by some models, are significantly more accurate than those made by other models. For nested models, we compute the adjusted Mean Squared Prediction Error (adjusted MSPE) of Clark and West (2007) (Table 2.7). Nested models are models with the same triggering function and the same function for the impact of sizes of past events on the intensity (e.g. A_e versus B_e), and models with the same triggering function with and without influence of the sizes of events on the triggering intensity (e.g. A_n versus A_e or B_e). According to the adjusted MSPEs, the probability predictions based on models with a size-dependent triggering probability ($\alpha \neq 0$) are significantly more accurate for extremes. For crashes this is also the case when considering the models with an exponential triggering function and unpredictable events sizes (model C), over the period from September 1, 2008, to January 1, 2013.

For the non-nested models, that is the models with a different triggering function or impact function (e.g. A_n versus C_n, A_e versus A_d), we compute the statistic of Diebold and Mariano (1995) (Table 2.8). According to the DM statistics the model with an exponential triggering function, without influence of the sizes on the triggering intensity and with unpredictable event sizes (model C_n), delivers significantly more accurate results for extremes than the same model with a power law triggering function (model A_n), over the period from September 1, 2008, to January 1, 2013. Also, models with an exponential impact function predict the probability on extremes less accurately than models with the other two impact functions.

The Tables 2.9 and 2.10 report the number of correct predictions, the number of false predictions, the hit rate, the false alarm rate and the Hanssen-Kuiper Skill Score of the EWS for (negative) extremes above the 95%, 97% and 99% in sample quantile. As a reference, no events are correctly predicted by an EWS using a homogeneous Poisson model. Here the intensity of the Poisson process is set equal to the number of events during the forecast period divided by the length of the forecast period. The Hanssen-Kuiper Skill Score (KSS) is computed as the hit rate minus the false alarm rate. The hit rate is the number of 'correct alarms' divided by the number of periods that there should be an alarm. The false alarm rate is the number of 'false alarms' divided by the number of periods that there should not be an alarm. Therefore, the KSS accounts for both sorts of errors, that is, for the number of periods there should be an alarm when there is none as well as for the number of periods there

should not be an alarm when there is one. In the first case the hit rate decreases, whereas in the second case the false alarm rate increases, both decreasing the KSS.

The KSS of all EWS are positive, meaning that the rate of correct predictions is higher than the rate of false predictions, whereas KSS of the Poisson-EWS is zero. The KSS of the models with predictable event sizes ($\eta \neq 0$) is in general lower than the the KSS of the same models with unpredictable event sizes. The KSS of models with a size-dependent triggering probability ($\alpha \neq 0$) are higher, except for the 97% and 99% quantile of extreme returns over the period from 1 September, 2008, to 1 January, 2013, and for the 95% quantile of extreme negative returns over the crisis period. The models with the impact function (2.6) perform best in terms of the KSS over the crisis period for both the crashes and the extremes. Over the period from 1 September, 2008, to 1 January, 2013, the KSS of models with the impact function (2.5) is higher for some sets of returns.

Figure 2.5 shows the predicted probability of a financial market crash occurring within 5 days based on the models A_d and C_d, from 1 September, 2008, to 1 January, 2013. The Figure shows that during the financial crisis of 2008 the crash probability is exceptionally high. The crash probability also becomes high when the stock market falls at the end of July/the beginning of August, 2011. Comparing the crash probabilities based on the different models, the model with the power law triggering function (model A_d) forecasts a higher probability on a crash than the model with the exponential triggering function (model C_d). Furthermore during the financial crisis of 2008, the risk of a crash decays more slowly according to model A_d than to C_d.

2.6 Comparison volatility models

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

Table 2.6: Quadratic and Log Probability Scores

				A_n	B_n	C_n	D_n	A_e	B_e	C_e	D_e	A_p	B_p	C_p	D_p	A_d	B_d	C_d	D_d
Crash	2008–2013	95%	QPS	0.389	0.387	0.386	0.391	0.387	0.384	0.383	0.389	0.390	0.385	0.383	0.390	0.387	0.385	0.382	0.391
			LPS	0.578	0.577	0.576	0.580	0.577	0.575	0.574	0.578	0.582	0.576	0.576	0.580	0.576	0.576	0.572	0.581
		97%	QPS	0.314	0.312	0.310	0.314	0.312	0.310	0.308	0.313	0.314	0.311	0.309	0.314	0.314	0.311	0.309	0.315
			LPS	0.487	0.487	0.485	0.488	0.485	0.485	0.483	0.487	0.489	0.486	0.485	0.489	0.488	0.487	0.485	0.490
		99%	QPS	0.208	0.209	0.208	0.209	0.206	0.207	0.206	0.207	0.205	0.206	0.205	0.205	0.203	0.205	0.203	0.205
			LPS	0.339	0.339	0.338	0.340	0.336	0.337	0.336	0.337	0.335	0.335	0.334	0.335	0.332	0.334	0.331	0.335
	2008–2010	95 %	QPS	0.409	0.399	0.399	0.411	0.411	0.398	0.397	0.412	0.421	0.400	0.402	0.416	0.414	0.400	0.400	0.417
			LPS	0.601	0.590	0.590	0.603	0.606	0.590	0.591	0.605	0.624	0.593	0.602	0.611	0.607	0.593	0.592	0.612
		97 %	QPS	0.407	0.400	0.399	0.407	0.404	0.396	0.394	0.405	0.407	0.396	0.395	0.406	0.400	0.397	0.392	0.407
			LPS	0.597	0.588	0.587	0.597	0.595	0.585	0.584	0.595	0.602	0.586	0.588	0.597	0.589	0.587	0.580	0.598
		99 %	QPS	0.375	0.380	0.378	0.377	0.371	0.377	0.375	0.373	0.368	0.375	0.373	0.371	0.365	0.374	0.369	0.370
			LPS	0.565	0.572	0.570	0.567	0.561	0.569	0.568	0.563	0.559	0.568	0.567	0.562	0.553	0.568	0.560	0.561
Extreme	2013–2008	95 %	QPS	0.324	0.324	0.320	0.327	0.318	0.318	0.313	0.321	0.315	0.316	0.310	0.319	0.312	0.315	0.308	0.319
			LPS	0.503	0.505	0.501	0.506	0.495	0.499	0.494	0.499	0.492	0.496	0.490	0.497	0.486	0.496	0.485	0.497
		97 %	QPS	0.258	0.256	0.254	0.260	0.254	0.252	0.249	0.256	0.252	0.250	0.247	0.254	0.251	0.250	0.247	0.255
			LPS	0.420	0.420	0.417	0.423	0.414	0.415	0.411	0.417	0.413	0.414	0.409	0.416	0.411	0.414	0.409	0.417
		99 %	QPS	0.146	0.149	0.148	0.147	0.144	0.148	0.147	0.145	0.144	0.148	0.147	0.145	0.142	0.148	0.145	0.145
			LPS	0.250	0.250	0.247	0.252	0.248	0.249	0.246	0.250	0.248	0.249	0.246	0.250	0.246	0.249	0.244	0.250
	2008–2010	95 %	QPS	0.289	0.286	0.282	0.292	0.283	0.278	0.274	0.286	0.280	0.275	0.270	0.284	0.273	0.275	0.267	0.284
			LPS	0.450	0.449	0.445	0.455	0.442	0.440	0.436	0.447	0.440	0.437	0.432	0.445	0.427	0.437	0.424	0.445
		97 %	QPS	0.301	0.293	0.292	0.303	0.296	0.288	0.288	0.298	0.294	0.286	0.285	0.296	0.287	0.285	0.283	0.295
			LPS	0.471	0.464	0.463	0.474	0.466	0.459	0.458	0.469	0.466	0.458	0.457	0.468	0.452	0.457	0.449	0.467
		99 %	QPS	0.243	0.249	0.245	0.246	0.244	0.252	0.248	0.247	0.246	0.253	0.250	0.249	0.242	0.253	0.246	0.250
			LPS	0.398	0.398	0.391	0.403	0.400	0.400	0.394	0.405	0.404	0.402	0.395	0.409	0.397	0.402	0.390	0.410

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

Table 2.7: Adjusted Mean Squared Prediction Errors

		A_n							B_n			C_n							D_n			A_e	A_p	A_d	C_e	C_p	C_d
		B_n	A_e	A_p	A_d	B_e	B_p	B_d	B_e	B_p	B_d	D_n	C_e	C_p	C_d	D_e	D_p	D_d	D_e	D_p	D_d	B_e	B_p	B_d	D_e	D_p	D_d
Crash	2008–2013	1.45	1.54	1.02	1.35	2.29	2.18	2.13	1.92	1.56	1.47	−0.90	2.18	1.92	2.16	0.02	0.06	0.01	1.39	0.82	0.70	1.60	2.24	1.56	−1.03	−0.94	−1.85
	2010–2013	2.39	0.37	0.04	−0.04	2.38	2.00	1.97	0.80	0.54	0.47	−1.54	0.91	0.68	0.43	−0.67	−0.58	−0.62	0.36	−0.01	−0.10	2.52	3.04	2.22	−1.42	−1.23	−2.00
Extreme	2008–2013	0.79	4.98	4.71	4.23	2.50	3.06	3.10	4.63	4.72	4.62	−1.90	4.54	4.72	4.20	0.29	1.17	1.24	5.04	4.70	4.46	0.85	0.75	−0.10	−1.66	−1.68	−1.79
	2010–2013	0.93	2.77	2.93	2.73	1.86	2.15	2.18	2.85	3.13	3.18	−1.13	2.62	2.89	2.53	−0.14	0.34	0.40	2.91	3.04	3.04	1.14	1.1	0.01	−1.00	−0.98	−0.88

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

Table 2.8: Diebold Mariano Statistics

		$A_n - C_n$	$A_n - D_n$	$B_n - C_n$	$B_n - D_n$	$A_e - A_p$	$A_e - A_d$	$A_p - A_d$	$B_e - B_p$	$B_e - B_d$	$B_p - B_d$	$C_e - C_p$	$C_e - C_d$	$C_p - C_d$	$D_e - D_p$	$D_e - D_d$	$D_p - D_d$
Crash	2008–2013	1.17	−2.09	1.44	−1.27	−1.15	−0.16	0.86	−0.37	−0.53	−1.48	−1.66	−0.74	0.97	−1.15	−1.30	−2.01
	2008–2010	1.87	−1.52	0.34	−1.99	−1.15	−0.16	0.86	−0.37	−0.53	−1.48	−0.90	−0.61	0.24	−1.15	−1.30	−2.01
Extreme	2008–2013	1.51	−1.66	2.16	−1.23	2.46	2.19	1.68	2.72	2.39	0.90	−0.87	−0.97	−1.14	2.22	1.85	0.58
	2008–2010	0.95	−1.52	1.20	−1.11	1.67	1.75	1.43	2.23	2.18	1.08	−1.48	−1.60	−2.00	1.65	1.55	0.88

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

Table 2.9: Results for the Early Warning Systems 2008–2013

			A_n	B_n	C_n	D_n	A_e	B_e	C_e	D_e	A_p	B_p	C_p	D_p	A_d	B_d	C_d	D_d
Crash	95 %	Hit	249	243	242	248	249	242	240	248	253	246	248	251	253	246	247	252
		False	137	135	134	140	137	135	133	139	144	137	135	143	146	137	132	144
		Hit rate	0.615	0.600	0.598	0.612	0.615	0.598	0.593	0.612	0.625	0.607	0.612	0.620	0.625	0.607	0.610	0.622
		False rate	0.188	0.186	0.184	0.193	0.188	0.186	0.183	0.191	0.198	0.188	0.186	0.197	0.201	0.188	0.182	0.198
		KSS	0.426	0.414	0.413	0.420	0.426	0.412	0.410	0.421	0.427	0.419	0.427	0.423	0.424	0.419	0.428	0.424
	97 %	Hit	180	170	174	179	184	177	181	182	193	183	186	191	189	184	184	191
		False	108	100	98	104	108	103	96	108	113	105	103	110	110	106	98	110
		Hit rate	0.564	0.533	0.545	0.561	0.577	0.555	0.567	0.571	0.605	0.574	0.583	0.599	0.592	0.577	0.577	0.599
		False rate	0.133	0.123	0.121	0.128	0.133	0.127	0.118	0.133	0.139	0.129	0.127	0.135	0.135	0.130	0.121	0.135
		KSS	0.431	0.410	0.425	0.433	0.444	0.428	0.449	0.438	0.466	0.445	0.456	0.463	0.457	0.446	0.456	0.463
	99 %	Hit	68	64	59	66	69	71	70	71	79	78	77	79	68	80	62	85
		False	34	35	34	35	37	37	36	37	43	41	40	41	37	43	33	47
		Hit rate	0.360	0.339	0.312	0.349	0.365	0.376	0.370	0.376	0.418	0.413	0.407	0.418	0.360	0.423	0.328	0.450
		False rate	0.036	0.037	0.036	0.037	0.039	0.039	0.038	0.039	0.046	0.043	0.042	0.043	0.039	0.046	0.035	0.050
		KSS	0.324	0.302	0.276	0.312	0.326	0.336	0.332	0.336	0.372	0.369	0.365	0.375	0.321	0.378	0.293	0.400
Extreme	95%	Hit	341	333	334	341	339	330	333	341	341	332	336	344	340	334	336	345
		False	120	113	107	118	112	106	100	114	112	105	98	115	114	106	100	117
		Hit rate	0.737	0.719	0.721	0.737	0.732	0.713	0.719	0.737	0.737	0.717	0.726	0.743	0.734	0.721	0.726	0.745
		False rate	0.179	0.169	0.160	0.176	0.167	0.158	0.149	0.170	0.167	0.157	0.146	0.172	0.170	0.158	0.149	0.175
		KSS	0.557	0.550	0.561	0.560	0.565	0.554	0.570	0.566	0.569	0.560	0.579	0.571	0.564	0.563	0.576	0.570
	97%	Hit	270	264	267	269	262	260	259	264	262	259	258	263	264	262	254	266
		False	108	106	102	109	99	97	95	103	100	98	91	103	97	99	90	102
		Hit rate	0.761	0.744	0.752	0.758	0.738	0.732	0.730	0.744	0.738	0.730	0.727	0.741	0.744	0.738	0.715	0.749
		False rate	0.139	0.136	0.131	0.140	0.127	0.125	0.122	0.133	0.129	0.126	0.117	0.133	0.125	0.127	0.116	0.131
		KSS	0.622	0.607	0.621	0.617	0.611	0.608	0.607	0.611	0.609	0.603	0.610	0.608	0.619	0.611	0.600	0.618
	99%	Hit	122	117	115	121	118	116	113	119	119	117	115	121	118	116	112	121
		False	62	58	57	60	55	53	52	56	54	53	52	53	52	54	51	54
		Hit rate	0.761	0.744	0.752	0.758	0.738	0.732	0.730	0.744	0.738	0.730	0.727	0.741	0.744	0.738	0.715	0.749
		False rate	0.139	0.136	0.131	0.140	0.127	0.125	0.122	0.133	0.129	0.126	0.117	0.133	0.125	0.127	0.116	0.131
		KSS	0.622	0.607	0.621	0.617	0.611	0.608	0.607	0.611	0.609	0.603	0.610	0.608	0.619	0.611	0.600	0.618

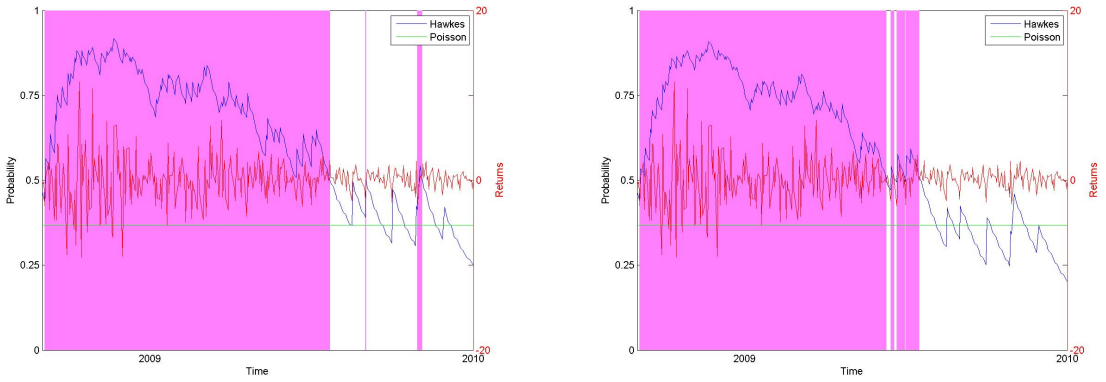
Hits, hit rates, false alarms, false alarm rates and the Hanssen-Kuiper Skill Scores (KSS) of the EWS predicting the occurrence of events within 5 days from 1 September, 2008, to 1 January, 2013. An event is predicted when the probability according to the models given the history of the event process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

Table 2.10: Results for the Early Warning Systems 2008–2010

			A_n	B_n	C_n	D_n	A_e	B_e	C_e	D_e	A_p	B_p	C_p	D_p	A_d	B_d	C_d	D_d
Crash	95 %	Hit	161	158	156	161	160	157	155	160	160	159	158	160	160	159	157	160
		False	66	63	62	68	68	64	63	69	74	65	65	71	74	65	64	72
		Hit rate	0.866	0.849	0.839	0.866	0.860	0.844	0.833	0.860	0.860	0.855	0.849	0.860	0.860	0.855	0.844	0.860
		False rate	0.402	0.384	0.378	0.415	0.415	0.390	0.384	0.421	0.451	0.396	0.396	0.433	0.451	0.396	0.390	0.439
		KSS	0.463	0.465	0.461	0.451	0.446	0.454	0.449	0.439	0.409	0.458	0.453	0.427	0.409	0.458	0.454	0.421
	97 %	Hit	140	133	134	140	139	133	133	140	142	134	134	143	139	134	133	143
		False	69	64	61	70	69	63	58	70	70	64	62	71	66	65	57	71
		Hit rate	0.800	0.760	0.766	0.800	0.794	0.760	0.760	0.800	0.811	0.766	0.766	0.817	0.794	0.766	0.760	0.817
		False rate	0.394	0.366	0.349	0.400	0.394	0.360	0.331	0.400	0.400	0.366	0.354	0.406	0.377	0.371	0.326	0.406
		KSS	0.406	0.394	0.417	0.400	0.400	0.400	0.429	0.400	0.411	0.400	0.411	0.411	0.417	0.394	0.434	0.411
	99 %	Hit	68	64	59	66	69	71	70	71	78	78	77	79	65	80	62	84
		False	34	35	34	35	37	37	36	37	41	41	40	41	34	43	33	44
		Hit rate	0.531	0.500	0.461	0.516	0.539	0.555	0.547	0.555	0.609	0.609	0.602	0.617	0.508	0.625	0.484	0.656
		False rate	0.153	0.158	0.153	0.158	0.167	0.167	0.162	0.167	0.185	0.185	0.180	0.185	0.153	0.194	0.149	0.198
		KSS	0.378	0.342	0.308	0.358	0.372	0.388	0.385	0.388	0.425	0.425	0.421	0.433	0.355	0.431	0.336	0.458
Extreme	95 %	Hit	218	218	215	221	217	214	212	220	217	213	212	220	215	214	211	220
		False	57	52	50	58	54	47	45	56	53	45	43	56	48	45	39	56
		Hit rate	0.924	0.924	0.911	0.936	0.919	0.907	0.898	0.932	0.919	0.903	0.898	0.932	0.911	0.907	0.894	0.932
		False rate	0.500	0.456	0.439	0.509	0.474	0.412	0.395	0.491	0.465	0.395	0.377	0.491	0.421	0.395	0.342	0.491
		KSS	0.424	0.468	0.472	0.428	0.446	0.494	0.504	0.441	0.455	0.508	0.521	0.441	0.490	0.512	0.552	0.441
	97 %	Hit	187	185	184	187	186	182	181	186	186	180	178	186	185	181	171	186
		False	51	47	45	52	47	44	43	50	46	43	39	48	42	43	36	47
		Hit rate	0.903	0.894	0.889	0.903	0.899	0.879	0.874	0.899	0.899	0.870	0.860	0.899	0.894	0.874	0.826	0.899
		False rate	0.357	0.329	0.315	0.364	0.329	0.308	0.301	0.350	0.322	0.301	0.273	0.336	0.294	0.301	0.252	0.329
		KSS	0.547	0.565	0.574	0.540	0.570	0.572	0.574	0.549	0.577	0.569	0.587	0.563	0.600	0.574	0.574	0.570
	99 %	Hit	111	109	107	111	109	108	105	110	110	109	106	112	108	108	103	112
		False	45	45	43	46	45	42	41	46	45	42	41	45	43	42	39	45
		Hit rate	0.860	0.845	0.829	0.860	0.845	0.837	0.814	0.853	0.853	0.845	0.822	0.868	0.837	0.837	0.798	0.868
		False rate	0.204	0.204	0.195	0.208	0.204	0.190	0.186	0.208	0.204	0.190	0.186	0.204	0.195	0.190	0.176	0.204
		KSS	0.657	0.641	0.635	0.652	0.641	0.647	0.628	0.645	0.649	0.655	0.636	0.665	0.643	0.647	0.622	0.665

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

Figure 2.5: Crash probability predictions 95% quantile



(a) Model A_d **(b)** Model C_d

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

In General AutoRegressive Conditional Heteroscedasticity (GARCH) type models, the variance of the current error term is a function of the error terms and innovations in previous periods. The time-varying conditional variance enables the models to capture the volatility clustering feature of stock market returns. After evaluating the performance of several GARCH-type models (GARCH, AGARCH, NAGARCH, GJR, EGARCH, and PGARCH) and error distributions (Normal, Student-t, and Gumbel), we continue to consider two GARCH-type models, that is the GARCH(1,1) model and the GJR(1,1) model, in combination with a Student-t distribution for the error terms. The heavy-tailed Student-t distribution accounts for the stylized fact that, even after correcting for volatility clustering, extreme returns occur more often than under normality (Cont, 2001). As shown by Hansen and Lunde (2005), it is difficult to beat the GARCH(1,1) model when forecasting conditional volatility. The GJR(1,1) model of Glosten et al. (1993) differs from the GARCH(1,1) model, as the model allows for separate influences of positive and negative innovations on future volatility. This asymmetric response to shocks, or so-called "leverage effect", is in line with the observation that measures of the volatility of assets tend to correlate negatively with the returns on those assets (Cont, 2001). This can be of an advantage when modelling returns (Hua and Manzan, 2013). The conditional variance σ_t^2 of the stochastic error process ϵ_t, when conditioned on the history of the process, in respectively the GARCH(1,1) and the GJR(1,1) model is

specified as follows

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{2.15}$$

$$\sigma_t^2 = \omega + \left(\alpha\epsilon_{t-1}^2 + \gamma\mathcal{I}\left[\epsilon_{t-1} < 0\right]\right)\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{2.16}$$

Autoregressive Conditional Duration type models first proposed by Engle and Russell (1998), focus on modeling the expected duration between events. As well as in the ETAS models, the conditional intensity in these models is a function of the time between past events. Furthermore the event process on which the models are based is self-exciting. Analogous to GARCH-type and ETAS models, ACD-type models are therefore able to pick up the characteristic clustering of extreme stock market returns. After evaluating the performance of several ACD-type models and error distributions we consider two ACD-type models, that is the ACD(1,1) model and the log-ACD(1,1) model, in combination with a Generalized Gamma distribution. Like Bauwens et al. (2004) and Allen et al. (2009), we find other ACD-type models much more costly to estimate and to evaluate, but not of superior performance, while the use of the Generalized Gamma distribution instead of the exponential or Weibull distribution does add to the performance of the models. A survey on ACD-type models is provided by Pacurar (2008).

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

$$\theta_t = \omega + \alpha\tau_{t-1} + \beta\theta_{t-1} \tag{2.17}$$

$$\log(\theta_t) = \omega + \alpha\log(\tau_{t-1}) + \beta\log(\theta_{t-1}) \tag{2.18}$$

We apply the models again for daily return data on the S&P 500 index between 2 January, 1957, and 1 September, 2008. Like the ETAS models, the volatility models are estimated by maximum likelihood in combination with the Nelder-Mead simplex direct search algorithm.

Consulting Andersen et al. (2006) and Lau and McSharry (2010) for forecasting multi-step ahead densities with GARCH-type models, we use Monte Carlo methods to derive the probability of the occurrence of no event within a certain period. For the ACD-type models this probability follows easily from the model specification, by noticing it is equal to the probability of the duration exceeding the time period.

2.6.1 In sample results

In order to evaluate the ability of ETAS models to exploit information in the returns series not captured by GARCH-type models, we apply the ETAS models to the standardized residuals from the GARCH-type models. The results of this exercise using ETAS models, with an exponential triggering function and no influence of the sizes of events on the triggering intensity, are reported in Table 2.11. We do not display the results for the models with influence of the sizes of events on the triggering intensity, as likelihood ratio tests show that the α parameter is not significant for all impact functions at a 5% level. However, all other estimated parameters are significant. Therefore, our evidence for ETAS models is not simply driven by volatility clustering. As the η parameter is significantly different from zero at a 5% level, and the models with a non-zero η parameter have a higher log-likelihood and a lower AIC value than their counterparts with $\eta = 0$, the size of future extreme residuals appears to be affected by the history of the event process.

We check the goodness-of-fit of the ETAS models by means of residual analysis, and verify whether the distribution of the transformed times is unit Poisson with the Kolmogorov-Smirnov (KS) test. The p-values of the KS tests applied to the 95%, 97% and 99% quantile of extreme (negative) standardized residuals, are reported in Table 2.12. For all sets of returns, the standardized residuals do not deviate from an event process specified by an ETAS model at a 5% level. Hence, we can conclude that the ETAS models are needed for modeling the standardized residuals from the GARCH-type models as the GARCH-type models cannot accommodate the self-exciting behavior of extremes.

2.6.2 Out-of-sample results

Table 2.13 reports the results of the EWS based on the GARCH-type, the ACD-type and two ETAS models for the extreme (negative) returns above the 95–99% in sample quantile.

Table 2.11: Estimation results standardized residuals

	Extreme				Crash			
	GARCH		GJR		GARCH		GJR	
	C_n	D_n	C_n	D_n	C_n	D_n	C_n	D_n
μ	0.0076	0.0076	0.0109	0.0102	0.0109	0.0106	0.0114	0.0112
	(0.0022)	(0.0022)	(0.0028)	(0.0027)	(0.0029)	(0.0028)	(0.0033)	(0.0033)
K_0	0.0100	0.0098	0.0131	0.0121	0.0098	0.0093	0.0063	0.0060
	(0.0016)	(0.0015)	(0.0024)	(0.0022)	(0.0020)	(0.0019)	(0.0012)	(0.0011)
β	0.0118	0.0115	0.0168	0.0152	0.0125	0.0118	0.0081	0.0078
	(0.0019)	(0.0018)	(0.0036)	(0.0031)	(0.0028)	(0.0026)	(0.0016)	(0.0015)
ξ	0.1541	0.1526	0.1651	0.1585	0.1059	0.1121	0.1464	0.1466
	(0.0412)	(0.0410)	(0.0406)	(0.0401)	(0.0374)	(0.0372)	(0.0388)	(0.0380)
ϕ	1.2262	0.9600	1.2320	0.9252	1.4441	1.1862	1.3266	0.9132
	(0.0685)	(0.1215)	(0.0682)	(0.1191)	(0.0769)	(0.1503)	(0.0718)	(0.1628)
η		0.0450		0.0716		0.0444		0.0544
		(0.0192)		(0.0258)		(0.0239)		(0.0221)
$\log L(\boldsymbol{\theta})$	-3562.29	-3559.41	-3591.94	-3587.93	-3666.76	-3664.92	-3673.76	-3670.39
AIC	7140.58	7132.82	7199.87	7189.85	7349.52	7343.83	7363.52	7354.79

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

Table 2.12: Kolmogorov-Smirnov tests

	Extreme				Crash			
	GARCH		GJR		GARCH		GJR	
	C_n	D_n	C_n	D_n	C_n	D_n	C_n	D_n
95%	0.174	0.280	0.189	0.316	0.182	0.154	0.181	0.146
97%	0.565	0.406	0.582	0.406	0.178	0.179	0.179	0.178
99%	0.583	0.121	0.583	0.122	0.050	0.204	0.050	0.204

The tests are performed on the transformed times $\{\tau_i\}$ specified by the models. The models are applied to the 95% to 99% quantile of the extreme (negative) standardized residuals from the GARCH(1,1) and GJR(1,1) model when estimated on the S&P returns over the period January 2, 1957, until September 1, 2008. The null hypothesis of the test is transformed times $\{\tau_i\}$ are distributed according to a homogeneous Poisson process with intensity 1. In the Table the p-values of the Kolmogorov-Smirnov tests for the 95%, 97% and 99% quantile are reported.

It is immediately apparent from the tables that the ACD-type models do not perform well. While the EWS based on the the ACD(1,1) model predicts the occurrence of an event almost every period, the EWS based on the log-ACD(1,1) model predicts far too few events over the out-of-sample period. This results in a KSS around zero (the KSS is even slightly negative for some model-quantile combinations).

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

However forecasting with the GARCH-type models is much more time consuming than forecasting with the other models. Using these models, the probability distribution of the occurrence of one or more events during an certain time period, has to be derived empirically by means of a Monte Carlo procedure. Roughly about one to two hours are needed to execute the Monte Carlo simulation and deduce alarm signals over the out-of-sample period (when

using $10,000$ replications), while with EWS based on ETAS models, it takes no more than half a second to compute forecasts.

2.7 Conclusion

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

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

The models are applied to the 95%–99% quantile of extreme (negative) returns on the S&P 500 index over a period from 2 January, 1957, to 1 September, 2008. The estimation results confirm that like earthquakes, crashes are reinforcing. The decay of the probability of triggering events seems better modeled by the power law distribution than by the exponential distribution. The sizes of events are history dependent, as on average larger extremes are observed after the occurrence of more and/or big events than after a tranquil period. The triggering probability is size-dependent, as larger events trigger on average more events than smaller events. However, while the exponential function is commonly used to model the impact of the sizes of events on the triggering of subsequent events, other impact functions better fit the data considered.

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

We develop an Early Warning System for events in the financial market based on the probability of the occurrence of an event within a certain time period predicted by the models. These are reviewed from 1 September, 2008, to 1 January, 2010, and from 1 September, 2008, to 1 January, 2013. Testing the EWS, the rate of correct predictions is higher than the rate of false predictions. Thus as our modeling framework exploits the self-exciting behavior

Table 2.13: Comparison EWS

		95%		97%		99%		95%		97%		99%		95%		97%		99%	
		GARCH	GJR	GARCH	GJR	GARCH	GJR	ACD	log-ACD	ACD	log-ACD	ACD	log-ACD	A_d	C_d	A_d	C_d	A_d	C_d
Crash	Hit	245	246	187	187	62	80	325	9	260	18	157	12	249	240	184	181	69	70
	False	131	136	101	103	54	52	662	32	699	67	719	27	137	133	108	96	37	36
	Hit rate	0.605	0.607	0.586	0.586	0.328	0.423	0.802	0.022	0.815	0.056	0.831	0.063	0.615	0.593	0.577	0.567	0.365	0.370
	False rate	0.180	0.187	0.124	0.127	0.057	0.055	0.911	0.044	0.860	0.082	0.762	0.029	0.188	0.183	0.133	0.118	0.039	0.038
	KSS	0.425	0.420	0.462	0.460	0.271	0.368	−0.108	−0.022	−0.045	−0.026	0.068	0.035	0.426	0.410	0.444	0.449	0.326	0.332
Extreme	Hit	217	222	153	157	52	52	303	23	248	8	0	11	339	333	262	259	118	113
	False	23	21	17	17	10	11	453	41	708	54	0	3	112	100	99	95	55	52
	Hit rate	0.469	0.479	0.431	0.442	0.294	0.294	0.654	0.050	0.699	0.023	0.000	0.062	0.732	0.719	0.738	0.730	0.738	0.730
	False rate	0.034	0.031	0.022	0.022	0.010	0.012	0.677	0.061	0.911	0.069	0.000	0.003	0.167	0.149	0.127	0.122	0.127	0.122
	KSS	0.434	0.448	0.409	0.420	0.283	0.282	−0.023	−0.012	−0.213	−0.047	0.000	0.059	0.565	0.570	0.611	0.607	0.611	0.607

Hits, hit rates, false alarms, false alarm rates and the Hanssen-Kuiper Skill Scores (KSS) of the EWS predicting the occurrence of a negative, absolute, return exceeding the $0.95 - 0.99\%$ in sample quantile from 1 September, 2008, to 1 January, 2013. An alarm is set when the probability of an event within 5 days is above 0.5 according to the models.

of stock returns around financial market crashes, it is capable of creating crash probability predictions in the medium term. Over all sets of returns, models with unpredictable event sizes and non-zero influence of the size of events on the triggering probability, perform best according to the Hanssen-Kuiper Skill Score.

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

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

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

2.A Simulation procedures

Discrete time procedure

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

2. For t_n after t_1 calculate the probability of the occurrence of no event in the interval $[t_{n-1}, t_n]$, that is (4.7). Simulate a random number u from a uniform distribution on the interval $[0, 1]$. When $u > P\left(N(t_n) - N(t_{n-1}) = 0\right)$ record the time point t_n as the time of an event and simulate the magnitude of the event from a General Pareto Distribution. When $u < P\left(N(t_n) - N(t_{n-1}) = 0\right)$ do nothing.

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

Continuous time procedure

1. Sample the background events

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

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

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

2. Sample the triggered events

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

$$G(T - t_{parent}, m_{parent}) = \frac{K_0}{\gamma\omega}\left(1 - (\gamma(T - t_{parent}) + 1)^{-\omega}\right) e^{\alpha(m_{parent} - M_0)} \tag{2.19}$$

$$G(T - t_{parent}, m_{parent}) = \frac{K_0}{\beta}\left(1 - e^{-\beta(T - t_{parent})}\right) e^{\alpha(m_{parent} - M_0)} \tag{2.20}$$

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

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

$$P\left(t^* \leq t_0 + t | t^* > t_0, t^* < T\right) = \frac{F(t_0 + t) - F(t_0)}{F(T) - F(t_0)} = \frac{S(t_0) - S(t_0 + t)}{S(t_0) - S(T)} \tag{2.21}$$

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

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

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

Chapter 3

Specification testing in Hawkes models

3.1 Introduction

Hawkes models are useful to model and to predict extreme events in financial markets.[1] We propose Lagrange Multiplier (LM) based specification tests for Hawkes models, where the specific focus is the potential extension of a univariate model to a multivariate model.[2] Even though the Hawkes models have become increasingly popular recently, there does not seem to be a general approach for testing different specifications of these models. Our main focus in this paper is on testing for (conditional) dependence between series. However the LM based specification tests can also be used to test for omitted explanatory variables, for breaks in the model parameters, for omitted impact of the sizes of events on the triggering of new events and for omitted predictability of event sizes. The performance of our LM tests is evaluated in a simulation study. Applying the specification test for dependence to US stocks and bonds data, we show that there are cross-excitation effects from one market to another.

The identification and prediction of crashes is important to traders, to regulators of financial markets and to risk management. A series of large negative price movements during a short time interval can have severe consequences (see Kindleberger and Aliber, 2011, among many others). Hawkes models can be well used to identify and predict large negative price

[1] See amongst others: Chavez-Demoulin et al. (2005), Hewlett (2006), Bowsher (2007), Bauwens and Hautsch (2009), Herrera and Schipp (2009), Embrechts et al. (2011), Chavez-Demoulin and McGill (2012), Aït-Sahalia et al. (2014), Gourieroux et al. (2014), Grothe et al. (2014), Aït-Sahalia et al. (2015), Bacry et al. (2015), Bormetti et al. (2015) and Gresnigt et al. (2015)

[2] This Chapter is based on Gresnigt et al. (2016a)

movements.[3] Of course, the Hawkes process should be properly specified, and hence specification tests are useful.

The Hawkes process, first proposed by Hawkes (1971), is an inhomogeneous Poisson process in which the event rate of the process increases with the arrival of an event and the event rate decays as a function of the time passed since the event. As the probability of events increases when an event occurs, the Hawkes process is called a self-exciting process. Hence, while events can be triggered by an exogenous factor, for a Hawkes process risk of events grows endogenously. Characteristics typically observed in data that fit Hawkes models are the clustering of events and serial dependence.

The earthquake literature, which brought the Hawkes models (Ogata, 1988), describes clustering behaviour of seismicity in time as well as in space, whereas financial series seem to exhibit clustering behaviour in the cross section on top of the clustering behaviour in the time-dimension. Hartmann et al. (2004), Gonzalo and Olmo (2005) and Hu (2006) conclude that when stocks crash, bonds are more likely to co-crash or otherwise to boom. There are efforts in the literature to detect dependence between series and to evaluate the fit of specific models, but an adequate assessment of the contribution of including cross-sectional dependence in a point process framework with serial dependence does not seem to exist. In this paper we therefore focus on specification tests for contagion of extreme events across series.

Our focal variable is the rate at which extreme events occur, and therefore we wish to examine if there are exogenous variables that have explanatory power for this rate. When modelling earthquakes it can be useful to consider precursors such as unusual animal behaviour and temperature changes which can signal upcoming disruptions between tectonical plates (Rikitake, 1978; Cicerone et al., 2009). Using the resemblance of financial crashes and earthquakes, as is done in Gresnigt et al. (2015), there may be precursors that are of interest when modelling financial series. Variables one can think of are liquidity measures (Baker and Stein, 2004), VIX options (Chung et al., 2011), inflation, industrial production, the three-month T-bill rate, the 12-month treasury bond, the dividend yield and the price earnings ratio (Longin and Solnik, 1995; Pesaran and Timmermann, 1995; Campbell and Thompson, 2008, see amongst others). In this paper we design Lagrange Multiplier based tests that can de-

[3] See amongst others: Chavez-Demoulin et al. (2005), Bowsher (2007), Bauwens and Hautsch (2009), Herrera and Schipp (2009), Embrechts et al. (2011), Chavez-Demoulin and McGill (2012), Grothe et al. (2014), Aït-Sahalia et al. (2015), Bacry et al. (2015), Bormetti et al. (2015) and Gresnigt et al. (2015)

termine whether adding explanatory variables to Hawkes models is beneficial. As far as we know, no such tests are available yet.

The LM principle can also be used to construct specification tests for the omitted impact of the sizes of events on the triggering of new events and for the omitted predictability of event sizes. The relevance of these tests is shown by Chavez-Demoulin et al. (2005), Herrera and Schipp (2009), and Gresnigt et al. (2015). They find that the probability that an extreme return on a stock market index triggers another extreme return, is larger when the initial event is larger. Also they document that current tension in a financial market, when measured by the history of the event process, has predictive value for the size of future events. Furthermore the LM tests can be used to investigate the presence or the timing of a structural break in the model parameter even when one is unsure about the timing of the break. For example, changing financial conditions could be reflected in the parameters of the Hawkes models when they are applied to financial series and calibrated over a longer time period.[4]

We use the Lagrange Multiplier principle (see Breusch and Pagan, 1980; Engle, 1982; Rogers, 1986; Hamilton, 1996, amongst many others) to design specification tests to assess the validity of different specifications of Hawkes models. An attractive feature of the LM test is that the models do not have to be estimated under the alternative hypothesis. We show that test statistics of a variety of Hawkes models can be computed from the estimation of the simplest Hawkes model. The LM test principle is particularly relevant as estimating the parameters of Hawkes models is quite difficult and time demanding, especially under extensive specifications. Like Hamilton (1996), for Markov-switching models, we provide expressions for the score in Hawkes models, where the score is defined as the derivative of the conditional log-likelihood of the n-th observation with respect to the parameter under investigation. When we test for dependence between series, we compare the results of the LM test with the regression method of Van Oordt and Zhou (2012) and the orthogonality test discussed by Lando and Nielsen (2010).

When we evaluate the size and power of the tests, by means of a Monte Carlo analysis, we find that the LM tests outperform the regression method and the orthogonality test in detecting cross-triggering effects. Over longer time periods, the power of the LM test is high and the LM test is properly sized. When the time period considered is shorter, the quality of

[4] See amongst others: Koutmos and Booth (1995), Lynch and Mendenhall (1996), Poon et al. (2004), Beine et al. (2010)

the LM test falls only a little. When we compare the LM test to the computationally more demanding LR test, we find that the performance of the LR test is only a bit superior to the performance of the LM test. As the LR test is a lot more time consuming and as the LM test performs very well, we recommend to use the LM test in empirical analysis.

We illustrate the use of the LM test by examining the cross-sectional dependence between 3 US stock indices, a US bond index and 3 exchange rates. Between markets, we find that crashes in the stock markets provoke extremes in the other markets occur where the triggering effect from the equity indices to the US bond market is the strongest. We do not find spillover effects from the bond and exchange rate markets to the stock market. We detect both stock-bond contagion and the flight-to-quality phenomenon, that is the outflow of capital from stock markets to bond markets when the first is facing crises periods. Furthermore, with the LM test we discover that booms in the US bond market trigger booms in the euro/$-rate. Within the stock market and exchange rate market segments, we also find strong evidence of cross-excitation. Moreover, filtering the series for volatility, we conclude that the extremal dependence found between the series cannot simply be explained by volatility spillovers.

Our paper is organized as follows. In Section 2 we go over some preliminaries on Hawkes models. Section 3 provides LM based specification tests for a variety of possible departures of Hawkes models, where the main focus is on dependence across events. Also other tests for dependence across extremes are discussed. Section 4 contains a simulations-based assessment of the empirical performance of various test statistics. In section 5 we illustrate the LM test for US stocks, bonds and exchange rate data. Section 6 summarizes.

3.2 Hawkes models

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

Consider an event process (t_1, m_1),...,(t_N, m_N) where t_n defines the time and m_n the mark of event n. Let $\mathcal{H}_t = \{(t_n, m_n) : t_n < t\}$ represent the entire history of events up to

time t. The conditional intensity of jump arrivals following a Hawkes process is given by

$$\lambda(t|\boldsymbol{\theta};\mathcal{H}_t) = \mu + \gamma \sum_{n:t_n<t} g(t-t_n, m_n) \tag{3.1}$$

where $\mu, \gamma > 0$ and $g(s,m) > 0$ whenever $s, m > 0$ and 0 elsewhere. The parameter γ controls maximum triggering intensity and the expected number of events directly triggered by an event. The conditional intensity consists of a constant term μ and a self-exciting function $\gamma g(s,m)$, which depends on the time passed since jumps that occurred before t and the size of these jumps. Thus the rate at which events arrive is separated into a long-term background component and a short-term clustering component describing the temporal distribution of aftershocks. The conditional intensity uniquely determines the distribution of the process.

As frequently done in the literature on Hawkes processes,[5] we specify the triggering function as suggested by Hawkes (1971), that is

$$g(s,m) = e^{-\beta s} c(m) \tag{3.2}$$

where β determines how fast the possibility of triggering events decays depending on the time passed since an event. The influence of the sizes of past events on the intensity is given by $c(m)$.[6]

To make our tests most suitable for financial series, we choose a Generalized Pareto Distribution for the sizes of the events above the threshold u following the literature on

[5] See amongst others: Hewlett (2006), Bowsher (2007), Herrera and Schipp (2009), Embrechts et al. (2011), Aït-Sahalia et al. (2014), Grothe et al. (2014), Gourieroux et al. (2014), Aït-Sahalia et al. (2015), Bormetti et al. (2015) and Gresnigt et al. (2015)

[6] We acknowledge that some papers use power-law kernels when modelling financial time series with Hawkes processes. Choosing the exponential decay function and making the additional assumption that the decay speeds among multiple series are the same, Hawkes processes become Markov processes. This facilitates simulation and estimation of the models numerically. Moreover the conditions for stationarity of multivariate Hawkes processes of Liniger (2009) and Embrechts et al. (2011), are not satisfied for multivariate models with the power-law triggering function such that stationarity of these models cannot easily be guaranteed. Furthermore the exponential specification enables one to derive the covariance density of the Hawkes process analytically (Hawkes, 1971).

extremes in finance,[7] that is

$$F(M \leq m|\boldsymbol{\theta}; \mathcal{H}_t) = 1 - \left(1 + \xi \frac{m-u}{\sigma(t)}\right)^{-1/\xi} \tag{3.3}$$

where $\sigma(t) = \phi + \eta \sum_{n:t_n<t} g(t - t_n, m_n)$ and $\xi \neq 0$. In case $\eta = 0$ the sizes of the events are unpredictable, whereas in case $\eta \neq 0$ the times and sizes of previous events affect the probability distribution of the sizes of subsequent events. The mean and variance of the distribution of the sizes of excess magnitudes events scale with $\sigma(t)$. Thus when $\eta > 0$ the magnitude of events is expected to be more extreme when the conditional intensity is high. The larger positive η is, the more pronounced is the influence of the history of events on the size of subsequent events.

For the influence of the sizes of past events on triggering of future events $c(m)$, we consider the specification proposed by Grothe et al. (2014). They use the probability that an event has a magnitude smaller than or equal to its' magnitude, $F(m)$, to the extract information from the magnitudes of events. The impact of events of size m on the triggering intensity is constructed as

$$c(m) = 1 - G^{\leftarrow}(F(m)) \tag{3.4}$$

where $G^{\leftarrow}$ is the inverse of the distribution function G of some continuous positive random variable with mean α. Combining (3.4) with conditional distribution (4.4) for the sizes of events, is particularly convenient, as stationarity conditions for the Hawkes process can be derived even when the sizes of events are not independent and identically distributed (IID) (Grothe et al., 2014, Appendix A).[8] Following Grothe et al. (2014) we choose for G the exponential distribution function such that using (4.4), (3.4) takes the form

$$c(m) = 1 + \frac{\alpha}{\xi} \log\left(1 + \xi \frac{m-u}{\sigma(t)}\right) \tag{3.5}$$

where α determines how the size of an event affects the probability of triggering other events. When $\alpha > 0$, larger events trigger more events than smaller events as the probability of

[7] See amongst others: Chavez-Demoulin et al. (2005), Poon et al. (2004), Herrera and Schipp (2009), Chavez-Demoulin and McGill (2012), Grothe et al. (2014) and Gresnigt et al. (2015)

[8] While the exponential form $c(m) = e^{\alpha(m-u)}$ is commonly used to define the influence of the sizes of events on the triggering intensity (Chavez-Demoulin et al., 2005; Herrera and Schipp, 2009; Chavez-Demoulin and McGill, 2012) this form does not guarantee stationarity of the Hawkes process when the sizes of events follow the Generalized Pareto Distribution (4.4).

triggering events increases with the size of the events (m). The larger positive α is, the more pronounced is the influence of the sizes of events. Whenever $\eta > 0$ the impact of the events' sizes depends on the history of the event process, such that when the conditional intensity is low and the sizes of events are expected to be less extreme, the impact of events of the same magnitude is higher.

The Hawkes process is stationary when the average number of descendants triggered by an event, that is the branching ratio n, is smaller than 1. When $n \geq 1$ the rate at which events arrive will grow to infinity over time, moreover, the number of events arriving within a finite time interval could go to infinity. This is an important issue, as our tests are only useful for stationary Hawkes processes.[9] In contrast to Hawkes processes in which the sizes of events have no impact or/and the sizes of events are IID, stationarity conditions for Hawkes process in which the sizes of events do influence the conditional intensity and the sizes of events are not IID, such as the Hawkes process defined above, are not easily derived. The sufficient condition for stationarity of the Hawkes process given by Daley and Vere-Jones (2005) requires that the expectation of the conditional intensity is constant. This ensures that the conditional intensity does not grow to infinity over time as the conditional intensity is mean reverting. From this condition it can be proven that the Hawkes model with triggering function (5.5) and impact function (4.3) is stationary when

$$\frac{(1+\alpha)\gamma}{\beta} < 1 \tag{3.6}$$

For the proof we refer to Grothe et al. (2014) (Appendix A).

The log-likelihood of the Hawkes model, specified in the conditional intensity and the probability distribution of the sizes of the events $f(m|\boldsymbol{\theta}; \mathcal{H}_t)$ (the derivative of (4.4)), is given by

$$\log L(\boldsymbol{\theta}) = \sum_{n=1}^{N} \log \lambda(t_n|\boldsymbol{\theta}; \mathcal{H}_t) - \int_0^T \lambda(t|\boldsymbol{\theta}; \mathcal{H}_t)dt + \sum_{n=1}^{N} \log f(m_n|\boldsymbol{\theta}; \mathcal{H}_t) \tag{3.7}$$

where t_n are the event arrival times in the interval $[0, T]$. The Hawkes models are estimated by maximizing the likelihood (5.30). Both in the simulation and estimation procedure we assume that at $t = 0$ no events have occurred yet, such that the Hawkes process starts at 0.

[9] We thank an anonymous referee for bringing this issue to our attention.

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

$$\begin{aligned}\Pr\left(N(t_n)-N(t_{n-1})>0\right) &= 1-\Pr\left(N(t_n)-N(t_{n-1})=0\right)\\ &= 1-\exp\left(-\int_{t_{n-1}}^{t_n}\lambda(t|\boldsymbol{\theta};\mathcal{H}_t)dt\right)\end{aligned} \tag{3.8}$$

The tests proposed here can be easily modified when one is interested in specification testing in Hawkes models with a different triggering function, impact function or distribution for the sizes of events.

3.3 LM specification tests

3.3.1 The LM test statistic

We propose a series of specification tests of Hawkes models. We use the approach of Hamilton (1996), who provides specification tests for Markov-switching models. First we develop a general framework for the specification tests. Next, we provide the specifics required to test for omitted explanatory variables, omitted influence of the size of past events on the intensity with which new events are triggered, omitted influence of the history of the event process on the sizes of future events and omitted structural breaks. In each case we derive the score from the extensive version of the model. Then we perform a Lagrange Multiplier test on the null hypothesis that the score is equal to zero. If the null hypothesis is rejected, we should use the full version of the model. Otherwise the more parsimonious version of the model, in which the parameters under study are set equal to zero, should be selected.

Suppose we want to estimate a $(r \times 1)$ vector of parameters $\boldsymbol{\theta}$ by Maximum Likelihood based on a series of N events $y_n = \{t_n, m_n\}$. Consider the distribution of y_n conditional on the values of y for $n-1,\ldots,1$ and on the realizations of a vector of observable exogenous variables $\mathbf{x}_t$, that is

$$p(y_n|\boldsymbol{\theta};\mathcal{H}_t) \tag{3.9}$$

where $\mathcal{H}_t = \{\mathbf{x}_t,\ldots,\mathbf{x}_1,y_{n-1},\ldots,y_1\}$. Our goal is to choose $\boldsymbol{\theta}$ such that the summation of logarithm of (3.9) over all events $1,\ldots,N$, which is equal to (5.30), is maximized.

As we deal with a count process we define the score in event numbers and not in time points. The score of the n-th event is defined as the derivative of the logarithm of the conditional likelihood (3.9) with respect to the parameter vector $\boldsymbol{\theta}$,

$$\mathbf{g}_n(\bar{\boldsymbol{\theta}}) = \left.\frac{\partial \log p(y_n|\boldsymbol{\theta}; \mathcal{H}_t)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} \tag{3.10}$$

which can be evaluated at the true parameter value $\bar{\boldsymbol{\theta}} = \boldsymbol{\theta_0}$ and at the Maximum Likelihood estimate $\bar{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}$.

As (3.9) is a density, it integrates to unity, such that using integration by parts we have

$$\mathrm{E}[\mathbf{g}_n(\boldsymbol{\theta})|\mathcal{H}_t] = \int \frac{\partial \log p(y_n|\boldsymbol{\theta}; \mathcal{H}_t)}{\partial \boldsymbol{\theta}} p(y_n|\boldsymbol{\theta}; \mathcal{H}_t) dy_n = \mathbf{0} \tag{3.11}$$

Therefore the score of an event n is impossible to predict on the basis of the information available at t_{n-1}, if the Hawkes model is correctly specified.

Suppose now the $(r \times 1)$ vector of parameters $\boldsymbol{\theta}$ is estimated subject to the constraint that the first q elements of this vector are zero resulting in $\tilde{\boldsymbol{\theta}}$. The standard errors of $\tilde{\boldsymbol{\theta}}$ are constructed as the diagonal elements of the inverse of the information matrix, which can be approximated by the inverse of the average outer product of the score (Davidson and MacKinnon, 2004),

$$\tilde{\mathbf{W}} = \left[\frac{1}{N}\sum_{n=1}^{N} \mathbf{g}_n(\tilde{\boldsymbol{\theta}})\mathbf{g}_n(\tilde{\boldsymbol{\theta}})'\right]^{-1} \tag{3.12}$$

Using (3.12) $\sqrt{N}\left(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta_0}\right)$ is approximately distributed as $\mathcal{N}\left(\mathbf{0}, \tilde{\mathbf{W}}\right)$.

The validity of the constraints can be assessed by evaluating how much the expected score of events (3.11) increases when the constraints are relaxed. Let $\tilde{\mathbf{g}} = \frac{1}{\sqrt{N}}\sum_{n=1}^{N} \mathbf{g}_n(\tilde{\boldsymbol{\theta}})$. The LM test is based on the following asymptotic relation,

$$\tilde{\mathbf{g}}'\tilde{\mathbf{W}}\tilde{\mathbf{g}} \to \chi^2(q) \tag{3.13}$$

The LM test statistic can be calculated easily as N times the regression of the constant unity on $\mathbf{g}_n(\tilde{\boldsymbol{\theta}})$ (Davidson and MacKinnon, 2004).

When performing the LM test against a one-sided alternative, the test can be modified such that the test in contrast to Karush-Kuhn-Tucker test, still only requires estimation under the null hypothesis and the distribution of the test statistic is not affected. Let $\mathbf{I}_{q|j}$ denote

the $(q \times q)$ identity matrix in which the (i,i)-th elements are zero if i is in the j-th subset of the 2^q subsets of the constrained parameters $\{1,\ldots,q\}$ of which at least one parameter θ_i is tested against the alternative hypothesis that $\theta_i > 0$. Furthermore let $\tilde{\mathbf{V}} = \left(\tilde{\mathbf{W}}_{[1:q;1:q]}\right)^{-1}$ denote the inverse of the $(q \times q)$ submatrix of $(r \times r)$ matrix given in (3.12). Given that

$$\begin{aligned}
\tilde{\mathbf{R}}_{\mathbf{j}} &= \left(\mathrm{I}_{\boldsymbol{q}} - \mathrm{I}_{\boldsymbol{q}|\boldsymbol{j}}\right)\tilde{\mathbf{V}} - \mathrm{I}_{\boldsymbol{q}|\boldsymbol{j}} \\
\tilde{\mathbf{C}}_{\mathbf{j}} &= \left\{x \in \mathbb{R}^q | x = \tilde{\mathbf{R}}_{\mathbf{j}} y, y \in \mathbb{R}^q_{++}\right\} \\
\tilde{\mathbf{Z}} &= \left[\mathrm{I}_{\boldsymbol{q}} \quad \tilde{\mathbf{V}}\tilde{\mathbf{W}}_{[1:q;q+1:r]}\right] \\
\tilde{\mathbf{B}}_{\mathbf{j}} &= \begin{bmatrix} \tilde{\mathbf{V}} - \left(\mathrm{I}_{\boldsymbol{q}|\boldsymbol{j}}\tilde{\mathbf{W}}_{[1:q;1:q]}\mathrm{I}_{\boldsymbol{q}|\boldsymbol{j}}\right)^{+} & \mathbf{0}_{q\times(r-q)} \\ \mathbf{0}_{(r-q)\times q} & \mathbf{0}_{(r-q)\times(r-q)} \end{bmatrix}
\end{aligned}$$

where $(.)^+$ represents the Moore-Penrose inverse, the LM test statistic is computed as follows (Rogers, 1986)

$$\tilde{\mathbf{g}}'\tilde{\mathbf{W}}\tilde{\mathbf{B}}_{\mathbf{j}}\tilde{\mathbf{W}}\tilde{\mathbf{g}} \to \chi^2(q) \qquad \tilde{\mathbf{Z}}\tilde{\mathbf{g}} \in \tilde{\mathbf{C}}_{\mathbf{j}} \tag{3.14}$$

Comparing the LM statistic for one-sided alternatives (3.14) to the LM statistic for two-sided alternatives (3.13), the only difference is that evidence of parameters falling outside the bounded parameter space, that is evidence of parameters being smaller than zero while under the alternative they are assumed to be larger than zero, does not increase the LM statistic.

The expected score (3.11) is zero in the parameters on which no constraints are imposed. This is particularly convenient when the constraints do not affect part of the conditional distribution (3.9). In this case the corresponding part in the score can be neglected. In the remainder let $\boldsymbol{\theta_c}$ and $\boldsymbol{\theta_m}$ denote the vectors with parameters that only affect the conditional intensity or the distribution of the event sizes. For the Hawkes model discussed in section 2, $\boldsymbol{\theta_c} = \{\mu, \gamma, \beta, \alpha\}$ and $\boldsymbol{\theta_m} = \{\xi, \phi\}$ in case $\eta = 0$ and $\boldsymbol{\theta_m} = \{\xi, \phi, \eta, \beta, \alpha\}$ in case $\eta \neq 0$.

We examine models in which an extensive version of the conditional intensity, (4.1), or the distribution of the event sizes, (4.4), is under investigation. In the first class of models, the relevant part of the score can be computed from the derivative of the conditional intensity and the integrated conditional intensity as follows

$$\mathbf{g}_n(\tilde{\boldsymbol{\theta}}_{\boldsymbol{c}}) = -\frac{\partial \int_{t_{n-1}}^{t_n} \lambda(t|\boldsymbol{\theta_c}; \mathcal{H}_t)dt}{\partial \boldsymbol{\theta_c}} + \frac{1}{\lambda(t_n|\boldsymbol{\theta_c}; \mathcal{H}_t)} \frac{\partial \lambda(t_n|\boldsymbol{\theta_c}; \mathcal{H}_t)}{\partial \boldsymbol{\theta_c}}\bigg|_{\boldsymbol{\theta_c}=\tilde{\boldsymbol{\theta}}_{\boldsymbol{c}}} \tag{3.15}$$

The score (3.15) consists of a term which reflects the change in the probability distribution of the time between two consecutive events and a term which reflects the relative change in the rate at which events occur when $\tilde{\boldsymbol{\theta}}_{\boldsymbol{c}}$ is shifted by a infinitesimal amount. By calculating (3.13) or (3.14) we test whether the probability to observe a process of events with event occurrence times $\{t_1, \ldots, t_N\}$, is significantly lower compared to when the conditional intensity is not constrained. In the case parameters are restricted to zero, we assess the added value of these parameters to the model in terms of their contribution to the prediction of the occurrence times of events t_n.

In the models in which an extensive version of the distribution of the event sizes is under investigation, (4.4), the relevant part of the score is equal to the derivative of the logarithm of the probability distribution of the event sizes $f(m|\boldsymbol{\theta}; \mathcal{H}_t)$,

$$\mathrm{g}_n(\tilde{\boldsymbol{\theta}}_{\boldsymbol{m}}) = \left. \frac{\partial \log f(m_n|\boldsymbol{\theta}_{\boldsymbol{m}}; \mathcal{H}_t)}{\partial \boldsymbol{\theta}_{\boldsymbol{m}}} \right|_{\boldsymbol{\theta}_{\boldsymbol{m}}=\tilde{\boldsymbol{\theta}}_{\boldsymbol{m}}} \tag{3.16}$$

The score (3.16) reflects the change in probability to observe an event of a certain magnitude when $\tilde{\boldsymbol{\theta}}_{\boldsymbol{m}}$ is shifted by a infinitesimal amount. With the LM test we evaluate the difference in probability between the occurrence of the event process with magnitudes $\{m_1, \ldots, m_N\}$ under the constrained size distribution and the unconstrained size distribution. An LM test on a constrained parameter vector in which parameters are restricted to zero, indicates whether the zero-restricted parameters provide any information on the magnitude of future events m_n

In our simulations we compare the performance of the LM test to the performance of the Likelihood Ratio test. The LR test statistic for testing $\mathcal{H}_0 : \boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$ against the alternative hypothesis $\mathcal{H}_A : \boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, in which the r_0 restrictions are not imposed upon the parameter vector $\boldsymbol{\theta}$, is given by

$$LR(\boldsymbol{\theta}) = -2 \left(\log L(\tilde{\boldsymbol{\theta}}) - \log L(\hat{\boldsymbol{\theta}}) \right) \tag{3.17}$$

Like the LM test statistic, the LR test statistic (3.17) has a $\chi^2(r_0)$-distribution.

Applying the LR test has as a disadvantage that the Hawkes model has to be estimated under the null and the alternative hypotheses. Estimation of the Hawkes model under the alternative is in general computationally quite demanding.

Omitted explanatory variables

Suppose we want to test the Hawkes model (4.1) against an alternative in which exogenous variables can help to predict occurrence times of events. When the conditional rate at which events arrive scales with variables, such that the variables affect the unconditional intensity regardless of the history of the event process, the conditional intensity has the form

$$\lambda(t|\boldsymbol{\theta};\mathcal{H}_t) = \mu + \mathbf{x}_t'\boldsymbol{\delta} + \sum_{k:t_k<t} g(t-t_k, m_k) \tag{3.18}$$

where $\boldsymbol{\delta}$ denotes the effect of the explanatory variables $\mathbf{x}_t$ or transformations of these variables on the conditional intensity $\lambda(t|\boldsymbol{\theta};\mathcal{H}_t)$.

The former section describes how to implement the LM test for the alternative hypothesis (3.18) against the null hypothesis (4.1), in which $\boldsymbol{\delta}$ equals zero. The expected score requires the computation of the following equations for the derivative of the conditional intensity and the integrated conditional intensity

$$\frac{\partial\lambda(t_n|\boldsymbol{\theta};\mathcal{H}_t)}{\partial\delta} = x_{t_n} \tag{3.19}$$

$$\frac{\partial\int_{t_{n-1}}^{t_n}\lambda(t|\boldsymbol{\theta};\mathcal{H}_t)dt}{\partial\delta} = \sum_{t_{n-1}}^{t_n} x_t \tag{3.20}$$

Omitted predictability of event sizes

When we observe events that are more extreme in magnitude when the event rate is high, this leads to the suspicion that the sizes of events are predictable. In case we want to test this hypothesis in the Hawkes framework (4.1), where the event sizes follow the General Pareto Distribution (4.4), this corresponds to testing whether η is larger than zero.

When we define

$$\tilde{\lambda}(t|\boldsymbol{\theta};\mathcal{H}_t) = \frac{\lambda(t|\boldsymbol{\theta};\mathcal{H}_t)-\mu}{\gamma} = \sum_{k:t_k<t} e^{-\beta(t-t_k)}c(m_k) \tag{3.21}$$

we can express the probability distribution of the event sizes as

$$f(m|\boldsymbol{\theta};\mathcal{H}_t) = \frac{1}{\phi+\eta\tilde{\lambda}(t|\boldsymbol{\theta};\mathcal{H}_t)}\left(1+\frac{\xi(m-u)}{\phi+\eta\tilde{\lambda}(t|\boldsymbol{\theta};\mathcal{H}_t)}\right)^{-\frac{1}{\xi}-1} \tag{3.22}$$

In case the sizes of events do not influence the conditional intensity($\alpha = 0$), the expected score for the predictability of event sizes equals the derivative of the probability distribution of the event sizes (3.22) with respect to η for $\eta = 0$, that is

$$\left.\frac{\partial \log f(m_n|\boldsymbol{\theta}; \mathcal{H}_t)}{\partial \eta}\right|_{\eta=0} = \frac{\tilde{\lambda}_\eta \left((m-u) - \phi\right)}{\phi\left(\phi + \xi(m_n - u)\right)} \tag{3.23}$$

where $\tilde{\lambda}_\eta = \tilde{\lambda}(t|\boldsymbol{\theta}; \mathcal{H}_t)\big|_{\eta=0}$, that is (3.21) in which η is set to zero.

In case the sizes of events do influence the conditional intensity($\alpha = 0$), one has to account for the effect of the history of the event process on the impact that the sizes of events have on the conditional intensity. This causes that the score of η given by (3.23) has to be extended with (3.15) in which $\boldsymbol{\theta_c} = \eta$, which can be derived from

$$\left.\frac{\partial \lambda(t_n|\boldsymbol{\theta}; \mathcal{H}_t)}{\partial \eta}\right|_{\eta=0} = -\sum_{k:t_k<t_n} \gamma\alpha e^{-\beta(t_n - t_k)} \frac{\tilde{\lambda}_\eta(m-u)}{\phi\left(\phi + \xi(m_n - u)\right)} \tag{3.24}$$

$$\left.\frac{\partial \int_{t_{n-1}}^{t_n} \lambda(t|\boldsymbol{\theta}; \mathcal{H}_t)dt}{\partial \eta}\right|_{\eta=0} = -\sum_{k:t_k<t_n} \frac{\gamma\alpha}{\beta} e^{-\beta(t_n - t_k)} \left(e^{\beta(t_n - t_{n-1})} - 1\right) \frac{\tilde{\lambda}_\eta(m-u)}{\phi\left(\phi + \xi(m_n - u)\right)} \tag{3.25}$$

Omitted impact of sizes events

The sizes of events influence the conditional intensity of the Hawkes process (4.1), when larger events trigger more events than smaller events. Under the assumption of (4.3) for the impact of the sizes of events on the event rate, this implies a α-parameter larger than zero.

When the sizes of events are unpredictable ($\eta = 0$), the expected score for α can be computed using following equations for the derivative of the conditional intensity and the integrated conditional intensity

$$\left.\frac{\partial \lambda(t_n|\boldsymbol{\theta}; \mathcal{H}_t)}{\partial \alpha}\right|_{\alpha=0} = \sum_{k:t_k<t_n} \frac{\gamma}{\xi} e^{-\beta(t_n - t_k)} \log\left(1 + \xi\frac{m_n - u}{\phi}\right) \tag{3.26}$$

$$\left.\frac{\partial \int_{t_{n-1}}^{t_n} \lambda(t|\boldsymbol{\theta}; \mathcal{H}_t)dt}{\partial \alpha}\right|_{\alpha=0} = \sum_{k:t_k<t_n} \frac{\gamma}{\beta\xi} e^{-\beta(t_n - t_k)} \left(e^{\beta(t_n - t_{n-1})} - 1\right) \log\left(1 + \xi\frac{m_n - u}{\phi}\right) \tag{3.27}$$

When the sizes of events are predictable ($\eta \neq 0$) one has to account for the impact the sizes of past events have on the probability distribution of the sizes of future events. Another

term has to be added to the score (3.15) of α. This term, given by (3.16), equals the derivative of (3.22) with respect to α,

$$\frac{\partial \log f(m_n|\boldsymbol{\theta};\mathcal{H}_t)}{\partial \alpha}\bigg|_{\alpha=0} = \frac{\eta\left((m-u)-(\phi+\eta\tilde{\lambda}_\alpha)\right)}{\gamma(\phi+\eta\tilde{\lambda}_\alpha)\left((\phi+\eta\tilde{\lambda}_\alpha+\xi(m_n-u)\right)}\frac{\partial\lambda(t_n|\boldsymbol{\theta};\mathcal{H}_t)}{\partial\alpha} \tag{3.28}$$

where $\lambda_\alpha = \tilde{\lambda}(t|\boldsymbol{\theta};\mathcal{H}_t)\big|_{\alpha=0}$ with $\tilde{\lambda}(t|\boldsymbol{\theta};\mathcal{H}_t)$ given in (3.21).

Omitted structural break

Suppose we want to test the Hawkes model (4.1) for a structural change in the constant unconditional intensity when we do know the time of this change. For this purpose we can use of the methods of Andrews (1993). In Section 3.1.1 on omitted explanatory variables, let x_t in (3.18) be a scalar that is equal to zero for $t < \tau$ and equal to unity when $t \geq \tau$. The constructed LM statistic (3.13) tests the null hypothesis that the data is accurately modeled by the simple version of the Hawkes model (4.1) against the alternative that there is a shift in the long-term background component of the intensity, μ, at date τ not captured by the simple model. One can calculate the LM statistic for all τ, however usually only values between $0.15N$ and $0.85N$ are considered. Here N corresponds to the number of events in the data under study. The asymptotic distribution of the value of τ that produces the largest LM statistic, has critical values corresponding to the π_0, $p = 1$ entry in Table 1 of Andrews (1993).

3.3.2 Tests for dependence between series

LM test

In this subsection we describe the LM test of our focal intent.

In case events across series tend to arrive around the same time, the occurrence of an event in one series could be increasing the probability that an event in another series arrives. A Hawkes model in which the conditional intensity of a series i is affected by the history of the event process of another series j is given by

$$\lambda_i(t|\boldsymbol{\theta};\mathcal{H}_t) = \mu + \gamma_i \sum_{k:t_k<t} g(t-t_k,m_k) + \gamma_j \sum_{l:t_l<t} h(t-t_l,m_l) \tag{3.29}$$

where $\gamma_j h(t - t_l, m_l)$ presents the effect of an event in series j on the intensity at time t in series i.

Equation (4.8) can easily be extended to include possible cross-triggering effects of more than one series, $j = \{1, \ldots, d\}$,

$$\lambda_i(t|\boldsymbol{\theta}; \mathcal{H}_t) = \mu + \Gamma_{ii} \sum_{k:t_{ik}<t} g_i(t - t_{ik}, m_{ik}) + \sum_{j=1, j\neq i}^{d} \Gamma_{ij} \sum_{k:t_{jk}<t} h_j(t - t_{jk}, m_{jk}) \tag{3.30}$$

where $\Gamma_{ij} h_j(t - t_{jk}, m_{jk})$ presents the effect of an event k in series j at time t_{jk} on the intensity at time t in series i. Given that c_i, f_i, represent respectively the impact of the event sizes on the triggering intensity and the probability distribution of the event sizes for a series i, we have

$$\frac{1}{1 + \alpha_i} \int_{u_i}^{\infty} c_i(m) f_i(m) dm = 1 \tag{3.31}$$

$$\beta_i \int_0^{\infty} e^{-\beta_i t} dt = 1 \tag{3.32}$$

where u_i is the threshold above which events are considered and α_i, β_i, are parameters of the model indicating respectively the strength of the influence of the event sizes on the triggering intensity and the decay speed for the series i. Given (3.31) and (3.32), stationarity for the multivariate Hawkes model can be proven whenever the (sufficient) conditions indicated by Liniger (2009) and Embrechts et al. (2011), are satisfied, that is whenever

$$Spr(\mathbf{Q}) < 1 \tag{3.33}$$

$$\beta_i \int_0^{\infty} t e^{-\beta_i t} dt < \infty \tag{3.34}$$

Here $Spr(\mathbf{Q})$ represents the spectral radius of the branching matrix $\mathbf{Q}$, which consists of the elements $\frac{(1+\alpha_j)\Gamma_{ij}}{\beta_j}$.

Specification (3.30) enables one to separately test for amplification effects from each of the series of $j = \{1, \ldots, d\}$ to the conditional intensity of a series i in one go. This attractive feature of the LM test originates from the fact that to compute the LM test no estimation under the alternative hypothesis is required. Let $x_j(t) = \sum_{k:t_{jk}<t} h_j(t - t_{jk}, m_{jk})$, denote the additional tension caused by all events k in a series j before time t. To test whether the cross-

triggering effects are significant for a series i, we perform the LM test on the coefficients of the explanatory variables $x_j(t)$ designed in the former section. First we estimate the Hawkes model under the null hypothesis of no cross-triggering effects, that is $\boldsymbol{\Gamma_{i,-i}} = \boldsymbol{\Gamma_i} \setminus \Gamma_{ii} = \mathbf{0}$, against the alternative hypothesis of cross-excitation $\boldsymbol{\Gamma_{i,-i}} > \mathbf{0}$. Hereafter we use (3.15) to compute the expected score from the derivative of the conditional intensity and the integrated conditional intensity

$$\frac{\partial \lambda(t_n|\boldsymbol{\theta}; \mathcal{H}_t)}{\partial \boldsymbol{\Gamma_{i,-i}}} = \mathbf{x}(t) \tag{3.35}$$

$$\frac{\partial \int_{t_{n-1}}^{t_n} \lambda(t|\boldsymbol{\theta}; \mathcal{H}_t) dt}{\partial \boldsymbol{\Gamma_{i,-i}}} = \int_{t_{n-1}}^{t_n} \mathbf{x}(t) dt \tag{3.36}$$

When there are $d \geq 1$ series involved, the LM test (3.14) is asymptotically $\chi^2(d-1)$ under the null of no cross-triggering effects for a series i.

The cross-triggering functions can be specified in various ways. For example

$$h_j(s, m) = e^{-\beta_j s} \tag{3.37}$$

Using cross-triggering functions of the form (3.37) extreme events in the series i can have a longer-lasting effect on triggering of subsequent events than extreme events in other series have on the triggering of events in i. However as the β_j-parameters are unidentified under the null hypothesis $\mathcal{H}_0 : \boldsymbol{\Gamma_{i,-i}} = \mathbf{0}$, the regular critical values of the χ^2-distribution cannot be used. This problem is recognized by Bowsher (2007). In Bowsher (2002) a solution is given to overcome this problem. The test statistic $T^{(m)}(\theta)$ is minimized with respect to the parameters θ. When

$$T^{(m)} = \min_{\theta \in \Theta} T^{(m)}(\theta) \tag{3.38}$$

exceeds the $(1-\alpha)$-quantile of the null distribution $T^{(m)}(\theta_0)$, where θ_0 is the true parameter value, the null hypothesis is rejected. The probability of falsely rejecting the null hypothesis bounded by α as $T^{(m)} \leq T^{(m)}(\theta_0)$. As the approach does not depend on regularly conditions and it is valid when θ_0 is on the boundary of the parameter space, the procedure is suitable to use for testing in Hawkes models. Other methods (Hansen, 1996; Andrews, 2001) can also be considered.

One could also restrict the decay rate of the effect of past events in all series to be the same. The cross-triggering functions are in this case specified as

$$h_j(s, m) = e^{-\beta_i s} \tag{3.39}$$

where the parameter β_i is equal to the β-parameter in the triggering function by which the effect of past events on subsequent events in the same series is described, $g(s, m)$. As under the null hypothesis $\mathcal{H}_0 : \mathbf{\Gamma_{i,-i}} = \mathbf{0}$, β_i still determines the decay rate of the triggering effect of events that occurred in the series i, we do not encounter the problem of unidentified nuisance parameters.

Aït-Sahalia et al. (2014) and Aït-Sahalia et al. (2015) define the intensity of new jumps $(J_{i,t})$ in series i at time t recursively as

$$d\lambda_{i,t} = \beta(\lambda_{i,\infty} - \lambda_{i,t})dt + \sum_{j=1}^{d} \Gamma_{ij} dJ_{j,t} \tag{3.40}$$

where $\lambda_{i,t}$ is mean reverting to λ_∞. As in (3.30) the intensity at which events occur in series i is amplified when an event in one of the series arrives. The amplification effect decays at rate β. Testing this specification for the presence of cross-excitation comes down to testing the former specification (3.30) with (3.39) for the presence of cross-excitation.

It is also possible to employ one of the following cross-triggering functions

$$h_j(s, m) = e^{-\beta_j s} c(m) \tag{3.41}$$

$$h_j(s, m) = e^{-\beta_i s} c(m) \tag{3.42}$$

where $c(m)$ represents the impact of the sizes of past events in the series j on the intensity with which future events are triggered a series i. Using (3.41) or (3.42), the effect of event occurrences in other series on the event rate in the series i depends on their magnitude. The impact of past events on the triggering of future events in series i differs across series when the parameters of the impact function are not treated as equal across series. In that case there are additional α_j, ξ_j, ϕ_j and η_j parameters unidentified under the null hypotheses.

Regression method

Van Oordt and Zhou (2012) propose a regression method for testing cross-sectional dependence. Their method consists of regressing the indicators of the occurrences of extreme events in the different time series at time t on lagged (products) of indicators of extreme events in these series. When considering one lag and d time series, the regression can be expressed as follows

$$\mathrm{I}_{t,i} = \sum_{j=1}^{d} \beta_j \mathrm{I}_{t-1,j} + \sum_{k>j} \beta_{j,k} \mathrm{I}_{t-1,j} \mathrm{I}_{t-1,k} + \sum_{l>j,k} \beta_{j,k,l} \mathrm{I}_{t-1,j} \mathrm{I}_{t-1,k} \mathrm{I}_{t-1,l} + \ldots + \prod_{x=1}^{d} \beta_{1,\ldots,d} \mathrm{I}_{t-1,x} \tag{3.43}$$

where $\mathrm{I}_{t,i}$ is the indicator function, which is one when an event arrives in market i at time t and zero otherwise. Equation (3.43) can easily be modified to include more lagged terms or indicators that identify the occurrence of events over a period of time in the past.

The β-coefficients and their standard deviations are easily estimated by Ordinary Least Squares (OLS). By performing a Student's t-test on the β coefficients, one can deduct whether the occurrence of extreme events in one or more series in the past contains 'significant information' on the occurrence of extreme events in another series in the future.

The focus is on detecting whether or not event occurrences in other series have explanatory power for the unconditional event rate in a series i. The aim of the Hawkes models is to model the conditional event rate in a series i, possibly using the event occurrences in other series as explanatory variables. In the Hawkes models events have the effect on the event rate that decays as a function of time, while the regression (3.43) describes a stepwise decay. Therefore the method of Van Oordt and Zhou (2012) does not fit into the Hawkes framework. However the method of Van Oordt and Zhou (2012) is meaningful in the Hawkes framework if the conditional and unconditional event rate are approximately equal. This is the case when the self-exciting part of the event rate is small relative to the constant part of the event rate, that is when the parameters μ and/or β are high and/or the parameter γ_i is low relative to the other parameters.

Orthogonality test

Testing for orthogonality means testing for cross-sectional dependence between the occurrence of extreme events at time t conditional on the information set $\mathcal{H}_{t-1}$, that is, the entire

history of the event process up to time t. When event processes are orthogonal their event rates at time t can be related through events that occurred in the past, however innovations to these event rates are not related. Therefore, the occurrence of multiple events at time $t+1$ is as likely as if the processes would be independent.

In some multivariate Hawkes models the event processes are not orthogonal. For example Grothe et al. (2014) combine the intensities, with which events in the margins arrive, in a copula function to derive the joint intensity with which in at least one margin an event arrives. As the joint intensity is smaller than the sum of the marginal intensities, the probability of events occurring simultaneously is larger than zero. Bormetti et al. (2015) let the marginal intensities at time t dependent on the intensity of a latent factor at time t. The intensities differ as their sensitivity to the intensity of this latent factor at time t differs. However the intensities share similar dynamics, which makes the occurrence of events at the same time more likely than in a model in which event processes are orthogonal.

Let $\left\{\tilde{t}_1, \ldots, \tilde{t}_N\right\}$ denote the combined event process which indicates the occurrence of one or more events. The orthogonality test described in Lando and Nielsen (2010) is based on the fact that in case of orthogonality, the process, $\left\{\tilde{t}_1, \ldots, \tilde{t}_N\right\}$, is generated by the sum of the univariate conditional intensities $\lambda_i(t|\boldsymbol{\theta}_i; \mathcal{H}_{i,t})$, $i \in \{1, \ldots, d\}$. The test prescribes to use residual analysis technique of Ogata (1988) on the marginal intensities as if they were independent, to compute the transformed times $\{\tilde{\tau}_n\}$, $1, \ldots, N$,

$$\tilde{\tau}_n = \int_0^{\tilde{t}_n} \sum_{i=1}^{d} \lambda_i(t|\boldsymbol{\theta}_i; \mathcal{H}_{i,t})dt = \sum_{i=1}^{d} \int_0^{\tilde{t}_n} \lambda_i(t|\boldsymbol{\theta}_i; \mathcal{H}_{i,t})dt \tag{3.44}$$

The test states that if the event processes are orthogonal the transformed times are distributed according to a homogeneous Poisson process with intensity 1. Also the transformed interarrival times,

$$\tilde{\tau}_n - \tilde{\tau}_{n-1} = \int_{\tilde{t}_{n-1}}^{\tilde{t}_n} \sum_{i=1}^{d} \lambda_i(t|\boldsymbol{\theta}_i; \mathcal{H}_{i,t})dt = \sum_{i=1}^{d} \int_{\tilde{t}_{n-1}}^{\tilde{t}_n} \lambda_i(t|\boldsymbol{\theta}_i; \mathcal{H}_{i,t})dt \tag{3.45}$$

are independent exponential random variables with mean 1. In order to verify whether $\{\tilde{\tau}_n\}$ is Poisson with unit intensity, we perform the Kolmogorov-Smirnov (KS) test.

When the probability of an extreme event arriving in one time series is influenced by extreme events that occurred in another series in the past, this time series is conditionally

dependent on the other series. When event processes are conditionally dependent, it is possible that the probability of events occurring simultaneously, is still zero. These processes are independent of each other when conditioned on all past information and thus orthogonal. Therefore, if event processes are orthogonal, this does not imply that these processes are conditionally independent. Examples of multivariate Hawkes processes that are orthogonal but not conditionally independent can be found in Bowsher (2007), Embrechts et al. (2011), Aït-Sahalia et al. (2014) and Aït-Sahalia et al. (2015).

3.4 Monte Carlo analysis of tests

Here we review the quality of the tests for detecting the dependence between series that follow a Hawkes process. We evaluate the power and size of the LM test, the LR test, the regression method and the orthogonality test, by means of a Monte Carlo analysis. We simulate from the Hawkes models under the null hypothesis of zero cross-excitation and under the alternative hypothesis, where we assume that at the beginning no events are present yet. Under the alternative hypothesis the occurrence of events in one series amplifies the probability that events arrive in another series. A Monte Carlo analysis of the tests for the omitted predictability of event sizes and the omitted impact of the sizes of past events on the probability that events arrive in the future, is attached in the appendix.

The considered Hawkes models differ as the parameters α and/or η are restricted to zero in some of the models, while in other models these parameters can take values different from zero. When $\alpha = 0$ the sizes of former events do not influence the conditional intensity of new event arrivals, while when $\alpha \neq 0$ the probability that events are triggered is affected by the sizes of previous events. When $\eta = 0$ the history of the event process does not have information on the sizes of events in the future, while when $\eta \neq 0$ the sizes of events can be predicted from the past.

For the simulation exercise we use parameter values derived from estimating the Hawkes models under the null and the alternative hypothesis using the data of Grothe et al. (2014). Grothe et al. (2014) consider extreme negative returns in European and US financial markets, which are approximated by daily log-returns of the MSCI-USA and MSCI-EU indices over the period January 1, 1990 to January 12, 2012. To estimate the models the threshold is set

at the 97.7% quantile of the empirical distributions of the stock market returns. Models are estimated by numerically maximizing their log-likelihood using the interior-point algorithm.

Table 3.1 presents the parameter estimates under the null of zero cross-excitation and the alternative which incorporates spillover effects between series. The parameter Γ_{12} (Γ_{21}) controls the cross-triggering of events in the US (EU) by the occurrence of events in the EU (US). Amongst the multivariate Hawkes models the instantaneous effect of the occurrence of events in the EU on the event rate in the US (Γ_{12}), is between 0.68 and 0.98 times as large as the effect of events in the series itself (Γ_{11}) in the Hawkes models in which the sizes of events do not influence the event rate. For the EU, the ratio between the instantaneous cross-excitation and self-excitation effects, ranges between 0.47 and 0.74. In the Hawkes models in which the event sizes do affect the event rate the same ranges of ratios are found when the effect of events of the minimum magnitude is examined. Furthermore the standard errors of the parameters that control the cross-excitation effects (Γ_{12} and Γ_{21}) are comparable in size to those of the self-exciting parameters (Γ_{11} and Γ_{22}). Therefore it seems that the impact of events in a series on the event rate in another series is quite substantial and should not be ignored. The parameters that control the self-excitation effects and the constant part of the event rate are higher when the models are estimated under the null. Hence, the impact of events on the event rate of the series itself and the number of spontaneously occurring events seem overestimated in the univariate models, accounting for self-excitation as well as the back and forth cross-excitation between series.

To quantify the effect of crashes in one market to another market, we examine the branching ratio of crashes in the markets. The branching ratio is the expected number of events triggered by an event. We compute the ratio by integrating the self-exciting part of the event rate from zero to infinity over time and size dimension. In the univariate case the branching ratio is given by (3.6). Amongst the Hawkes models, a US crash triggers on average 0.38 to 0.46 EU crashes, which is approximately equal to range of the average number of crashes provoked by EU crashes themselves $(0.36, 0.46)$. When the EU market crashes, this leads on average to 0.21 to 0.28 US crashes amongst the Hawkes models, while the expected number of US crashes generated by a US crash varies between 0.46 to 0.56 in the Hawkes models. Thus, crashes in the US market have a substantial effect on the occurrence of crashes in the EU market. The effect is approximately as large as the effect of crashes in the EU market. The effect of crashes in the EU market on the occurrence of crashes in the US market not

so large as the cross-triggering effect vice versa. However, this cross-triggering effect seems also quite substantial as one US crash is expected after the occurrence of about 3 tot 5 EU crashes.

Comparing the Hawkes models with the parameter restriction $\alpha = 0$ to the Hawkes models without this restriction, the magnitude of an extreme has a significant positive influence on the probability of triggering another extreme for both sets of returns. This means that on average larger events trigger more events than smaller events. Also the estimates for η are positive and significant for both sets of returns. This suggests a model which incorporates the history of the event process to prospect the sizes of subsequent events, matches the crashes closer than a model which assumes the sizes of events are independent of the past. When $\eta > 0$, the mean and variance of the distribution of the excess magnitudes of the events scale with the value of the cumulative triggering function, and thus the probability of the arrival of an event triggered by another event. This means that on average larger extremes are observed after the occurrence of more and/or big events than after a tranquil period.

We test for the existence of cross-dependence among the simulated series. The Figures 3.1 and 3.2 display Hawkes processes simulated from respectively models under the null hypothesis of no cross-excitation and models under the alternative hypothesis of cross-excitation. The magnitude-time plots show that the simulated series share the major feature characteristic of the Hawkes models, that is the clustering of events over time. However, looking at Figure 3.1, different series simulated from a model in which the occurrence of events in one series has no influence on the occurrence of events in another series, do not necessarily exhibit clustering behavior at the same time. This is further illustrated by the intensity plots as the event rates of the series do not spike at the same time. In the case of spillover effects between series, events cluster not only through time but also in the cross section. Therefore one can see that the derived conditional intensities of the different series get high and low simultaneously.

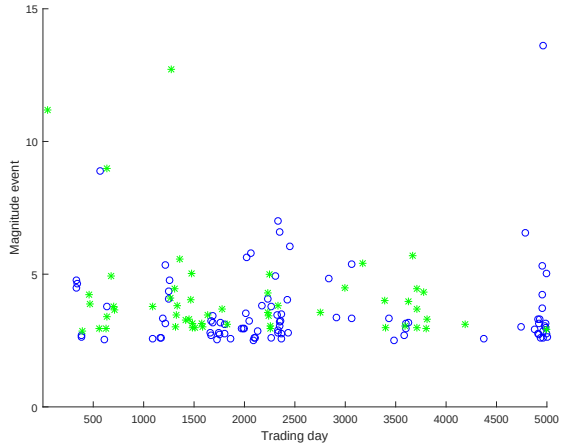
We simulate $1,000$ series from the Hawkes models under the null hypothesis of zero cross-excitation and the alternative over a period of $5,000$ and of $10,000$ days. The number of generated events differs among the simulated series, but in expectation it is equal to 2.3% of the number of days over which we simulate. The simulation procedure can be found in the appendix. In the following the parameters Γ_{12} and Γ_{21} control the cross-triggering effects from one series to another. In the case the occurrence of an event in series 1 (2) amplifies

Table 3.1: Parameter estimates used in simulation

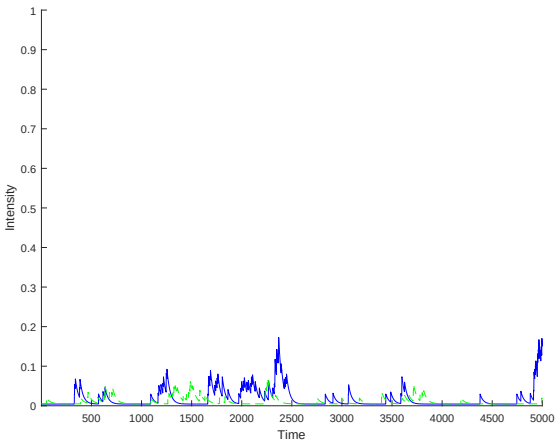
	Zero cross-excitation				Cross-excitation			
	$\alpha = 0$		$\alpha \neq 0$		$\alpha = 0$		$\alpha \neq 0$	
	$\eta = 0$	$\eta \neq 0$	$\eta = 0$	$\eta \neq 0$	$\eta = 0$	$\eta \neq 0$	$\eta = 0$	$\eta \neq 0$
Γ_{11}	0.0290	0.0333	0.0205	0.0164	0.0198	0.0216	0.0139	0.0115
	(0.0076)	(0.0074)	(0.0177)	(0.0117)	(0.0071)	(0.0071)	(0.0056)	(0.0083)
Γ_{12}					0.0135	0.0212	0.0099	0.0104
					(0.0058)	(0.0085)	(0.0046)	(0.0077)
Γ_{21}					0.0188	0.0265	0.0143	0.0072
					(0.0090)	(0.0097)	(0.0080)	(0.0090)
Γ_{22}	0.0447	0.0455	0.0321	0.0167	0.0400	0.0360	0.0283	0.0109
	(0.0089)	(0.0080)	(0.0196)	(0.0045)	(0.0123)	(0.0124)	(0.0119)	(0.0305)
ξ_1	0.2886	0.1614	0.2884	0.1670	0.2886	0.1302	0.2894	0.1465
	(0.0559)	(0.1074)	(0.1372)	(0.0435)	(0.1345)	(0.1053)	(0.1378)	(0.1332)
ξ_2	0.2364	0.1635	0.2455	0.1868	0.2364	0.1293	0.2496	0.1625
	(0.1248)	(0.0817)	(0.4475)	(0.0985)	(0.1973)	(0.0818)	(0.1095)	(0.3077)
ϕ_1	0.8501	0.4599	0.8503	0.4701	0.8501	0.3908	0.8495	0.4075
	(0.0387)	(0.1024)	(0.0281)	(1.2899)	(0.1387)	(0.0776)	(0.2117)	(0.1320)
ϕ_2	0.8867	0.4479	0.8787	0.4293	0.8867	0.4078	0.8750	0.3936
	(0.0543)	(0.2405)	(1.4997)	(0.1091)	(0.1498)	(0.1082)	(0.2363)	(0.5246)
μ_1	0.0059	0.0063	0.0065	0.0063	0.0053	0.0061	0.0056	0.0057
	(0.0015)	(0.0015)	(0.0037)	(0.0067)	(0.0015)	(0.0015)	(0.0017)	(0.0113)
μ_2	0.0055	0.0055	0.0057	0.0057	0.0038	0.0041	0.0041	0.0039
	(0.0013)	(0.0012)	(0.0178)	(0.0013)	(0.0012)	(0.0012)	(0.0013)	(0.0115)
β_1	0.0389	0.0458	0.0411	0.0403	0.0357	0.0463	0.0338	0.0359
	(0.0087)	(0.0105)	(0.0346)	(0.1528)	(0.0117)	(0.0116)	(0.0122)	(0.0294)
β_2	0.0582	0.0594	0.0600	0.0579	0.0871	0.0979	0.0970	0.0987
	(0.0134)	(0.0105)	(0.1916)	(0.0134)	(0.0421)	(0.0255)	(0.0292)	(0.0411)
α_1			0.4416	0.7921			0.3024	0.5626
			(0.4722)	(1.4302)			(0.2123)	(1.0501)
α_2			0.4107	1.6228			0.5103	2.4323
			(0.3305)	(0.3405)			(0.1638)	(5.2226)
η_1		0.2597		0.1146		0.1979		0.0914
		(0.0803)		(0.1080)		(0.0671)		(0.0372)
η_2		0.3164		0.1051		0.2582		0.0712
		(0.0947)		(0.0374)		(0.0944)		(0.0843)

The parameters estimates are derived applying the Hawkes models to daily log-returns of the MSCI-USA and MSCI-EU indices over the period January 1, 1990 to January 12, 2012. In the Hawkes models that enable cross-excitation, the parameters Γ_{12} and Γ_{21} control the cross-triggering effect from the EU to the US and the cross-triggering effect from the US to the EU. In the Hawkes models with the parameter restriction $\alpha = 0$, the magnitude of events have no influence on the triggering of subsequent events. In the Hawkes models with the parameter restriction $\eta = 0$, the history of the events has no influence on the magnitude of subsequent events. Standard deviations are shown in between parentheses.

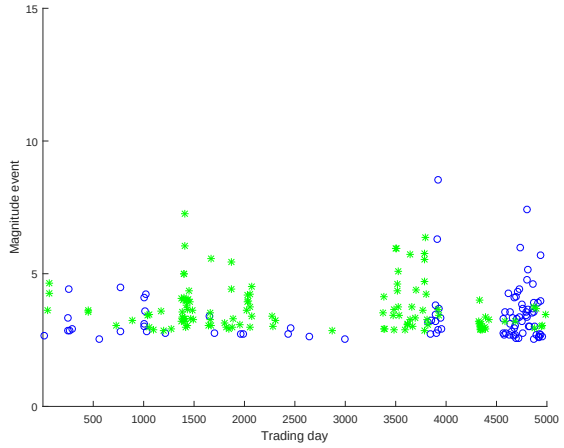
Figure 3.1: Simulation Hawkes models without cross-excitation



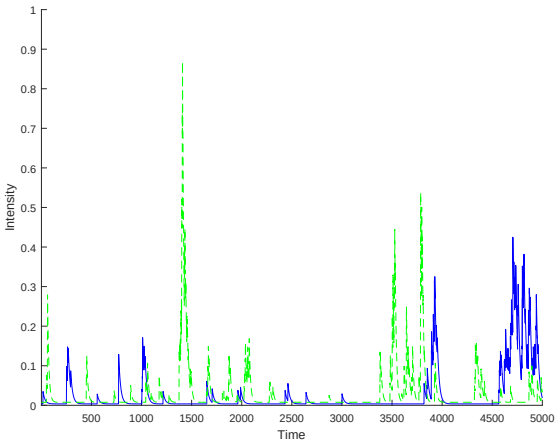
(a) $\alpha = 0, \eta = 0$



(b) $\alpha = 0, \eta = 0$



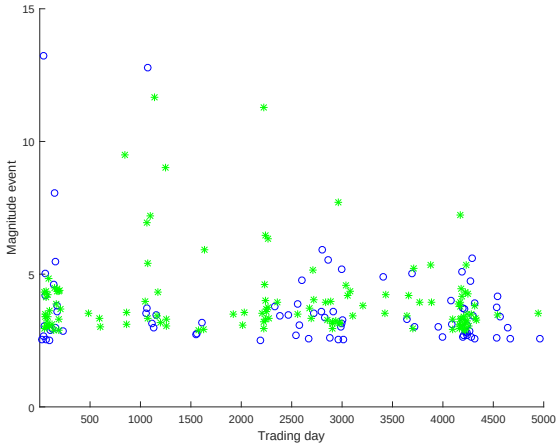
(c) $\alpha \neq 0, \eta \neq 0$



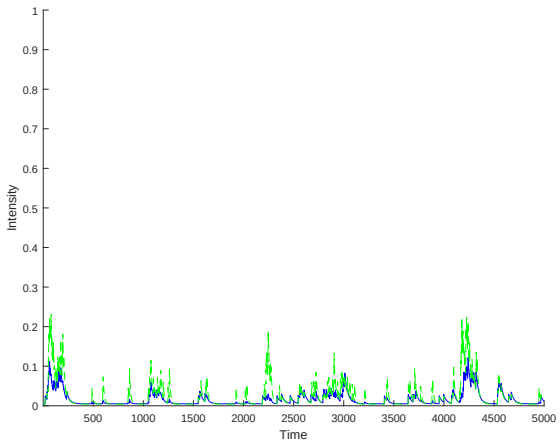
(d) $\alpha \neq 0, \eta \neq 0$

Simulation in discrete time of genuine multivariate Hawkes models without cross-excitation. Parameters and the minimum magnitude of events under consideration are set on the models estimated from the data set of Grothe et al. (2014). In Hawkes models with the parameter restriction $\alpha = 0$, the magnitude of events have no influence on the triggering of subsequent events. In the Hawkes models with the parameter restriction $\eta = 0$, the history of the events has no influence on the magnitude of subsequent events. Magnitudes and times of the events and plots of the conditional intensity are presented for the two series. The blue solid lines and circles correspond with the first series, the green dashed lines and stars correspond with the second.

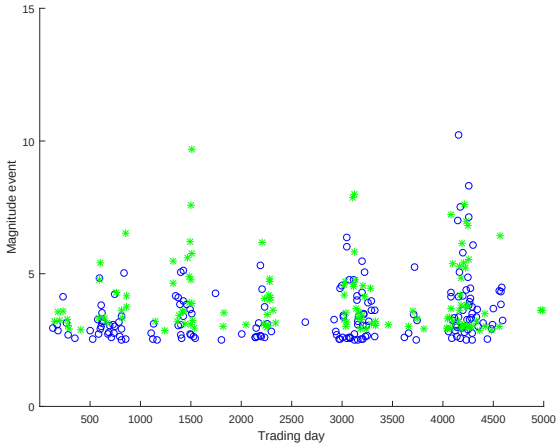
Figure 3.2: Simulation Hawkes models with cross-excitation



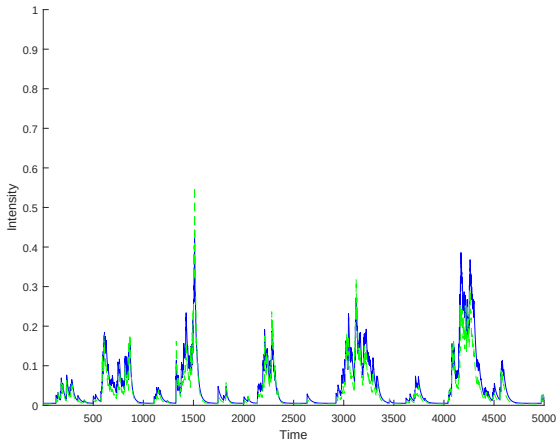
(a) $\alpha = 0, \eta = 0$



(b) $\alpha = 0, \eta = 0$



(c) $\alpha \neq 0, \eta \neq 0$



(d) $\alpha \neq 0, \eta \neq 0$

Simulation in discrete time of genuine multivariate Hawkes models with cross-excitation. Parameters and the minimum magnitude of events under consideration are set on the models estimated from the data set of Grothe et al. (2014). In Hawkes models with the parameter restriction $\alpha = 0$, the magnitude of events have no influence on the triggering of subsequent events. In the Hawkes models with the parameter restriction $\eta = 0$, the history of the events has no influence on the magnitude of subsequent events. Magnitudes and times of the events and plots of the conditional intensity are presented for the two series. The blue solid lines and circles correspond with the first series, the green dashed lines and stars correspond with the second.

the probability that an event arrives in series 2 (1), this is indicated by a parameter Γ_{12} (Γ_{21}), that is different from zero.

Table 3.2 reports the percentage rejections of the null hypothesis of zero cross-excitation at a 5% level when performing the LM test. Under the alternative hypothesis, the events in one series affect the conditional intensity in another series as specified by cross-triggering function (3.39). Using (3.39), the impact of event occurrences among series on the conditional intensity of one series decays at the same rate. Reviewing the size and power of the LM test we conclude the LM test is quite accurate. For $T = 10,000$ the LM test is slightly undersized with percentage rejections that vary between 2.2% and 4.6% over the different Hawkes models. Also the power of the test is high as in 95.0% to 100.0% cases the Hawkes model with zero cross-excitation is rightly rejected. When the time period over which the Hawkes models are simulated and estimated is short, the quality of the LM test falls only a little bit. This can be explained by the small number of events generated, such that the distribution of the LM test statistic has not entirely converged to the asymptotic $\chi^2(1)$-distribution. For $T = 5,000$, the test is less undersized as the test does not reject the null in 3.4% to 5.6% of the cases. Although the power of the test is a little lower compared to when the LM test is reviewed over the longer time period of $T = 10,000$, the power is still high with percentage rejections that range from 75.1% to 99.0%.

Table 3.2 presents the results of the LM test in which the information matrix is approximated by the average outer product of the score as outlined in the previous section. Results of the LM test in which the information matrix is approximated by the Hessian matrix, that is the negative of the matrix with as elements the average second-order partial derivatives, are presented in the appendix. We prefer to use the average outer product of the score as the LM test in this form does not require the derivation of second derivatives. Also the test, although less undersized, has more power for the Hawkes models than the LM test in which the Hessian is used.

We compare the LM test to the computationally more demanding LR test. The LR test is supposed to outperform the LM test, as the LR test uses information on the likelihood under both the null and the alternative hypothesis, while for the LM test only the first is needed. Table 3.3 reports the percentage rejections of the null hypothesis of zero cross-excitation at a 5% level when carrying out the LR test. The LR test is more undersized than the LM test. Percentage rejections of the null when performing the LR test on simulations from Hawkes

Table 3.2: Size and power LM tests

		T=5,000				T=10,000			
		$\alpha = 0$		$\alpha \neq 0$		$\alpha = 0$		$\alpha \neq 0$	
		$\eta = 0$	$\eta \neq 0$	$\eta = 0$	$\eta \neq 0$	$\eta = 0$	$\eta \neq 0$	$\eta = 0$	$\eta \neq 0$
Size	Γ_{12}	4.5	3.4	4.0	5.6	2.2	3.0	4.6	2.6
	Γ_{21}	3.8	4.0	3.9	4.5	2.5	2.2	2.8	2.5
Power	Γ_{12}	75.1	87.9	74.5	79.2	95.0	99.0	96.1	96.2
	Γ_{21}	96.5	99.0	94.8	98.6	99.9	100.0	100.0	100.0

Percentage rejections of the null hypothesis of zero cross-excitation, when performing the LM test on $1,000$ simulations from Hawkes models over a time period of $5,000$ and $10,000$ time instances using 5% critical values. For the power of the tests the models are simulated under the alternative, while for the size the models are simulated under the null. The parameters and the minimum magnitude of events are set on values estimated from the data set of Grothe et al. (2014). The parameters Γ_{12} and Γ_{21} control respectively the cross-triggering effect from series 2 to series 1 and the cross-triggering effect from series 1 to series 2 (3.30). In the Hawkes models with the parameter restriction $\alpha = 0$, the magnitude of events have no influence on the triggering of subsequent events. In the Hawkes models with the parameter restriction $\eta = 0$, the history of the events has no influence on the magnitude of subsequent events.

models under the null, vary between 1.3% and 3.4%. The LR test is somewhat more powerful than the LM test. When the Hawkes models are simulated under the alternative over $5,000$ days, this results in rejection percentages between 80.6% and 99.4%, when simulated over $10,000$ days rejection percentages range from 96.7% to 100.0%. The rejection percentages of the LR test are not so different from the percentages found when carrying out the LM test, such that we conclude the performance of the LR test is only a bit superior to the performance of the LM test. As the LR test is a much more time consuming and as the LM test performs very well, we prefer to use the LM test.

Table 3.4 reports the percentage rejections of the null hypothesis of zero cross-excitation when applying the regression method on $1,000$ simulations from Hawkes models under the null and the alternative. The regression method is described in Section 3.2.2. Table 3.4 reveals that the regression method cannot be used to detect dependence between series as the method indicates presence of cross-excitation while there is no cross-excitation in at least 49% of the cases amongst models and simulation periods. Section 3.2.2 already mentioned that the method of Van Oordt and Zhou (2012) does not fit into the Hawkes framework as it focuses on the unconditional event rate using a stepwise decay function for the effect of events on the conditional intensity, while in Hawkes models the effect of events on the conditional intensity decays smoothly as a function of time. Here we conclude indeed that

Table 3.3: Size and power LR tests

		T=5,000				T=10,000			
		$\alpha = 0$		$\alpha \neq 0$		$\alpha = 0$		$\alpha \neq 0$	
		$\eta = 0$	$\eta \neq 0$	$\eta = 0$	$\eta \neq 0$	$\eta = 0$	$\eta \neq 0$	$\eta = 0$	$\eta \neq 0$
Size	Γ_{12}	1.5	1.6	2.3	3.1	1.3	2.2	2.9	2.1
	Γ_{21}	2.0	3.4	2.2	1.7	2.0	1.9	2.2	2.2
Power	Γ_{12}	81.9	91.7	80.6	83.7	96.7	99.5	98.2	98.0
	Γ_{21}	97.7	99.4	95.7	99.0	99.9	100.0	100.0	100.0

Percentage rejections of the null hypothesis of zero cross-excitation, when performing the LR test on 1000 simulations from Hawkes models over a time period of $5,000$ and $10,000$ time instances using 5% critical values. The parameters and the minimum magnitude of events are set on values estimated from the data set of Grothe et al. (2014). The parameters Γ_{12} and Γ_{21} control respectively the cross-triggering effect from series 2 to series 1 and the cross-triggering effect from series 1 to series 2 (3.30). In the Hawkes models with the parameter restriction $\alpha = 0$, the magnitude of events have no influence on the triggering of subsequent events. In the Hawkes models with the parameter restriction $\eta = 0$, the history of the events has no influence on the magnitude of subsequent events.

the regression method is a bad approach to examine the presence of spillovers in a Hawkes framework.

Table 3.5 shows the results of the orthogonality test on $1,000$ simulations from Hawkes models under the null and the alternative. The Table reports the percentage rejections of the null hypothesis of zero cross-excitation applying a Kolmogorov-Smirnov test on the transformed times (3.44). A description of the orthogonality test can be found in Section 3.2.3. As explained in Section 3.2.3, the tests are not able to capture conditional dependence such that the power of the orthogonality tests is low.

3.5 Application to Financial Data

To illustrate the convenient use of the LM test for dependence when more than 2 series are involved, we apply the test to 3 stock market indices, a bond market index and 3 exchange rate markets. Hartmann et al. (2004) report that the stock-bond contagion is approximately as frequent as the flight-to-quality phenomenon, where the latter is the outflow of capital from stock markets to bond markets when the first are facing crises periods. Therefore we examine the spillover effects between extreme negative price movements in the equity indices and extreme negative price movements as well as extreme positive price movements in the bond

Table 3.4: Size and power regression method

			T=5,000				T=10,000			
			$\alpha = 0$		$\alpha \neq 0$		$\alpha = 0$		$\alpha \neq 0$	
			$\eta = 0$	$\eta \neq 0$	$\eta = 0$	$\eta \neq 0$	$\eta = 0$	$\eta \neq 0$	$\eta = 0$	$\eta \neq 0$
Size	Margin 1	β_1	80.9	79.9	80.5	78.4	95.7	97.4	96.7	94.8
		β_2	54.1	51.5	52.0	52.3	80.9	81.2	82.8	80.3
		$\beta_{1,2}$	45.5	47.8	47.0	45.0	68.9	73.3	72.4	70.3
	Margin 2	β_1	50.8	51.8	51.0	49.3	80.6	80.1	81.9	77.4
		β_2	82.0	80.6	80.2	78.9	95.7	97.4	96.8	94.8
		$\beta_{1,2}$	57.6	58.6	59.1	57.8	79.5	80.5	80.1	79.0
Power	Margin 1	β_1	92.1	94.0	92.4	93.1	99.3	99.8	99.4	99.9
		β_2	93.6	96.1	94.0	95.0	99.8	99.9	99.3	100.0
		$\beta_{1,2}$	75.3	75.3	74.7	75.7	79.2	79.2	77.1	75.8
	Margin 2	β_1	92.8	95.7	93.4	93.9	99.6	99.6	99.6	99.9
		β_2	95.2	96.7	95.0	95.4	99.8	99.9	99.6	100.0
		$\beta_{1,2}$	76.4	78.6	74.9	78.6	76.9	76.0	75.2	75.8

Percentage rejections of the null hypothesis of zero cross-excitation, when applying the regression method on $1,000$ simulations from Hawkes models over a time period of $5,000$ and $10,000$ time instances using 5% critical values. The parameters and the minimum magnitude of events are set on values estimated from the data set of Grothe et al. (2014). In the Hawkes models with the parameter restriction $\alpha = 0$, the magnitude of events have no influence on the triggering of subsequent events. In the Hawkes models with the parameter restriction $\eta = 0$, the history of the events has no influence on the magnitude of subsequent events.

Table 3.5: Size and power orthogonality tests

	T=5,000				T=10,000			
	$\alpha = 0$		$\alpha \neq 0$		$\alpha = 0$		$\alpha \neq 0$	
	$\eta = 0$	$\eta \neq 0$	$\eta = 0$	$\eta \neq 0$	$\eta = 0$	$\eta \neq 0$	$\eta = 0$	$\eta \neq 0$
Size	2.2	1.1	1.0	2.0	1.5	1.8	2.1	1.9
Power	12.2	14.1	14.8	14.2	18.2	19.5	21.3	19.8

Percentage rejections of the null hypothesis of zero cross-excitation, when performing Kolmogorov-Smirnov test on the transformed times (3.44) of $1,000$ simulations from Hawkes models over a time period of $5,000$ and $10,000$ time instances using 5% critical values. The parameters and the minimum magnitude of events are set on values estimated from the data set of Grothe et al. (2014). In the Hawkes models with the parameter restriction $\alpha = 0$, the magnitude of events have no influence on the triggering of subsequent events. In the Hawkes models with the parameter restriction $\eta = 0$, the history of the events has no influence on the magnitude of subsequent events.

index. As Gonzalo and Olmo (2005) find for the Dow Jones Stock Price Index and Dow Jones Corporate Bonds Indices that causality of extremes seems to depend on the maturity of the bonds, and hence to identify dependence between different stock and bond extremes, testing is necessary. Also, we investigate the cross-dependence between the stock and bond markets on one hand and exchange rate markets on the other hand. For the bond market we include series of both positive and negative exchange rate extremes in our application. Furthermore we test for dependence within the segment of stock markets and the segment of exchange rate markets.

With the LM test we are able to separately test for cross-excitation to one series from all the other series at once because no estimation under the alternative hypothesis is required. In contrast to the LR test for n series, $n \times (n+1)$ models need to be estimated instead of n, which is quite time-consuming in our case as it includes 11 series. The equity indices we consider are the S&P index, the NASDAQ and the Dow Jones Industrial index. For the bond index we take the US aggregate government bond series. The exchange rates we look at are the \$/€-rate, the \$/£-rate and the \$/¥-rate. Our data consists of daily prices between January 1, 1990, and July 1, 2015. The Hawkes models are applied to extreme price movements above the 95% and the 97.5% quantile.

Table 3.6 reports the LM test statistics for dependence between extremes in the equity, the bond and the exchange rate series. Here the element (i, j) corresponds to the cross-triggering effect to the series in row i from the series in column j. Here we only present the results for the simplest Hawkes models, in which event sizes are unpredictable and the sizes of past events do not affect the probability of triggering new events. Results for the other models are similar.

Between markets, we find that crashes in the stock markets amplify the probability that extremes in the other markets occur. However, according to the LM test, there is a lack of spillover effects the other way around. The LM test detects both stock-bond contagion as the flight-to-quality phenomenon. Moreover, Table 3.6 shows that the effects from the equity indices to the US bond market are strongest, as the LM test statistics for the triggering of crashes in the US bond market range from 14.45 to 26.95 and the LM test statistics for the triggering of booms in the US bond market by crashes in the S&P500 index and the DJI index range from 13.96 to 17.27 among the equity indices and thresholds. Crashes in all the equity indices also provoke both crashes and booms in the \$/€-rate and crashes in \$/¥-rate

using both thresholds, while only the LM statistics for spillover effects from the S&P 500 index and the DJI index to the \$/£-rate are consistently high and significant. The rate at which booms in the \$/¥-rate occur, seems to be unaffected by the occurrence of extremes in the other markets. Furthermore with the LM test we discover that, using both thresholds, crashes in the \$/€-rate and the \$/¥rate trigger crashes in the US bond market and that booms in the US bond market trigger booms in the \$/€-rate.

Within the stock market segment, we find, using both thresholds, highly significant LM test statistics indicating that crashes in the NASDAQ provoke crashes in the S&P 500 index, and that crashes in both the S&P 500 index and the NASDAQ increase the probability of a crash in the DJI index. We do not encounter evidence of cross-excitation from the DJI index to the other indices. Therefore, we conclude that crashes in the NASDAQ have the largest influence among the stock market indices. Within the exchange rate segment, the LM test points out that the occurrences of crashes in the \$/€-rate and the \$/£-rate are strongly related. Also, booms in \$/¥-rate trigger booms in the \$/€-rate. However, as the LM test suggests, extremes in the \$/¥-rate and the \$/£-rate are independent of each other, and spillover effects from other two exchange rates to the \$/¥-rate are absent. Hence the \$/¥-rate seems to behave differently from the other two exchange rates.

Comparing the the LM test statistics for extremes above the 95% and the 97.5% quantile, most statistics are quite close or only barely significant. The most striking differences occur for the spillover effects to booms in the US bond market. Apparently US bond booms further in the tail behave differently from smaller US bond booms. Furthermore the highly significant LM test statistic for the spillover effect from crashes in the \$/€-rate to crashes in the \$/£-rate disappears when extremes above the 97.5% threshold are studied instead of extremes above the 95% threshold.

All in all we conclude that there is substantial cross-dependence of extreme events in financial markets, both within and between segments. To account for the cross-triggering phenomenon, we therefore recommend that univariate Hawkes models be extended.

3.5.1 Comparision to GARCH

We are interested in whether the indicated cross dependence between extremes in the stock, bond and exchange rate series modeled by Hawkes models is a sign of extremal dependence

Table 3.6: LM test dependence stock, bond and exchange markets

		S&P	NDQ	DJI	BND(-)	$/€(-)	$/£(-)	$/¥(-)	BND(+)	$/€(+)	$/£(+)	$/¥(+)
95%	S&P		**7.97**	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08
	NDQ	3.82		2.28	0.00	0.00	0.00	0.00	**4.27**	0.29	0.34	1.14
	DJI	**13.44**	**12.25**		0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.20
	BND(-)	**26.95**	**14.45**	**21.58**		**5.63**	0.74	**4.48**				
	$/€(-)	**5.21**	**6.43**	**5.89**	0.25		**11.51**	0.33				
	$/£(-)	**5.79**	1.96	**9.36**	0.07	**14.36**		0.00				
	$/¥(-)	**12.43**	**7.09**	**14.03**	0.38	0.00	1.05					
	BND(+)	**16.23**	**8.38**	**13.96**						1.80	0.04	3.22
	$/€(+)	**8.99**	**6.98**	**13.08**					**5.83**		1.65	**6.13**
	$/£(+)	**3.88**	0.84	**8.05**					2.87	3.62		1.52
	$/¥(+)	0.00	0.00	0.13					0.27	0.00	0.00	
97.5%	S&P		**10.39**	1.90	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.84
	NDQ	**4.03**		3.56	0.02	0.00	0.00	0.00	2.68	0.00	0.00	2.51
	DJI	**7.89**	**11.78**		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.79
	BND(-)	**23.21**	**15.65**	**21.06**		**6.09**	**3.94**	1.73				
	$/€(-)	**3.89**	**7.06**	**4.06**	0.00		**8.57**	2.09				
	$/£(-)	**6.58**	0.07	**6.52**	0.15	1.91		1.07				
	$/¥(-)	**7.66**	**8.41**	**10.18**	0.74	0.78	3.47					
	BND(+)	**15.34**	1.95	**17.27**						**5.26**	0.36	**11.24**
	$/€(+)	**11.16**	**15.79**	**11.68**					1.81		2.92	**7.43**
	$/£(+)	**10.71**	**4.01**	**8.46**					**4.97**	2.43		3.31
	$/¥(+)	0.00	0.00	0.00					0.00	0.00	0.00	

LM test statistics based on extreme negative price movements in the equity indices, the S&P 500 index, the NASDAQ and the Dow Jones Industrial index, and extreme negative as well as extreme positive price movements in the US bond index and the exchange rates, that is the $/€-rate, the $/£-rate and the $/¥-rate. The element (i, j) corresponds to the cross-triggering effect to the series in row i from the series in column j. The tests are applied to extreme price movements above the 95% and the 97.5% quantile over a period that starts at January 1, 1990 and ends at July 1, 2015. The boldface numbers indicate significant cross-triggering at a 5% level.

not explained by volatility spillovers between the series. In order to assess whether this is the case we use an approach similar to Gresnigt et al. (2015). That is, we remove potential volatility effects by first estimating GARCH models that account for spillovers on the series, after which we proceed as before using the residuals.

In the General AutoRegressive Conditional Heteroscedasticity (GARCH) models, the variance of the current error term is a function of the innovations and the variance of the error terms in previous periods. In particular, in GARCH models that allow volatility from other series to spillover (Hamao et al., 1990, see for example), in addition to innovations in the series itself, also innovations in another series affect the conditional variance. The time-varying conditional variance enables the models to capture the volatility clustering feature observed in and across financial series. In the bivariate GARCH(1,1) model, the conditional variances $\sigma_{i,t}^2$, $\sigma_{j,t}^2$, of the stochastic error processes $\epsilon_{i,t}$, $\epsilon_{j,t}$, when conditioned on the history of the processes, are specified as follows

$$\sigma_{i,t}^2 = \omega_i + \alpha_{ii}\epsilon_{i,t-1}^2 + \alpha_{ij}\epsilon_{j,t-1}^2 + \beta_i\sigma_{i,t-1}^2 \tag{3.46}$$

$$\sigma_{j,t}^2 = \omega_j + \alpha_{ji}\epsilon_{i,t-1}^2 + \alpha_{jj}\epsilon_{j,t-1}^2 + \beta_j\sigma_{j,t-1}^2 \tag{3.47}$$

To account for the stylized fact that, even after correcting for volatility clustering, extreme returns occur more often than under normality, the heavy-tailed Student-t distribution is used for the error terms.

We evaluate whether extremal dependence between the stock, bond and exchange rate series found in the previous section cannot be explained by volatility spillovers between the series as follows. First we estimate a bivariate GARCH(1,1) model on each couple of the series. Next we apply the Hawkes models without cross-excitation to the extreme standardized residuals from the GARCH models, after which perform the LM test for dependence on these residuals. The results of this exercise for extremes above the 95% and the 97.5% quantile in the S&P 500 index, the NASDAQ, the DJI index, the US bond market, the \$/€-rate, the \$/£-rate and the \$/¥-rate using the simplest Hawkes models as in the previous section, are presented in Table 3.10.

Comparing Table 3.10 with Table 3.6, which contains the LM test statistics for the unfiltered series, we find that extremal dependence is not found in $14 - 15$ cases when we first remove the volatility spillovers by taking the standardized residuals from the bivari-

Table 3.7: LM test dependence residuals bivariate GARCH

		S&P	NDQ	DJI	BND(-)	$/€(-)	$/£(-)	$/¥(-)	BND(+)	$/€(+)	$/£(+)	$/¥(+)
95%	S&P		**7.58**	**5.55**	**14.04**	2.53	0.00	**10.71**	0.08	0.16	0.55	0.00
	NDQ	0.00		0.57	**9.01**	0.77	0.00	1.93	0.00	0.00	0.00	0.00
	DJI	1.38	**9.56**		**7.20**	1.55	0.00	**4.62**	0.00	0.00	0.35	0.00
	BND(-)	2.45	2.45	2.07		0.32	0.00	3.26				
	$/€(-)	**7.03**	**6.05**	**8.87**	**8.06**		**8.88**	**6.72**				
	$/£(-)	2.88	2.15	**4.58**	**7.35**	**13.07**		**6.87**				
	$/¥(-)	0.00	0.00	0.01	2.73	1.34	0.35					
	BND(+)	**16.41**	**25.89**	**27.83**						3.15	0.32	2.40
	$/€(+)	3.56	**8.55**	**6.42**					**6.54**		**15.62**	**5.79**
	$/£(+)	**3.89**	2.30	2.06					3.41	2.78		0.23
	$/¥(+)	**14.13**	**15.90**	**7.14**					1.75	1.82	0.87	
97.5%	S&P		**16.42**	0.00	**12.98**	1.54	0.17	**8.20**	2.93	0.96	0.61	1.51
	NDQ	0.00		0.00	**8.21**	0.06	0.00	0.57	0.15	0.00	0.00	0.00
	DJI	**13.17**	**21.14**		1.87	1.78	1.20	0.00	3.39	0.04	0.00	0.00
	BND(-)	**8.94**	**7.20**	**7.91**		0.83	0.55	**9.71**				
	$/€(-)	**8.64**	**5.34**	3.80	**15.43**		**7.17**	**3.92**				
	$/£(-)	**4.17**	**5.00**	2.59	**17.98**	**11.41**		**8.36**				
	$/¥(-)	0.00	0.12	0.00	0.00	0.00	0.99					
	BND(+)	**24.01**	**20.53**	**10.88**						0.01	0.00	2.19
	$/€(+)	**7.24**	**8.25**	2.78					**5.58**		**8.38**	0.12
	$/£(+)	**5.41**	2.34	0.56					1.04	3.50		1.88
	$/¥(+)	3.59	0.64	0.00					0.00	0.00	0.00	

LM test statistics based on extreme standardized residuals above the 95% and the 97.5% quantile from the bivariate GARCH(1,1) models, estimated on every couple of series over a period that starts at January 1, 1990 and ends at July 1, 2015. The element (i, j) corresponds to the cross-triggering effect to the series in row i from the series in column j. The boldface numbers indicate significant cross-triggering at a 5% level.

ate GARCH models. However removing volatility effects, extremes in other series seem to cross-excite each other while no significant evidence of cross triggering was found before in $13 - 14$ cases. In 28 cases the LM test statistics give the same result for the unfiltered and the filtered series. We also performed the LM test for dependence on the standardized residuals from estimating an univariate GARCH(1,1) model on the individual series, which corresponds (3.46) with α_{ij} and α_{ji} set to zero. Results can be found in the appendix. Again, while the LM test indicated cross-excitation between the unfiltered series, we do not discover extremal dependence with the LM test in at least 14 cases. Moreover, in a number of cases it is the other way around. Therefore we conclude that the GARCH residuals show a different kind of dependence and that the extremal dependence found between the unfiltered series cannot simply be explained by volatility spillovers.

3.6 Conclusion

We proposed Lagrange Multiplier (LM) based specification tests for univariate Hawkes models, where the specific focus is on tests for contagion of extreme events across series. Thereby we provided a general approach for testing different specifications of Hawkes models as well as an adequate assessment of the contribution of including cross-sectional dependence in a point process framework with serial dependence.

The models examined are based on the self-exciting Hawkes point process. The univariate Hawkes process is characterized by an event rate that increases when an event arrives, after which it decays as a function of the time passed since the event. In case of cross-triggering effects between series, the event rate also amplifies when an event occurs in another series such that series exhibit clustering of events through time within series as well as clustering of events among series.

The LM principle can be used to test for omitted explanatory variables, which includes spillover effects caused by the occurrence of events in other series, for omitted impact of the sizes of events on the triggering of new events, for omitted predictability of event sizes, and for breaks in the model parameters. When an extensive version of the conditional intensity is under investigation, we notice that the score consists of a term which reflects the change in the probability distribution of the time between two consecutive events and a term which reflects the relative change in the rate at which events occur. In case the distribution of the event sizes is extended, the score reflects the change in probability to observe an event of a certain magnitude. The LM test constructed from the scores, enables us to verify whether the extension contributes to the prediction of occurrence times or the magnitude of future events.

Reviewing the quality of the LM test for cross-triggering effects by means of a Monte Carlo analysis, we conclude the LM test outperforms the regression method of Van Oordt and Zhou (2012) and the orthogonality test discussed by Lando and Nielsen (2010). Over longer time periods, the power of the LM test is high and the LM test is seems properly sized. When the time period considered is shorter, the quality of the LM test falls only a little. We also compare the LM test to the computationally more demanding LR test. The LR test is supposed to outperform the LM test, as the LR test uses information on the likelihood under both the null and the alternative hypothesis, while for the LM test only the first is needed.

The LM test facilitates to separately test for amplification effects from multiple series to the conditional intensity of a series in one go. Therefore, in contrast to the LR test that requires the estimation of $n(n+1)$ models, only n models have to be estimated when n series are involved. The LR test is more undersized but somewhat more powerful than the LM test. All in all, as the LR test is a lot more time consuming and as the LM test performs very well, we recommend to use the LM test in empirical analysis.

We illustrated the use of the LM test by examining the cross-sectional dependence between 3 US stock indices, a US bond index and 3 exchange rates between January 1, 1990, and April 10, 2015. Between markets, we conclude that crashes in the stock markets provoke extremes in the other markets while the other way around spillover effects seem absent. The triggering effect from the equity indices to the US bond market seems to be the strongest, where the LM test detects both stock-bond contagion and the flight-to-quality phenomenon. Furthermore, with the LM test we discovered that crashes in the the \$/€-rate trigger crashes in the US bond market and that booms in the US bond market trigger booms in the \$/€-rate. Within the stock market and exchange rate market segments, we also found strong evidence of cross-excitation. Therefore, we conclude that cross-dependence of extreme events in financial markets is high and that univariate Hawkes models better be extended to account such a cross-triggering phenomenon.

We evaluate whether the indicated cross dependence between extremes in the stock, bond and exchange rate series modeled by Hawkes models is a sign of extremal dependence not explained by volatility spillovers between the series. Therefore we remove potential volatility effects by first estimating GARCH models that do and do not account for spillovers on the series, and apply the LM test for dependence to the extreme standardized residuals from this estimation. For the GARCH models that account for spillovers, the LM test statistics become insignificant and significant in about the same number of cases, while for the GARCH models that do not account for spillovers at least as many LM test statistics become insignificant. Hence, we conclude that the GARCH residuals show a different kind of dependence and that the extremal dependence found between the unfiltered series cannot simply be explained by volatility spillovers.

3.A Simulation procedures

3.A.1 Univariate procedure

In the simulation procedures we do not use a burn-in period, as we assume no events occurred before $t = 0$.

1. When no other events are present yet, the event process is equal to a homogeneous Poison process with μ, the constant background rate. Simulate the time till the first event from an exponential distribution with parameter μ. The time of the event t_1 recorded is the end of the interval in which the event occurs. Simulate the magnitude of the event from an independent General Pareto Distribution.

2. For t_n after t_1 calculate the probability of the occurrence of no event in the interval $[t_{n-1}, t_n]$ using (4.7). Simulate a random number r from a uniform distribution on the interval $[0, 1]$. When $r > \Pr\left(N(t_n) - N(t_{n-1}) = 0\right)$ record the time point t_n as the time of an event and simulate the magnitude of the event from a General Pareto Distribution. When $r < P\left(N(t_n) - N(t_{n-1}) = 0\right)$ do nothing.

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

3.A.2 Multivariate procedure

1. When no other events are present yet, the event process is equal to a homogeneous Poison process with μ_i, the constant background rate of series i. Obtain the time till the first event by simulating from exponential distributions with the parameters μ_i, where i ranges from 1 to the number of simulated series d, and selecting the lowest value among the simulated values. The time of the first event t_1 recorded is the end of the interval in which this event occurs. Simulate the magnitude of this event from an independent General Pareto Distribution with parameters that are characteristic for the size distribution of events in the market in which the first event occurs.

2. For every series $i \in \{2, \ldots, d\}$, for t_n after t_1 calculate the probability of the occurrence of no event in the interval $[t_{n-1}, t_n]$ in the series using (4.7). Simulate d random numbers $r_1, \ldots, r_d$ from a uniform distribution on the interval $[0, 1]$. For every series

$i \in \{2, \ldots, d\}$, when $r_i > \Pr_i\left(N(t_n) - N(t_{n-1}) = 0\right)$ record the time point t_n as the time of an event in series i and simulate the magnitude of the event from a General Pareto Distribution with parameters that are characteristic for the size distribution of events in series i. When $r_i < \Pr_i\left(N(t_n) - N(t_{n-1}) = 0\right)$ do nothing.

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

At every step in the simulation procedure we have to be able to recover the entire history of the Hawkes process as the history of the process incorporated in the conditional distribution of the sizes of events and therefore needed to simulate the event sizes. Using the branching structure of the Hawkes process to simulate events, we are not able to retrieve the entire history of the event process at the branches (there could be events triggered in other branches with occurrence times prior to the occurrence time of some event in a certain branch in a later stage of the simulation procedure). Therefore the simulation procedure cannot be based on the branching structure of the Hawkes process and "events have to be simulated over time". An alternative procedure to generate events is the modified thinning algorithm of Ogata (1981).

3.B Simulation results tests omitted predictability of event sizes

3.C Simulation results tests omitted impact of sizes events

3.D Simulation results LM test for dependence between series using the Hessian matrix

3.E Results LM test for dependence residuals univariate GARCH

Table 3.8: Size and power LM and LR tests for the omitted predictability of event sizes

	LM				LR			
	T=5,000		T=10,000		T=5,000		T=10,000	
	$\alpha = 0$	$\alpha \neq 0$	$\alpha = 0$	$\alpha \neq 0$	$\alpha = 0$	$\alpha \neq 0$	$\alpha = 0$	$\alpha \neq 0$
Size	3.9	3.4	3.1	2.2	3.5	3.1	2.6	2.4
Power	91.5	89.8	97.6	95.7	92	90.1	98.4	96.4

Percentage rejections of the null hypothesis of no influence of the history of the event process on the sizes of subsequent events ($\eta = 0$), when performing the LM and the LR test on $1,000$ simulations from Hawkes models over a time period of $5,000$ and $10,000$ time instances using 5% critical values. For the power of the tests the models are simulated under the alternative, while for the size the models are simulated under the null. The parameters and the minimum magnitude of events are set on values estimated from the data set of Grothe et al. (2014). In the Hawkes models with the parameter restriction $\alpha = 0$, the magnitude of events have no influence on the triggering of subsequent events.

	LM				LR			
	T=5,000		T=10,000		T=5,000		T=10,000	
	$\eta = 0$	$\eta \neq 0$	$\eta = 0$	$\eta \neq 0$	$\eta = 0$	$\eta \neq 0$	$\eta = 0$	$\eta \neq 0$
Size	4.3	4.0	3.2	2.4	2.3	3.2	1.8	2.0
Power	37.2	57.2	49.5	77.2	31.6	53.8	44.8	76.5

Percentage rejections of the null hypothesis of no impact of the sizes of events on the rate at which subsequent events arrive ($\alpha = 0$), when performing the LM and the LR test on $1,000$ simulations from Hawkes models over a time period of $5,000$ and $10,000$ time instances using 5% critical values. For the power of the tests the models are simulated under the alternative, while for the size the models are simulated under the null. The parameters and the minimum magnitude of events are set on values estimated from the data set of Grothe et al. (2014). In the Hawkes models with the parameter restriction $\eta = 0$, the history of the events has no influence on the magnitude of subsequent events.

Table 3.9: Size and power LM tests using the Hessian matrix

		T=5,000				T=10,000			
		$\alpha = 0$		$\alpha \neq 0$		$\alpha = 0$		$\alpha \neq 0$	
		$\eta = 0$	$\eta \neq 0$	$\eta = 0$	$\eta \neq 0$	$\eta = 0$	$\eta \neq 0$	$\eta = 0$	$\eta \neq 0$
Size	Γ_{12}	0.8	1.3	0.8	0.1	0.5	0.7	1.3	0.0
	Γ_{21}	0.7	1.2	1.3	0.6	1.1	0.7	0.7	0.1
Power	Γ_{12}	61.8	77.7	63.9	58.9	91.7	97.9	93.3	92.5
	Γ_{21}	89.9	94.9	89.2	89.0	99.3	98.7	99.5	99.9

Percentage rejections of the null hypothesis of zero cross-excitation, when performing the LM test on $1,000$ simulations from Hawkes models over a time period of $5,000$ and $10,000$ time instances using 5% critical values. In the tests the information matrix is approximated by the negative of the Hessian matrix instead of the average outer product of the score. For the power of the tests the models are simulated under the alternative, while for the size the models are simulated under the null. The parameters and the minimum magnitude of events are set on values estimated from the data set of Grothe et al. (2014). The parameters Γ_{12} and Γ_{21} control respectively the cross-triggering effect from series 2 to series 1 and the cross-triggering effect from series 1 to series 2 (3.30). In the Hawkes models with the parameter restriction $\alpha = 0$, the magnitude of events have no influence on the triggering of subsequent events. In the Hawkes models with the parameter restriction $\eta = 0$, the history of the events has no influence on the magnitude of subsequent events.

Table 3.10: LM test dependence residuals univariate GARCH

		S&P	NDQ	DJI	BND(-)	$/€(-)	$/£(-)	$/¥(-)	BND(+)	$/€(+)	$/£(+)	$/¥(+)
95%	S&P		**6.57**	2.34	5.31	0.00	0.00	**6.07**	0.00	0.00	0.00	0.00
	NDQ	0.00		0.00	0.44	0.00	0.00	0.21	0.00	0.00	0.00	0.00
	DJI	1.08	**4.09**		3.20	0.00	0.00	1.93	0.00	0.00	0.00	0.00
	BND(-)	**4.51**	2.19	**6.19**		0.11	0.00	1.54				
	$/€(-)	**4.85**	2.15	**6.46**	2.70		1.08	**4.26**				
	$/£(-)	3.48	1.32	3.17	2.33	**7.00**		3.75				
	$/¥(-)	0.00	0.00	0.00	1.20	0.00	0.01					
	BND(+)	**15.19**	**20.59**	**26.13**						1.30	0.15	1.68
	$/€(+)	**4.77**	**7.95**	**4.58**					2.70		**3.88**	**4.72**
	$/£(+)	1.67	0.91	0.57					1.74	0.16		0.00
	$/¥(+)	**14.43**	**15.55**	**7.90**					0.95	0.04	0.34	
97.5%	S&P		**8.45**	0.49	2.48	0.04	0.00	2.34	0.00	0.00	0.00	0.00
	NDQ	0.00		0.00	0.82	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	DJI	2.84	**17.39**		2.36	0.92	0.00	0.78	0.00	0.00	0.00	0.00
	BND(-)	0.89	0.64	1.99		1.54	1.02	**7.08**				
	$/€(-)	2.52	**4.07**	1.16	**8.39**		**5.81**	**4.98**				
	$/£(-)	3.11	3.53	2.17	**8.66**	**5.31**		**4.28**				
	$/¥(-)	0.00	0.00	0.00	0.00	0.00	0.33					
	BND(+)	**14.74**	**7.23**	**5.67**						0.00	0.00	0.03
	$/€(+)	**7.17**	**8.62**	2.08					0.80		**4.36**	0.00
	$/£(+)	**4.03**	1.96	0.40					0.13	0.87		0.00
	$/¥(+)	0.00	0.00	0.00					0.00	0.00	0.00	

LM test statistics based on extreme standardized residuals above the 95% and the 97.5% quantile from the univariate GARCH(1,1) models, estimated on every series over a period that starts at January 1, 1990 and ends at July 1, 2015. The element (i, j) corresponds to the cross-triggering effect to the series in row i from the series in column j. The boldface numbers indicate significant cross-triggering at a 5% level.

Chapter 4

Exploiting spillovers to forecast crashes

4.1 Introduction

We develop Hawkes models in which events are triggered through self as well as cross-excitement.[1] Exploiting the cross-sectional dependence between financial series, we aim to improve the forecasts of extremes and their sizes. It has already been shown that Hawkes models contribute to the identification and prediction of extreme events in financial markets.[2] Currently, research in finance pays much attention to the modeling of extremal dependence between financial markets, though with an in-sample focus. We extend these studies on contagion, as we examine whether incorporating this dependence improves forecasts. For a wide range of assets, we find that Hawkes models with spillover effects forecast the probability of crashes and the Value-at-Risk significantly more accurately than models without.

Traders, regulators of financial markets and risk management benefit greatly from the accurate forecasts of extreme price movements in financial markets. Nowadays, a large literature focuses on modeling extremal dependence between financial markets.[3] This topic gained interest since the financial crisis of 2008 as this crisis demonstrated the consequences of the cohesion between the behavior of the prices in the financial markets. For example on September 29, October 15 and December 1 in 2008 the S&P 500, the Dow Jones Industrial

[1] This Chapter is based on Gresnigt et al. (2016b)

[2] See amongst others: Chavez-Demoulin et al. (2005), Hewlett (2006), Bowsher (2007), Bauwens and Hautsch (2009), Herrera and Schipp (2009), Embrechts et al. (2011), Chavez-Demoulin and McGill (2012), Aït-Sahalia et al. (2014), Gourieroux et al. (2014), Grothe et al. (2014), Aït-Sahalia et al. (2015), Bacry et al. (2015), Bormetti et al. (2015), Gresnigt et al. (2015) and Gresnigt et al. (2016a)

[3] See amongst others: Longin and Solnik (1995), Poon et al. (2003), Poon et al. (2004), Bekaert et al. (2010), Grothe et al. (2014), and Aït-Sahalia et al. (2015)

Average (DJI) and the NASDAQ, all suffered top 20 percentage losses. Furthermore on September 29 the euro/dollar rate and the pound/dollar rate also dropped by a large amount, while the US bond market boomed. On the 16th of October, just one day after the major US stock markets crashed, and on the 1th of December both currencies fell again sharply. Moreover 4 days after these dates US bond prices shifted significantly. These joint extremes demonstrate the overlap of periods in which financial markets are subject to tension with extreme price movements as a result.

A Hawkes model uses an inhomogeneous Poisson process to model the occurrence of events above a certain threshold. The event rate increases with the arrival of a new event whereas the event rate decays with the time passed since an event. As the probability of events increases when an event occurs, the Hawkes process (Hawkes, 1971) is called a self-exciting process. Characteristics typically observed in processes that fit Hawkes models are the clustering of extremes and serial dependence.

Earthquakes, for which the Hawkes models were originally designed, exhibit clustering behaviour in space as well as in time (Ogata, 1998). Similarities between the behavior of stock market returns around crashes and the dynamics of earthquake sequences have been noted in the so-called econophysics literature, in which physics models are applied to economics.[4] This literature indicates that it is likely that speculative bubbles which lead to crashes in the stock market, whether or not triggered by an exogenous factor, are caused by the positive herding behavior of investors. As the positive herding behavior of investors is locally self-enforcing, instability in the financial markets grows endogenously. Such a self-excitation can also be observed in seismic behavior around earthquake sequences, where an earthquake usually generates aftershocks which in turn can generate new aftershocks and so on. Earthquakes and stock returns therefore share characteristics typically observable as the clustering of extremes.

Like earthquake sequences, financial series seem to cluster in a dimension other than the time dimension.[5] Extreme stock returns across markets are found to be more correlated than small returns (Bae et al., 2003). They occur more frequently at the same time than expected under the assumption of a normal dependence structure (Mashal and Zeevi, 2002; Hartmann

[4] See amongst others: Sornette (2003), Weber et al. (2007), Petersen et al. (2010), Baldovin et al. (2011), Baldovin et al. (2013), Baldovin et al. (2015), and Bormetti et al. (2015)

[5] See amongst others: Eun and Shim (1989), Fischer and Palasvirta (1990), King and Wadhwani (1990), Lin et al. (1994) and Connolly and Wang (2003)

et al., 2004; Sun et al., 2009). This suggests that financial markets experience stress at the same time. For example, using the multivariate GARCH framework, volatility spillover effects between stock markets have been detected in numerous studies.[6] Interpreting volatility as a measure for the tension, these findings indicate that stress from financial markets pours over to other financial markets.

In this paper we extend Hawkes models to include contagion in financial markets. In the models we allow extreme events in one financial market to trigger both the occurrence and the magnitude of extreme events in other markets. Studies on Hawkes models for financial series in a multivariate setting include Bormetti et al. (2015), Grothe et al. (2014) and Aït-Sahalia et al. (2015). Different from these and other in-sample studies on financial contagion, we explicitly examine the effects of cross-excitation on out-of-sample forecasts. We assess its added value for the probability forecast of an extreme event and for Value-at-Risk. Thereby we extend Chavez-Demoulin et al. (2005), Herrera and Schipp (2009), and Gresnigt et al. (2015), who showed using Value-at-Risk and Expected Shortfall that univariate Hawkes models contribute to the modeling and prediction of risk in finance. Moreover Byström (2004) finds that conditional models based on Extreme Value Theory give particularly accurate VaR measures, which are superior to traditional approaches such as GARCH for both standard and more extreme VaR quantiles.

In somewhat more detail, we use the Lagrange Multiplier principle (see Breusch and Pagan, 1980; Engle, 1982; Hamilton, 1996) as in Gresnigt et al. (2016a) to test whether the spillovers contribute to the model fit. The correctness of the model specifications is further evaluated by means of the residual analysis methods as proposed in (Ogata, 1988). We assess the quality of the probability forecasts by the Quadratic and Log Probability Scores and the test of Diebold and Mariano (1995) (DM test) with an asymmetric loss function as proposed by Dijk and Franses (2003). For the evaluation of the VaR forecasts we use the unconditional coverage, independence and conditional coverage test of Christoffersen (1998), a Dynamic Quantile test in the line of Engle and Manganelli (2004) and the DM test based on the asymmetric loss function of Giacomini and Komunjer (2005).

We apply the models to extreme losses in three stock markets and to extreme losses and gains in the US bond market and two exchange rates. Hence, for the stock market

[6] See amongst others: Hamao et al. (1990), Bae and Karolyi (1994), Koutmos and Booth (1995), Booth et al. (1997) and Kanas (1998)

our analysis focuses on long positions, whereas we consider both long and short positions for bond and FX markets. We forecast the probability of crashes and the Value-at-Risk over a period that includes the financial crisis of 2008 using a moving window. Estimating the models, we find that on average the LM tests reject the absence of cross-excitation. Performing residual analysis, the fit of the models for the various series is good. Out-of-sample, models with spillover effects provide significantly more accurate forecasts of the occurrence of an extreme return and of the Value-at-Risk than the models without spillover effects for almost all series. We conclude that including cross-sectional dependence improves the forecasts of crashes, and hence cross-sectional dependence should not be ignored.

The rest of our paper is organized as follows. In Section 2 we give a brief introduction on Hawkes models. Furthermore we propose an extension of the univariate Hawkes model which incorporates cross-excitation and we discuss a Lagrange Multiplier test for dependence across series. In section 3 we apply the models and the LM test to US stocks, bonds and exchange rate data. Section 4 contains an in-sample assessment of the models by means of a residual analysis. The models are evaluated out-of-sample on the basis of the prediction of the probability of an extreme and the Value-at-Risk in Section 5. Section 6 concludes.

4.2 Hawkes models

4.2.1 Univariate model

The Hawkes model is a branching model. Each event can trigger subsequent events, and these can again trigger subsequent events. The model is based on the mutually self-exciting Hawkes point process, which is an inhomogeneous Poisson process. For the Hawkes process, the intensity at which events arrive at time t depends on the history of events prior to time t.

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

$$\lambda(t|\boldsymbol{\theta}; \mathcal{H}_t) = \mu + \gamma_1 \sum_{i:t_i<t} g(t - t_i, m_i), \tag{4.1}$$

where $\mu > 0$, $\gamma_0 > 0$ and $g(s, m) > 0$ whenever $s, m > 0$ and 0 elsewhere. The parameter γ_1 controls the maximum triggering intensity and the expected number of events directly triggered by an event. The conditional intensity consists of a constant term μ and a self-exciting function $\gamma_1 g(s, m)$, the latter function depends on the time passed since the jumps which occurred before t and the size of these jumps. Thus the rate at which events arrive is separated in a long-term background component and a short-term clustering component describing the temporal distribution of aftershocks. The conditional intensity uniquely determines the distribution of the process.

It is usual to specify the triggering function as

$$g(s, m) = e^{-\beta s} c(m) \tag{4.2}$$

where β determines how fast the possibility of triggering events decays depending on the time passed since an event. The influence of the sizes of past events on the intensity is given by $c(m)$.

For the influence of the sizes of past events on triggering of future events, $c(m)$, it is common to use the exponential form

$$c(m) = e^{\alpha(m-u)/u}, \tag{4.3}$$

where u represents the minimum magnitude of an event and α determines how the size of an event affects the probability of triggering other events. When $\alpha > 0$, larger events trigger more events than smaller events, because the probability of triggering events increases with the size of past events (m). The larger positive α is, the more pronounced is the influence of the size of events.

To enable an application to financial data, we choose a Generalized Pareto Distribution for the sizes of the events[7]

$$\Pr\left(m \leq M | m \geq u; \boldsymbol{\theta}; \mathcal{H}_t\right) = 1 - \left(1 + \xi \frac{m-u}{\sigma(t)}\right)^{-1/\xi}, \tag{4.4}$$

[7]See amongst others: Chavez-Demoulin et al. (2005), Poon et al. (2004), Herrera and Schipp (2009), Chavez-Demoulin and McGill (2012), Grothe et al. (2014), Gresnigt et al. (2015) and Gresnigt et al. (2016a)

where $\sigma(t) = \phi + \frac{\eta}{\gamma_1}(\lambda(t|\boldsymbol{\theta}; \mathcal{H}_t) - \mu)$ and $\xi \neq 0$. In case $\eta = 0$ the size of an event is unpredictable, whereas in case $\eta > 0$ the arrival times and sizes of previous events affect the distribution of the sizes of subsequent events. We scale the mean and variance of the distribution of the sizes of events with $\sigma(t)$. When $\eta > 0$ the magnitude of events is expected to be more extreme, when the conditional intensity is high. The larger positive η is, the more pronounced is the influence of the history of events on the size of subsequent events.

The expected number of off-spring of an event, is given by the branching ratio. Using (5.5) the branching ratio is equal to

$$\gamma_1 \int_0^\infty g(s, m)ds = \frac{\gamma_1}{\beta} c(m). \tag{4.5}$$

A Hawkes process is stationary when the branching ratio is smaller than 1. When the ratio exceeds 1, the rate at which events arrive will grow to infinity over time.

The log-likelihood of the Hawkes model, specified in the conditional intensity and the probability distribution of the sizes of the events, is given by

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{N} \log \lambda(t_i|\boldsymbol{\theta}; \mathcal{H}_t) - \int_0^T \lambda(s|\boldsymbol{\theta}; \mathcal{H}_s)ds + \sum_{i=1}^{N} \log f(m_i|\boldsymbol{\theta}; \mathcal{H}_t) \tag{4.6}$$

where t_i are the event arrival times in the interval $[0, T]$.

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

$$\begin{aligned} \Pr\left(N(t) - N(t - \delta t) > 0|\boldsymbol{\theta}; \mathcal{H}_t\right) &= 1 - \Pr\left(N(t) - N(t - \delta t) = 0|\boldsymbol{\theta}; \mathcal{H}_t\right) \\ &= 1 - F\left(t^* > \delta t|\boldsymbol{\theta}; \mathcal{H}_t\right) \\ &= 1 - \exp\left(-\int_{t-\delta t}^{t} \lambda(s|\boldsymbol{\theta}; \mathcal{H}_s)ds\right) \end{aligned} \tag{4.7}$$

4.2.2 Spillover model

When events across series tend to arrive around the same time, the occurrence of an event in one series could increase the probability that an event in another series arrives. A Hawkes model in which the conditional intensity of a series i is affected by the history of the event

process of another series j is represented as

$$\lambda(t|\boldsymbol{\theta};\mathcal{H}_t) = \mu + \gamma_1 \sum_{i:t_i<t} g(t-t_i, m_i) + \gamma_2 \sum_{j:t_j<t} h(t-t_j, m_j), \tag{4.8}$$

where $\gamma_2 h(t-t_j, m_j)$ presents the effect of an event in series j on the intensity at time t in series i. The cross-exciting function $h(s,m)$ can be specified in various ways. Here, we set $h(s,m) = g(s,m)$ to keep the model parsimonious and model parameters identified when there are no spillovers, $\gamma_2 = 0$. The events in the series j also affect the probability distributions of the sizes of events in series i through (4.4) when $\eta \neq 0$.

We perform a Lagrange Multiplier test on the null-hypothesis that $\gamma_2 = 0$ against the alternative hypothesis that $\gamma_2 > 0$ Gresnigt et al. (2016a). As we deal with a count process we define the score used for the LM test in event numbers and not in time points. The score of the n-th event is defined as the derivative of the logarithm of the conditional likelihood (5.30) with respect to the parameter vector $\boldsymbol{\theta}$,

$$\mathrm{g}_i(\bar{\boldsymbol{\theta}}) = \left.\frac{\partial \log p(y_i|\boldsymbol{\theta};\mathcal{H}_t)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}}, \tag{4.9}$$

which can be derived to be equal to

$$\mathrm{g}_i(\tilde{\boldsymbol{\theta}}) = -\frac{\partial \int_{t_{i-1}}^{t_i} \lambda(s|\boldsymbol{\theta};\mathcal{H}_s)ds}{\partial \boldsymbol{\theta}} + \frac{1}{\lambda(t_i|\boldsymbol{\theta};\mathcal{H}_t)} \left.\frac{\partial \lambda(t_i|\boldsymbol{\theta};\mathcal{H}_t)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} + \left.\frac{\partial \log f(m_i|\boldsymbol{\theta};\mathcal{H}_t)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}. \tag{4.10}$$

The score (4.10) consists of three terms that respectively reflect the change in the probability distribution of the time between two consecutive events, the relative change in the rate at which events occur and the change in probability to observe an event of a certain magnitude, when $\tilde{\boldsymbol{\theta}}$ is shifted by a infinitesimal amount. We assess the added value of the parameters to the model in terms of their contribution to the prediction of the occurrence times and magnitudes of events t_i and m_i.

Let $x(t) = \sum_{j:t_j<t} h(t-t_j, m_j)$ denote the additional tension caused by all events in a series j before time t. To test whether the cross-triggering effects are significant for a series i, we first estimate the Hawkes model under the null-hypothesis of no cross-excitation, $\gamma_2 = 0$. Next, we use (4.10) to compute the expected score from the derivative of the conditional intensity, the integrated conditional intensity and the probability distribution of event sizes,

given by

$$\frac{\partial \lambda(t_i|\boldsymbol{\theta};\mathcal{H}_t)}{\partial \gamma_2} = x(t_i) \tag{4.11}$$

$$\frac{\partial \int_{t_{i-1}}^{t_i} \lambda(s|\boldsymbol{\theta};\mathcal{H}_s)ds}{\partial \gamma_2} = \int_{t_{i-1}}^{t_i} x(s)ds \tag{4.12}$$

$$\frac{\partial \log f(m_i|\boldsymbol{\theta};\mathcal{H}_t)}{\partial \gamma_2} = \frac{\eta x(t_i)}{\gamma_1 \sigma(t_i)} \left(\frac{(m_i - u_i) - \sigma(t_i)}{\sigma(t_i) + \xi(m_i - u_i)} \right). \tag{4.13}$$

The LM test is asymptotically $\chi^2(1)$ under the null of no cross-excitation.

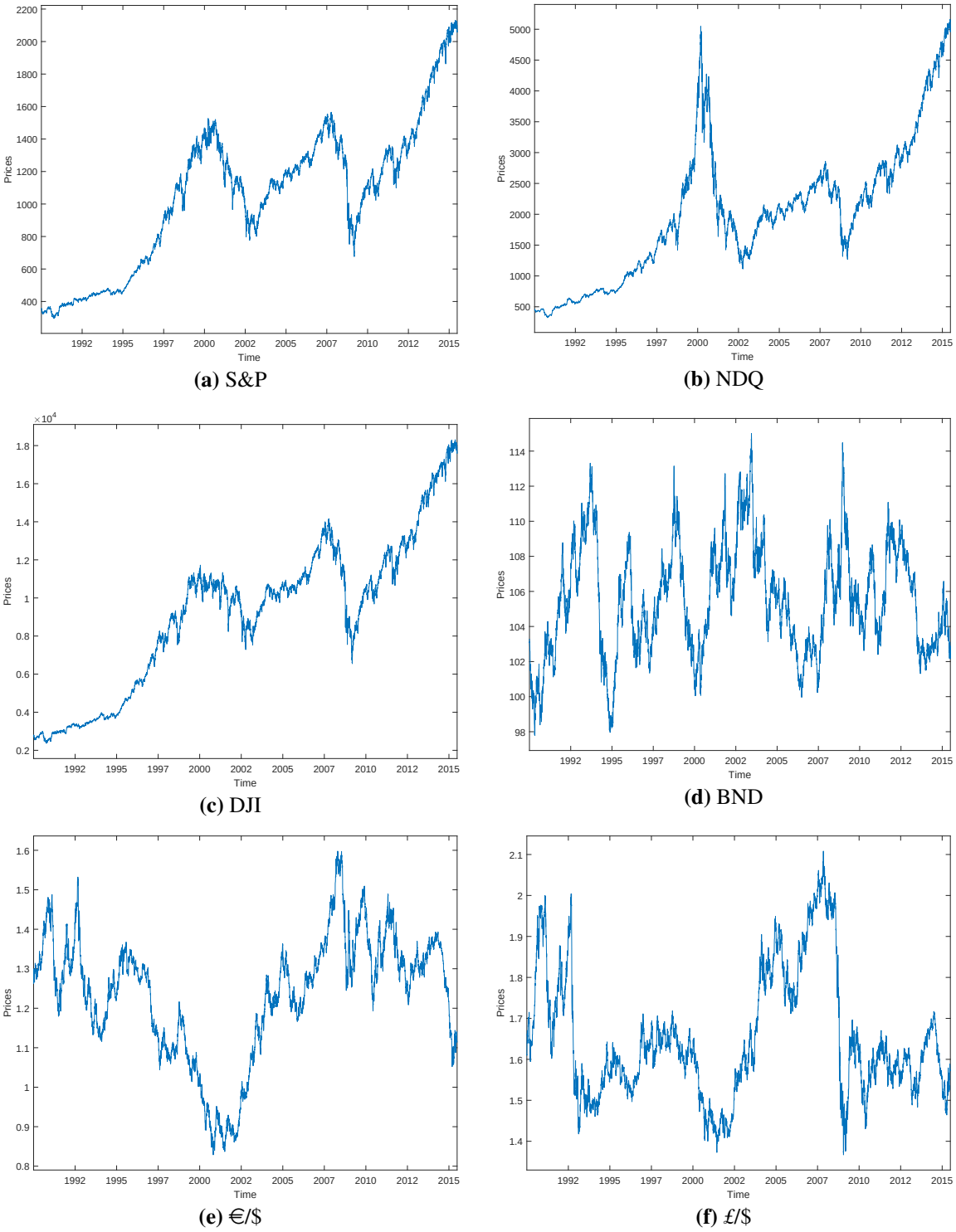
4.3 Application to Financial Data

We consider daily returns of three US stock market indices, the S&P 500 (S&P), the NASDAQ (NDQ) and the Dow Jones Industrial Average (DJI), and furthermore the US bond index (BND), and two exchange rates, the euro/dollar rate (€/$) and the pound/dollar rate (£/$) over the period from 1 January, 1990, to 1 July, 2015. We construct discrete returns as $R_t = (p_t/p_{t-1} - 1) \cdot 100$, where p_t denotes the value of a series at t. The Hawkes models are used to forecast the probability of crashes and the Value-at-Risk from 1 January, 2008, to 1 July, 2015, such that it includes the interesting period in which many crashes occurred. We use a moving window of $4,000$ trading days before we make our predictions to estimate the Hawkes models and we update our estimates every 40 trading days. Figure 4.1 displays the evolution of the prices of the stock market indices, the bond market index and the exchange rates from 1 January, 1990, to 1 July, 2015.

We apply the Hawkes models to negative extremes in the stock market indices and to negative and positive extremes in the US bond index and the exchange rates beyond the 95% and the 97.5% quantile. We set the thresholds for each series at the value that corresponds to the quantiles of their empirical distributions in the moving window.

The estimated Hawkes models differ as the parameters α, η and/or γ_2 are restricted to zero in some of the models, while in other models these parameters are unrestricted. When $\alpha = 0$ the sizes of former events do not influence the conditional intensity of new event arrivals, whereas they do so when $\alpha > 0$. When $\eta = 0$ the history of the event process does not influence the sizes of future events, while when $\eta > 0$ the sizes of events can be predicted from the past.

Figure 4.1: Prices



(a) S&P

(b) NDQ

(c) DJI

(d) BND

(e) €/$

(f) £/$

This Figure displays the values of the S&P 500 index, the NASDAQ, the Dow Jones Industrial Average, the US bond market, the euro/dollar rate and the pound/dollar from 1 January, 1990, to 1 July, 2015.

Our specific focus is on comparing the models in which $\gamma_2 = 0$ to the models in which $\gamma_2 > 0$. When $\gamma_2 = 0$, events in series i and their sizes are not effected by events in series j. When $\gamma_2 > 0$, the arrival rate of events in series i is amplified by the arrival of events in series j. The effect scales with the size of the events in the series j whenever $\alpha > 0$. Moreover, in case $\eta > 0$ the occurrence of events in the series j influences the distribution of the sizes of events in the series i. For illustrative purposes, in this paper we examine the impact of crashes in the S&P 500 index on extremes in the other series. We choose to investigate the spillovers from the S&P 500 index as the index is one of the most commonly followed equity indices and considered as one of the best representations of the US stock market. Furthermore we investigate whether crashes in the NASDAQ provoke crashes in the S&P 500 index. The NASDAQ is very well suited to serve as a proxy for spillover effects to the S&P 500 index from other stock market indices as the NASDAQ is the second-largest exchange in the world by market capitalization.

Tables 4.1 and 4.2 present the mean of parameter estimates across the estimation windows for the Hawkes model in which $\gamma_2 > 0$, $\eta > 0$ and $\alpha > 0$ when applied to extremes beyond the 95% and 97.5% quantiles. As $\gamma_2 > 0$, crashes in the S&P 500 index can provoke extremes in the other series, while crashes in the S&P 500 index can be triggered by crashes in the NASDAQ. Due to space constraints we only display the parameter estimates for this model.

The relatively large estimates for α for crashes in the S&P 500, the DJI index, and crashes and booms in the €/$-rate and the £/$-rate, indicate that a model in which the triggering intensity depends on the size of past events, matches the extremes closer than a model in which all events have the same impact on the triggering intensity. In the model where $\alpha > 0$, larger events trigger on average more events than smaller events.

The relatively large estimates for η for the S&P 500, the NASDAQ, the €/$-rate and the £/$-rate, provide evidence that the history of the event process influences the size of subsequent events. As $\eta > 0$, the size distribution has a higher mean and variance when the conditional intensity is high such that the magnitude of events is expected to be more extreme in that case. This means that, on average, larger events are observed in turbulent periods in which more events occur than in tranquil periods.

The parameter γ_2 in the bottom row of Tables 4.1 and 4.2 controls for the cross-triggering of events. For most series, the estimated parameters are relatively large compared to their

Table 4.1: Parameter estimates 95%

	S&P	NDQ	DJI	BND(-)	$/€(-)	$/£(-)	BND(+)	$/€(+)	$/£(+)
ξ	0.0599	0.0712	0.1268	0.1316	0.0004	0.0007	0.0657	0.0164	0.1009
	(0.0712)	(0.0737)	(0.0748)	(0.0883)	(0.0908)	(0.0894)	(0.0806)	(0.0686)	(0.0780)
ϕ	0.4877	0.5371	0.3888	0.1069	0.2220	0.1657	0.1088	0.2665	0.1870
	(0.0864)	(0.1130)	(0.0767)	(0.0246)	(0.0361)	(0.0324)	(0.0193)	(0.0431)	(0.0316)
μ	0.0089	0.0039	0.0114	0.0125	0.0187	0.0193	0.0197	0.0237	0.0246
	(0.0029)	(0.0021)	(0.0032)	(0.0048)	(0.0064)	(0.0052)	(0.0058)	(0.0068)	(0.0059)
γ_1	0.0240	0.0244	0.0133	0.0109	0.0091	0.0114	0.0091	0.0055	0.0046
	(0.0060)	(0.0060)	(0.0078)	(0.0039)	(0.0030)	(0.0038)	(0.0039)	(0.0029)	(0.0026)
β	0.0424	0.0357	0.0460	0.0248	0.0281	0.0332	0.0282	0.0227	0.0201
	(0.0095)	(0.0067)	(0.0102)	(0.0088)	(0.0087)	(0.0092)	(0.0091)	(0.0077)	(0.0067)
α	0.3267	0.3629	0.2731	0.3798	0.7478	0.6179	0.2603	0.6632	0.8304
	(0.1546)	(0.2005)	(0.1600)	(0.2576)	(0.2363)	(0.1861)	(0.2929)	(0.2550)	(0.2044)
η	0.1123	0.1181	0.0658	0.0096	0.0163	0.0373	0.0100	0.0154	0.0115
	(0.0401)	(0.0393)	(0.0422)	(0.0057)	(0.0077)	(0.0145)	(0.0060)	(0.0097)	(0.0069)
γ_2	0.0041	0.0030	0.0168	0.0048	0.0029	0.0033	0.0061	0.0028	0.0015
	(0.0027)	(0.0024)	(0.0078)	(0.0024)	(0.0016)	(0.0016)	(0.0032)	(0.0015)	(0.0008)

Parameter estimates for the Hawkes models with spillover effects ($\gamma_2 > 0$), impact of the magnitude of events on the triggering subsequent events ($\alpha > 0$) and influence of the history of the event process on the event sizes ($\eta > 0$). The models are applied to extreme negative returns in three US stock market indices and extreme negative as well as extreme positive returns in the US bond market and two exchange rates beyond the 95% quantile to forecast over a period that starts at 1 January, 2008, and ends at 1 July, 2015, using a moving window of $4,000$ trading days that is updated every 40 trading days. For the S&P 500 index, γ_2 controls the spillover effect from crashes in the NASDAQ, while for the DJI index, the US bond index, the euro/dollar rate and the pound/dollar this parameter controls the spillover effect from crashes in the S&P 500. The table presents the mean of the parameter estimates across the estimation windows. The mean of the standard errors of the parameter estimates is shown in between parentheses.

Table 4.2: Parameter estimates 97.5%

	S&P	NDQ	DJI	BND(-)	$/€(-)	$/£(-)	BND(+)	$/€(+)	$/£(+)
ξ	0.0814	0.0620	0.1339	0.0036	0.0001	0.0242	0.0336	0.0656	0.0850
	(0.1113)	(0.1250)	(0.1141)	(0.1028)	(0.1257)	(0.1688)	(0.1117)	(0.1047)	(0.1130)
ϕ	0.5682	0.7248	0.6586	0.2035	0.2005	0.1699	0.1369	0.2922	0.2219
	(0.1316)	(0.1970)	(0.1342)	(0.0360)	(0.0437)	(0.0491)	(0.0275)	(0.0512)	(0.0419)
μ	0.0064	0.0036	0.0068	0.0049	0.0087	0.0060	0.0133	0.0127	0.0133
	(0.0019)	(0.0014)	(0.0019)	(0.0023)	(0.0035)	(0.0027)	(0.0031)	(0.0033)	(0.0029)
γ_1	0.0266	0.0348	0.0026	0.0106	0.0099	0.0124	0.0068	0.0046	0.0022
	(0.0086)	(0.0113)	(0.0049)	(0.0044)	(0.0039)	(0.0045)	(0.0050)	(0.0032)	(0.0016)
β	0.0523	0.0540	0.0470	0.0238	0.0322	0.0294	0.0400	0.0276	0.0143
	(0.0141)	(0.0128)	(0.0119)	(0.0081)	(0.0103)	(0.0088)	(0.0142)	(0.0108)	(0.0052)
α	0.4132	0.3006	0.5281	0.0837	0.7083	0.5254	0.0817	0.9676	1.5137
	(0.2675)	(0.3912)	(0.2328)	(0.5126)	(0.4497)	(0.3736)	(0.5628)	(0.3486)	(0.2485)
η	0.1848	0.2011	0.0081	0.0000	0.0292	0.0514	0.0131	0.0115	0.0052
	(0.0843)	(0.0965)	(0.0176)	(0.0086)	(0.0162)	(0.0252)	(0.0118)	(0.0112)	(0.0058)
γ_2	0.0040	0.0063	0.0224	0.0079	0.0053	0.0052	0.0117	0.0036	0.0011
	(0.0034)	(0.0042)	(0.0080)	(0.0037)	(0.0029)	(0.0024)	(0.0067)	(0.0023)	(0.0007)

Parameter estimates for the Hawkes models with spillover effects ($\gamma_2 > 0$), impact of the magnitude of events on the triggering subsequent events ($\alpha > 0$) and influence of the history of the event process on the event sizes ($\eta > 0$). The models are applied to extreme negative returns in three US stock market indices and extreme negative as well as extreme positive returns in the US bond market and two exchange rates beyond the 97.5% quantile to forecast over a period that starts at 1 January, 2008, and ends at 1 July, 2015, using a moving window of $4,000$ trading days that is updated every 40 trading days. For the S&P 500 index, γ_2 controls the spillover effect from crashes in the NASDAQ, while for the DJI index, the US bond index, the euro/dollar rate and the pound/dollar this parameter controls the spillover effect from crashes in the S&P 500. The table presents the mean of the parameter estimates across the estimation windows. The mean of the standard errors of the parameter estimates is shown in between parentheses.

standard errors. The ratio between the instantaneous cross-excitation and the self-excitation effect is, except for the S&P 500 index and the NASDAQ, larger than 0.3 and 0.4, for the 95% and 97.5% thresholds. The instantaneous effect of crashes in the S&P 500 index on the rate at which crashes in the DJI index occur is on average even larger than the instantaneous effect of crashes in the series itself. Comparing the parameter estimates for the different thresholds, it seems that larger events are triggered more frequently by events in another series than smaller events.

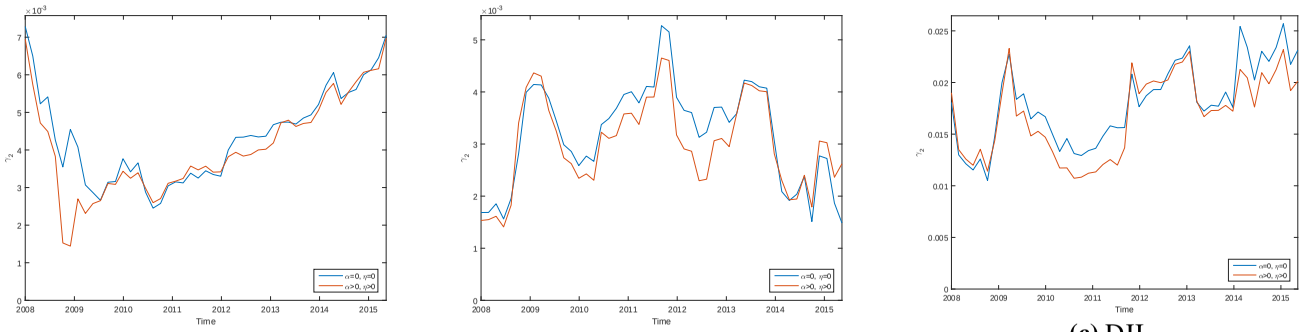
To quantify the effects that extremes in one market have on another market, we examine the branching ratio of extremes of the minimum magnitude in the markets, see (4.5). Amongst series and thresholds, an event triggers on average 0.07 to 0.48 events in another series, where the smallest amount of crashes is provoked between the NASDAQ and the S&P 500 index. The expected number of DJI crashes following a S&P 500 crash is the highest. The percentage of events induced by events in another series ranges between 11% and 90%.

Figure 4.2 shows the estimates of the spillover parameter γ_2 of the Hawkes models with and without a size-dependent triggering effect and predictable event sizes for the different series over time. It illustrates that for most series the estimates increase during crisis periods, and in particular during the financial crisis of 2008. The γ_2 estimates for the S&P 500 index are an exception, the estimates decrease during the 2008 crisis and increase quite steadily after 2011. While the γ_2 parameter for the other series mostly stays at a higher or equal level after the financial crisis, the γ_2 parameter for the NASDAQ deviates quite a bit and reaches the level estimated before the financial crisis.

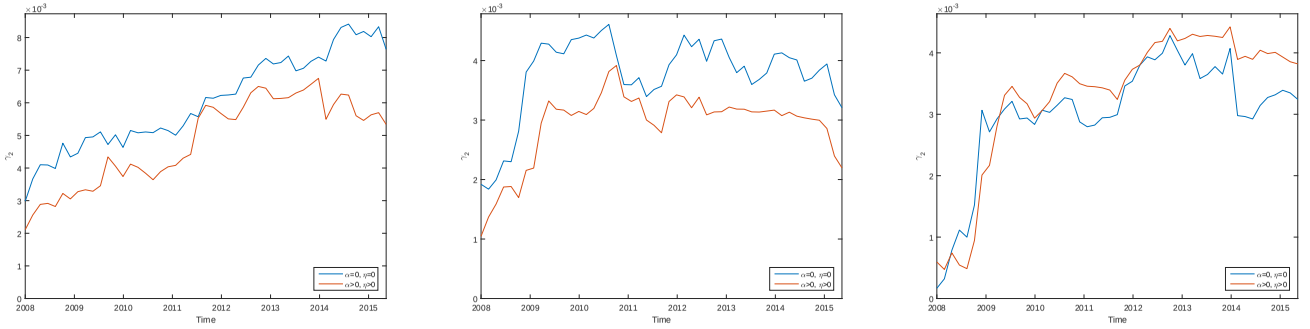
Looking at the parameter estimates, the impact of crashes in the S&P 500 index on extremes in other series and the impact of crashes in the NASDAQ on crashes in the S&P 500 index, seems quite substantial and hence should not be ignored. We perform an LM test to confirm the indicated cross-excitation.

Table 4.3 reports the mean, the minimum and the maximum of LM statistics for testing the null hypothesis of no cross-excitation ($\gamma_2 = 0$) across the estimation windows. In accordance with the parameter estimates, we find that crashes in the S&P 500 index amplify the probability that crashes as well as booms in the bond market and the exchange rate series occur. Thus, the LM test detects both stock-bond contagion as the flight-to-quality phenomenon, that is, the outflow of capital from stock markets to bond markets when the first are facing crisis periods. The evidence for the effect of crashes in the S&P 500 to extremes

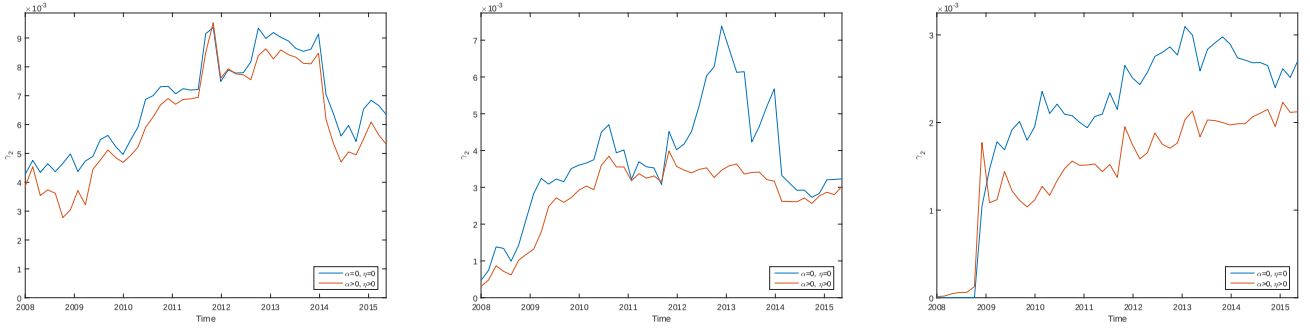
Figure 4.2: γ_2 estimates for the 95% threshold



(a) S&P **(b)** NDQ **(c)** DJI



(d) BND(-) **(e)** €/$(-) **(f)** £/$(-)



(g) BND(+) **(h)** €/$(+) **(i)** £/$(+)

Estimates of the spillover parameter γ_2 for the Hawkes models with and without a size-dependent triggering effect (α) and predictable event sizes (η).

Table 4.3: LM statistics dependence test

		$\alpha = 0$				$\alpha \neq 0$			
		$\eta = 0$		$\eta \neq 0$		$\eta = 0$		$\eta \neq 0$	
95%	S&P	7.12	[2.79 , 14.43]	10.15	[4.88 , 15.1]	6.45	[2.83 , 13.04]	5.01	[0.78 , 10.23]
	NDQ	4.34	[1.06 , 10.02]	4.01	[0.75 , 8.63]	4.15	[0.87 , 10.00]	3.27	[0.60 , 13.33]
	DJI	9.03	[5.61 , 17.43]	9.05	[5.82 , 17.55]	9.12	[5.84 , 17.70]	6.07	[2.86 , 12.65]
	BND(-)	24.29	[13.13 , 31.73]	23.84	[15.84 , 31.05]	7.99	[2.00 , 19.88]	6.89	[2.25 , 15.22]
	€/$(-)	9.42	[2.58 , 18.80]	7.98	[1.81 , 16.37]	6.48	[2.27 , 16.35]	7.68	[0.93 , 12.78]
	£/$(-)	8.13	[0.03 , 14.79]	6.51	[0.00 , 11.99]	6.86	[0.68 , 11.13]	2.08	[0.13 , 3.94]
	BND(+)	13.57	[7.12 , 19.51]	11.80	[0.00 , 17.07]	13.59	[7.80 , 18.54]	7.53	[2.60 , 13.83]
	€/$(+)	11.04	[0.79 , 16.40]	8.74	[0.78 , 13.37]	9.88	[0.63 , 16.3]	5.97	[0.94 , 9.56]
	£/$(+)	5.63	[0.00 , 9.39]	5.13	[0.00 , 8.44]	1.51	[0.00 , 3.41]	4.72	[0.02 , 9.28]
97.5%	S&P	3.15	[0.75 , 4.79]	3.13	[0.85 , 4.73]	2.50	[0.32 , 4.12]	2.28	[0.31 , 3.86]
	NDQ	4.76	[0.81 , 9.09]	4.81	[1.16 , 8.49]	4.51	[1.80 , 8.99]	4.14	[1.31 , 7.7]
	DJI	7.67	[5.22 , 11.25]	7.65	[5.49 , 11.19]	7.90	[5.06 , 11.35]	7.97	[5.59 , 10.53]
	BND(-)	16.64	[12.81 , 19.07]	16.27	[12.31 , 18.96]	10.24	[5.57 , 16.14]	10.00	[4.95 , 15.96]
	€/$(-)	8.00	[1.53 , 12.45]	6.93	[0.73 , 10.94]	8.21	[2.08 , 18.27]	8.57	[1.52 , 11.21]
	£/$(-)	7.10	[2.05 , 13.3]	7.73	[1.14 , 18.88]	7.69	[4.48 , 11.98]	7.75	[1.41 , 16.8]
	BND(+)	15.24	[5.05 , 22.98]	12.03	[4.33 , 23.18]	14.55	[5.07 , 21.48]	5.66	[2.00 , 10.63]
	€/$(+)	14.48	[3.37 , 19.00]	12.67	[3.32 , 17.29]	9.13	[5.05 , 12.65]	7.27	[4.18 , 10.69]
	£/$(+)	10.87	[0.00 , 16.64]	11.81	[0.00 , 18.26]	4.39	[0.14 , 7.37]	6.02	[0.26 , 9.36]

LM statistics for testing the null-hypothesis of no spillover effects from the NASDAQ to the S&P 500 index and from the S&P 500 index to the NASDAQ, the DJI index, the US bond index, the euro/dollar rate and the pound/dollar rate. The models are applied to extremes beyond the 95% and the 97.5% quantile to forecast over a period that starts at 1 January, 2008, and ends at 1 July, 2015, using a moving window of $4,000$ trading days that is updated every 40 trading days. In the Hawkes models with the parameter restriction $\alpha = 0$, the magnitude of events have no influence on the triggering subsequent events. In the Hawkes models with the parameter restriction $\eta = 0$, the history of the events has no influence on the magnitude of subsequent events. The Table presents the mean of the LM statistics across the estimation windows. The minimum and maximum of the statistics are shown in between brackets. The critical value corresponding to a 5% significance level is equal to 3.84.

in the US bond market is the most pronounced as the LM test statistics are the highest among these series.

The LM test indicates that, on average, spillover effects are significant, except for the spillover effects from the NASDAQ to the S&P 500 index when events beyond the 97.5% quantile are studied. Moreover, the spillover effects from the S&P 500 index to the DJI index and the US bond market are significant across all estimation windows. Looking at the LM statistics, the cross-excitation effects between the S&P 500 index and the NASDAQ seem the least powerful. Both findings are reflected in the parameter estimates in Tables 4.1 and 4.2.

From the parameter estimates and the LM tests, we conclude that, in-sample, the impact of the occurrence of crashes in the NASDAQ on the occurrence of crashes in the S&P 500 index and the impact of the occurrence of crashes in the S&P 500 index on the occurrence of extremes in other financial series is sizeable and important. We may thus expect that the Hawkes models that account for the cross-triggering phenomenon perform better out-of-sample than the Hawkes models that ignore it.

4.4 Residual analysis

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

$$\tau_i = \int_0^{t_i} \lambda(s)ds \tag{4.14}$$

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

$$\tau_i - \tau_{i-1} = \int_{t_{i-1}}^{t_i} \lambda(s)ds \tag{4.15}$$

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

our test is that the distribution of the residual process and the unit Poisson distribution are equal.

The KS tests are performed on the transformed times by applying the models to the extremes. The p-values of the tests are reported in Table 4.4. The p-values indicate that, on average, the extremes do not deviate from an event process specified by the models at a 5% level. Though some p-values are higher for the models without cross-excitation, they cannot be interpreted as a preference for these models.

4.5 Forecasting

4.5.1 Probability predictions

Traders, risk managers and regulators of financial markets can benefit from the accurate forecast of a crash. The probability of an extreme event occurring between $t - \delta t$ and t is given by (4.7). We evaluate the probability forecast of an extreme event beyond the 95% and the 97.5% quantile during the next day in stock, bond and FX markets from 1 January, 2008, to 1 July, 2015, using a moving window of $4,000$ trading days that is updated every 40 trading days. We estimate the thresholds and parameters within the moving window and start forecasting thereafter to prevent any look-ahead bias.

Figure 4.3 shows the predictions of the probability of extremes above the 95% threshold for the different series over time. It seems that the probability predictions produced by models with and without a size-dependent triggering effect and predictable event sizes deviate quite a bit, especially during crisis periods in which predictions of the extensive models are the highest. Comparing the models with and without spillover effects, the models that allow for series to cross-excite each other deliver the highest probability predictions during crisis periods.

To compare the probability forecasts, we compute the Quadratic Probability Score (QPS) and the Log Probability Score (LPS) for each model, that is

$$QPS = \frac{2}{T}\sum_{t=1}^{T}\left(\hat{p}(t) - \mathrm{I}(r(t) > u)\right)^2 \tag{4.16}$$

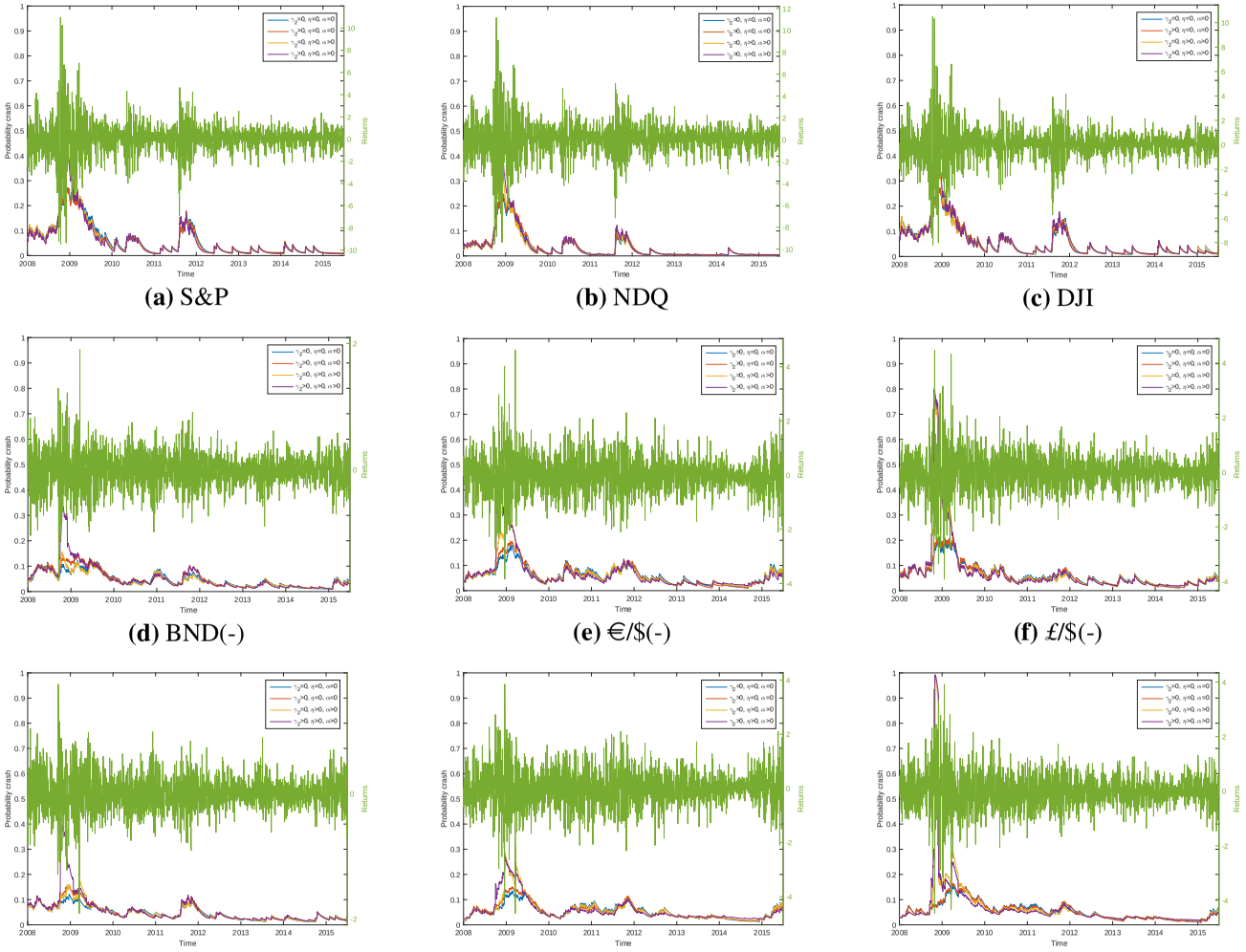
$$LPS = -\frac{1}{T}\sum_{t=1}^{T}\left[(1 - \mathrm{I}(r(t) > u))\log\left(1 - \hat{p}(t)\right) + \mathrm{I}(r(t) > u)\log\left(\hat{p}(t)\right)\right], \tag{4.17}$$

Table 4.4: Results residual analysis

		$\alpha = 0$				$\alpha > 0$			
		$\eta = 0$		$\eta > 0$		$\eta = 0$		$\eta > 0$	
		$\gamma_2 = 0$	$\gamma_2 > 0$	$\gamma_2 = 0$	$\gamma_2 > 0$	$\gamma_2 = 0$	$\gamma_2 > 0$	$\gamma_2 = 0$	$\gamma_2 > 0$
95%	S&P	0.68	0.49	0.76	0.44	0.46	0.77	0.48	0.84
	NDQ	0.43	0.17	0.49	0.22	0.43	0.17	0.50	0.21
	DJI	0.69	0.56	0.70	0.70	0.66	0.48	0.70	0.60
	BND(-)	0.72	0.66	0.57	0.44	0.67	0.60	0.65	0.66
	€/$(-)	0.64	0.83	0.63	0.82	0.79	0.74	0.77	0.53
	£/$(-)	0.26	0.22	0.23	0.12	0.46	0.29	0.36	0.17
	BND(+)	0.60	0.60	0.56	0.50	0.65	0.60	0.64	0.57
	€/$(+)	0.61	0.78	0.60	0.73	0.77	0.70	0.77	0.57
	£/$(+)	0.28	0.22	0.25	0.13	0.43	0.36	0.39	0.20
97.5%	S&P	0.39	0.37	0.39	0.27	0.55	0.54	0.56	0.49
	NDQ	0.53	0.18	0.58	0.26	0.64	0.21	0.67	0.29
	DJI	0.40	0.42	0.39	0.42	0.47	0.50	0.47	0.57
	BND(-)	0.72	0.33	0.72	0.28	0.65	0.34	0.65	0.34
	€/$(-)	0.59	0.82	0.59	0.77	0.69	0.65	0.69	0.48
	£/$(-)	0.31	0.35	0.30	0.24	0.49	0.38	0.48	0.29
	BND(+)	0.40	0.62	0.40	0.62	0.46	0.62	0.48	0.62
	€/$(+)	0.56	0.33	0.58	0.25	0.63	0.24	0.65	0.18
	£/$(+)	0.27	0.51	0.26	0.38	0.85	0.85	0.87	0.72

The p-values of the Kolmogorov-Smirnov tests performed on the transformed times $\{\tau_i\}$ specified by the models. The null-hypothesis of the test states that the transformed times $\{\tau_i\}$ are distributed according to a homogeneous Poisson process with intensity 1. The models are applied to extremes beyond the 95% and the 97.5% quantile to forecast over a period that starts at 1 January, 2008, and ends at 1 July, 2015, using a moving window of $4,000$ trading days that is updated every 40 trading days. The Table presents the mean of the p-values across the estimation windows. For the S&P 500 index, γ_2 controls the spillover effect from crashes in the NASDAQ, while for the DJI index, the US bond index, the euro/dollar rate and the pound/dollar this parameter controls the spillover effect from crashes in the S&P 500. In the Hawkes models with the parameter restriction $\alpha = 0$, the magnitude of events have no influence on the triggering subsequent events. In the Hawkes models with the parameter restriction $\eta = 0$, the history of the events has no influence on the magnitude of subsequent events.

Figure 4.3: Probability predictions for the 95% threshold



(a) S&P **(b)** NDQ **(c)** DJI

(d) BND(-) **(e)** €/$(-) **(f)** £/$(-)

(g) BND(+) **(h)** €/$(+) **(i)** £/$(+)

Probability predictions of crashes the next day in the S&P 500 index, the NASDAQ and the Dow Jones Industrial Average, and crashes and booms in the US bond index, the euro/dollar rate and the pond/dollar rate beyond the 95% quantile from 1 January, 2008, to 1 July, 2015. The predictions are made by Hawkes models with and without cross-excitation (γ_2), a size-dependent triggering effect (α) and predictable event sizes (η).

Table 4.5: Probability scores 95%

		$\alpha = 0$				$\alpha > 0$			
		$\eta = 0$		$\eta > 0$		$\eta = 0$		$\eta > 0$	
		$\gamma_2 = 0$	$\gamma_2 > 0$	$\gamma_2 = 0$	$\gamma_2 > 0$	$\gamma_2 = 0$	$\gamma_2 > 0$	$\gamma_2 = 0$	$\gamma_2 > 0$
S&P	QPS	0.1108	0.1108	0.1106	0.1107	0.1098	0.1100	0.1098	0.1099
	LPS	0.2059	0.2062	0.2052	0.2056	0.2040	0.2046	0.2032	0.2038
NDQ	QPS	0.0615	0.0616	0.0611	0.0611	0.0614	0.0615	0.0609	0.0611
	LPS	0.1213	0.1207	0.1204	0.1196	0.1210	0.1202	0.1198	0.1191
DJI	QPS	0.1057	0.1059	0.1055	0.1057	0.1052	0.1053	0.1050	0.1051
	LPS	0.1982	0.1982	0.1979	0.1978	0.1971	0.1970	0.1965	0.1964
BND(-)	QPS	0.0922	0.0917	0.0925	0.0921	0.0920	0.0925	0.0919	0.0931
	LPS	0.1861	0.1841	0.1886	0.1879	0.1855	0.1851	0.1853	0.1854
€/$(-)	QPS	0.1278	0.1275	0.1278	0.1275	0.1268	0.1269	0.1268	0.1280
	LPS	0.2443	0.2438	0.2441	0.2438	0.2420	0.2423	0.2418	0.2435
£/$(-)	QPS	0.1263	0.1259	0.1263	0.1259	0.1286	0.1283	0.1288	0.1286
	LPS	0.2387	0.2381	0.2391	0.2383	0.2397	0.2393	0.2402	0.2402
BND(+)	QPS	0.0878	0.0874	0.0878	0.0883	0.0879	0.0885	0.0881	0.0889
	LPS	0.1795	0.1779	0.1794	0.1829	0.1797	0.1793	0.1800	0.1797
€/$(+)	QPS	0.1134	0.1130	0.1133	0.1129	0.1136	0.1129	0.1137	0.1129
	LPS	0.2223	0.2213	0.2223	0.2214	0.2224	0.2211	0.2225	0.2215
£/$(+)	QPS	0.1194	0.1191	0.1195	0.1191	0.1309	0.1321	0.1313	0.1322
	LPS	0.2271	0.2262	0.2272	0.2263	0.2456	0.2524	0.2475	0.2521

QPS and LPS of the probability predictions of crashes the next day in the S&P 500 index, the NASDAQ and the Dow Jones Industrial Average, and crashes and booms in the US bond index, the euro/dollar rate and the pond/dollar rate beyond the 95% quantile from 1 January, 2008, to 1 July, 2015. The QPS and LPS are given in (4.16) and (4.17). The QPS and LPS range respectively from 0 to 1 and from 0 to ∞, with 0 indicating perfect accuracy.

where t is a day and T is the total number of days. Here p_t represents the probability forecast of an extreme on day t, and $\mathrm{I}(r(t) > u)$ is an indicator function taking the value 1 when the return at time t exceeds the threshold and the value 0 otherwise. The QPS and LPS range respectively from 0 to 1 and from 0 to ∞, with 0 indicating perfect accuracy. When the QPS or the LPS are higher, the probability forecasts deviate more from a binary variable indicating the occurrence of events. The LPS punishes large deviations heavier than small deviations.

The QPS and the LPS of the probability predictions of extremes above the 95% and the 97.5% threshold made by the various Hawkes models, are displayed in Table 4.5 and Table 4.6, respectively. When models are employed in which the impact of events on the rate at which new events arrive is not size-dependent ($\alpha = 0$), we observe that the QPS and LPS are lower when spillover effects are included in the models for most series. When models are employed in which the impact of events is size-dependent, the QPS and the LPS of the

Table 4.6: Probability scores 97.5%

		$\alpha = 0$				$\alpha > 0$			
		$\eta = 0$		$\eta > 0$		$\eta = 0$		$\eta > 0$	
		$\gamma_2 = 0$	$\gamma_2 > 0$	$\gamma_2 = 0$	$\gamma_2 > 0$	$\gamma_2 = 0$	$\gamma_2 > 0$	$\gamma_2 = 0$	$\gamma_2 > 0$
S&P	QPS	0.0634	0.0634	0.0633	0.0633	0.0635	0.0636	0.0634	0.0635
	LPS	0.1219	0.1217	0.1219	0.1218	0.1217	0.1219	0.1217	0.1218
NDQ	QPS	0.0357	0.0357	0.0354	0.0354	0.0357	0.0358	0.0354	0.0353
	LPS	0.0733	0.0720	0.0724	0.0712	0.0732	0.0718	0.0723	0.0709
DJI	QPS	0.0602	0.0597	0.0601	0.0597	0.0599	0.0590	0.0597	0.0588
	LPS	0.1161	0.1138	0.1161	0.1137	0.1152	0.1126	0.1150	0.1123
BND(-)	QPS	0.0462	0.0460	0.0462	0.0462	0.0461	0.0463	0.0461	0.0463
	LPS	0.1060	0.1032	0.1060	0.1075	0.1054	0.1037	0.1054	0.1037
€/$(-)	QPS	0.0731	0.0729	0.0732	0.0729	0.0726	0.0724	0.0724	0.0728
	LPS	0.1555	0.1549	0.1556	0.1553	0.1545	0.1539	0.1540	0.1545
£/$(-)	QPS	0.0730	0.0725	0.0730	0.0725	0.0733	0.0752	0.0737	0.0750
	LPS	0.1493	0.1479	0.1495	0.1481	0.1478	0.1493	0.1484	0.1495
BND(+)	QPS	0.0452	0.0447	0.0452	0.0447	0.0458	0.0449	0.0460	0.0452
	LPS	0.1041	0.1015	0.1039	0.1015	0.1057	0.1017	0.1068	0.1025
€/$(+)	QPS	0.0614	0.0609	0.0614	0.0609	0.0617	0.0618	0.0617	0.0619
	LPS	0.1368	0.1342	0.1368	0.1343	0.1365	0.1354	0.1365	0.1357
£/$(+)	QPS	0.0635	0.0631	0.0635	0.0631	0.0656	0.0646	0.0657	0.0646
	LPS	0.1362	0.1341	0.1362	0.1339	0.1360	0.1342	0.1360	0.1350

QPS and LPS of the probability predictions of crashes the next day in the S&P 500 index, the NASDAQ and the Dow Jones Industrial Average, and crashes and booms in the US bond index, the euro/dollar rate and the pond/dollar rate beyond the 97.5% quantile from 1 January, 2008, to 1 July, 2015. The QPS and LPS are given in (4.16) and (4.17). The QPS and LPS range respectively from 0 to 1 and from 0 to ∞, with 0 indicating perfect accuracy.

models are closer, and most of the QPS and LPS of models with predictable events increase when allowing for cross-excitation.

The QPS and LPS of the models, in which the size of events and the conditional intensity affect the one and/or the other, are lower for crashes in the stock markets and €/$-rate. For the other series, the QPS and LPS increase when one allows the size of past events to influence the rate at which new events arrive.

Following the approach of Dijk and Franses (2003), we use the test of Diebold and Mariano (1995) (DM test) based on an average weighted loss differential to check whether the probability forecasts of models are significantly more accurate than the probability forecasts of other models. We modify the DM test statistic using the weight function $w(t) = (\mathrm{I}(r(t) < u) - q)$, where q is the threshold quantile, such that the loss function becomes

$$\mathrm{L}(e_q^k(t)) = (\mathrm{I}(r(t) < u) - q)e_q^k(t) \tag{4.18}$$

where $e_q^k(t) = (\hat{p}(t) - \mathrm{I}(r(t) > u))^2$. The loss function (4.18) rewards correct probability forecasts of extreme observations more heavily than it penalizes false signals. The use of the quadratic loss function $e_q^k(t)$ is particularly convenient, as the test does not have to be adjusted to account for parameter uncertainty arising from the models, which contain unknown parameters that have to be estimated before forecasts can be constructed. Furthermore, as we use a moving window to estimate the parameters of the model, the models are estimated over a finite sample period, such that the asymptotic normal distribution remains valid for nested models (Giacomini and White, 2006).

The DM statistics comparing probability predictions of models with and without cross-excitation are given in Table 4.7. The DM statistics comparing probability predictions of models with and without influence of the event sizes on the conditional intensity, and models with and without influence of the history of the event process on the event sizes are given in Table 4.8.

From Table 4.7 we conclude that the probability predictions of models with spillovers are significantly more accurate for the extremes examined at a 5% level, except for the S&P 500 index. The choice of models matters for the DJI index and for booms in the bond market and the exchange rates.

Table 4.7: DM statistics testing cross-excitation probability predictions

	95%				97.5%			
	$\alpha = 0$		$\alpha > 0$		$\alpha = 0$		$\alpha > 0$	
	$\eta = 0$	$\eta > 0$	$\eta = 0$	$\eta > 0$	$\eta = 0$	$\eta > 0$	$\eta = 0$	$\eta > 0$
S&P	-7.63	-7.61	-9.31	-9.03	-3.98	-3.48	-6.69	-6.73
NDQ	6.75	6.48	5.94	5.94	5.36	5.35	4.84	4.79
DJI	0.81	0.66	4.23	3.41	3.70	3.59	5.41	5.26
BND(-)	5.07	5.04	4.81	4.45	4.02	4.06	4.03	4.03
€/$(-)	3.65	2.89	2.75	2.14	3.38	2.85	2.44	2.02
£/$(-)	4.55	4.16	2.48	2.10	5.38	5.00	3.64	3.23
BND(+)	5.64	1.35	3.76	3.65	3.96	3.69	3.30	3.34
€/$(+)	2.78	2.03	1.36	1.76	3.47	3.24	1.85	2.23
£/$(+)	2.67	-0.08	1.40	-2.16	3.67	2.51	-0.22	-1.86

DM statistics based on the weighted loss function (4.18) comparing probability predictions of models with and without cross-excitation. The models predict the probability of crashes the next day in the S&P 500 index, the NASDAQ and the Dow Jones Industrial Average, and crashes and booms in the US bond index, the euro/dollar rate and the pond/dollar rate beyond the 97.5% quantile from 1 January, 2008, to 1 July, 2015. The critical value corresponding to a 5% significance level is equal to ± 1.96 (asymptotic $\mathcal{N}(0,1)$-distribution).

According to the the DM statistics in Table 4.8, the probability predictions of models with predictable event sizes are significantly more accurate for the extremes examined at a 5% level. The DM statistics show that, overall, the incorporation of a triggering effect that depends on the size of events ($\alpha > 0$) leads to more accurate probability predictions. For the incorporation of a distribution of the event sizes that depends on the history of the event process ($\eta > 0$), the results are mixed. Only for the NASDAQ, we find convincing evidence that the predictability of event sizes adds to the models' performance forecasting the probability of crashes in the series.

Our main result in this section is that, all in all, the Hawkes models that account for cross-triggering effects deliver more accurate forecasts of the occurrence of extremes in financial markets than the Hawkes models that ignore the existence of spillovers.

4.5.2 Value-at-Risk

To avoid large losses due to price changes and to meet regulatory requirements that limit exposure to risk, investors in financial markets use Value-at-Risk to quantify the maximum

Table 4.8: DM statistics testing the influence on/of sizes probability predictions

		η ($\alpha = 0$)		α ($\eta = 0$)		η ($\alpha > 0$)		α ($\eta > 0$)	
		$\gamma_2 = 0$	$\gamma_2 > 0$	$\gamma_2 = 0$	$\gamma_2 > 0$	$\gamma_2 = 0$	$\gamma_2 > 0$	$\gamma_2 = 0$	$\gamma_2 > 0$
95%	S&P	-1.38	0.19	5.59	5.58	3.31	3.97	5.25	5.09
	NDQ	2.58	2.79	4.29	4.72	3.30	3.13	4.25	4.61
	DJI	-1.21	-1.28	4.42	4.97	1.90	-0.39	4.58	4.75
	BND(-)	2.29	2.03	-0.90	2.21	0.19	1.58	-1.78	1.36
	€/$(-)	1.71	-2.48	4.21	4.56	3.82	2.49	4.15	4.05
	£/$(-)	-4.40	-2.90	4.85	4.94	-2.01	-0.40	4.84	4.87
	BND(+)	1.13	-3.04	3.41	2.96	0.97	1.55	2.15	3.68
	€/$(+)	2.99	-0.93	3.46	3.85	2.53	2.35	3.44	3.75
	£/$(+)	-5.15	-5.25	3.04	3.11	-2.85	-6.58	3.00	2.54
97.5%	S&P	-1.71	-1.70	4.34	4.08	3.06	2.91	4.39	4.32
	NDQ	3.25	2.99	3.12	3.90	2.96	0.93	3.48	3.88
	DJI	-0.79	-0.68	3.72	4.69	1.69	2.66	3.84	4.59
	BND(-)	-2.11	0.47	-1.04	2.30	-1.33	2.04	-1.04	2.22
	€/$(-)	-0.59	-1.66	3.24	3.93	3.68	2.38	3.44	3.59
	£/$(-)	-3.28	-3.06	4.66	4.65	2.99	-5.26	4.63	4.68
	BND(+)	1.79	-1.29	2.39	2.25	-1.46	2.14	0.18	2.24
	€/$(+)	2.31	-0.09	3.45	3.37	1.40	3.38	3.45	3.62
	£/$(+)	1.26	-2.79	3.17	3.13	2.92	-2.28	3.27	3.06

DM statistics based on the weighted loss function (4.18) comparing probability predictions of models with and without influence of the the history of the events on the magnitude of subsequent events (η) and models with and without influence of the magnitude of events triggering subsequent events (α). of crashes the next day in the S&P 500 index, the NASDAQ and the Dow Jones Industrial Average, and crashes and booms in the US bond index, the euro/dollar rate and the pond/dollar rate beyond the 97.5% quantile from 1 January, 2008, to 1 July, 2015. The critical value corresponding to a 5% significance level is equal to ± 1.96 (asymptotic $\mathcal{N}(0,1)$-distribution).

of returns associated with a certain confidence level. The $VaR_\alpha(t)$ is defined as the α-th quantile of a distribution at time t which is the solution to $\Pr(m \geq q|\boldsymbol{\theta}; \mathcal{H}_t) = 1 - \alpha$. As $\Pr(m \geq q|\mathcal{H}_t) = \Pr(N(t) - N(t - \delta t) > 0|\boldsymbol{\theta}; \mathcal{H}_t) \times \Pr(m \geq q|m \geq u; \boldsymbol{\theta}; \mathcal{H}_t)$ which are given in (4.7) and (4.4), we have that

$$VaR_\alpha(t) = u + \frac{\sigma(t)}{\xi}\left(\left(\frac{1-\alpha}{1 - \exp\left(-\int_{t-\delta t}^{t} \lambda(s|\boldsymbol{\theta}; \mathcal{H}_s)ds\right)}\right)^{-\xi} - 1\right), \tag{4.19}$$

where $\sigma(t) = \phi + \frac{\eta}{\gamma_1}(\lambda(t|\boldsymbol{\theta}; \mathcal{H}_t) - \mu)$ and $\lambda(t|\boldsymbol{\theta}; \mathcal{H}_t)$ is given by (4.8).

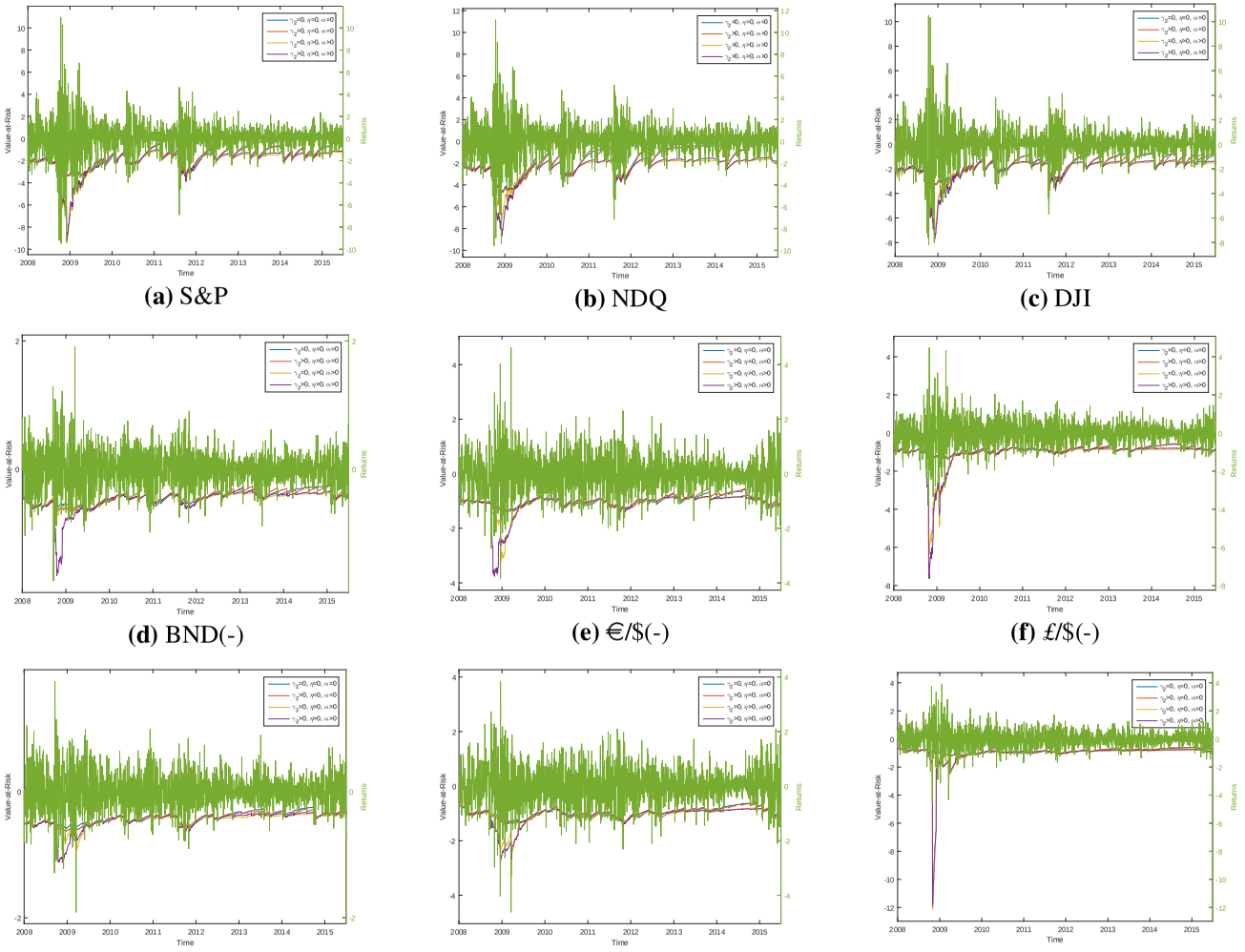
Figure 4.4 shows the VaR predictions for the 95% confidence level from 1 January, 2008, to 1 July, 2015. Like the probability predictions, the models with a size-dependent triggering effect and predictable event sizes produce higher VaR predictions during crisis periods than models without. Also, in accordance with the probability predictions, the models that allow for series to cross-excite each other, deliver the highest VaR predictions during crisis periods.

We evaluate the VaR forecasts (4.19) by comparing the estimated conditional VaR for a one-day horizon, that is $\delta t = 1$, for the confidence levels $\alpha = \{0.95, 0.975\}$ with the actual returns. To quantify the quality of the VaR forecasts we use the unconditional coverage (LR_{uc}) test, the independence (LR_{ind}) test and the conditional coverage (LR_{cc}) test of Christoffersen (1998) and a Dynamic Quantile (DQ) test in the line of Engle and Manganelli (2004). Furthermore, to compare the VaR forecasts of different models, we define an asymmetric loss function as Giacomini and Komunjer (2005) and test whether the loss differentials are significant using the test of Diebold and Mariano (1995).

With the unconditional coverage (LR_{uc}) test we check whether the actual fraction of VaR violations equals the theoretical proportion of $1 - \alpha$. We let $x_\alpha(t)$ denote an indicator function taking the value 1 when the extreme is larger than the $VaR_\alpha(t)$ and the value 0 otherwise. If the models are correct $\mathrm{E}[x_\alpha(t)] = 1 - \alpha$. The independence ($LR_{ind}$) test concerns independence among the VaR violations such that a violation today has no influence on the probability of a violation tomorrow. A combination of the LR_{uc} and LR_{ind} is given by the conditional coverage (LR_{cc}) test. The LR_{uc}, LR_{ind} and LR_{cc} test statistics for the confidence levels 95% and 97.5% are given in Table 4.9 and Table 4.10.

Tables 4.9 and 4.10 show that at the 95% and 97.5% confidence levels, the unconditional and thereby conditional coverage of the VaR predictions for investors with long positions in

Figure 4.4: $VaR_{0.95}$ **predictions**



(a) S&P **(b)** NDQ **(c)** DJI

(d) BND(-) **(e)** €/$(-) **(f)** £/$(-)

(g) BND(+) **(h)** €/$(+) **(i)** £/$(+)

VaR predictions at the 95% confidence level for investors with a long position in the S&P 500 index, the NASDAQ and the Dow Jones Industrial Average, and investors with a short and a long position in the US bond index, the euro/dollar rate and the pond/dollar rate from 1 January, 2008, to 1 July, 2015. The predictions are made by Hawkes models with and without cross-excitation (γ_2), a size-dependent triggering effect (α) and predictable event sizes (η).

Table 4.9: LR_{uc}**,** LR_{ind} **and** LR_{cc} **test statistics** $VaR_{0.95}$

		$\alpha = 0$				$\alpha > 0$			
		$\eta = 0$		$\eta > 0$		$\eta = 0$		$\eta > 0$	
		$\gamma_2 = 0$	$\gamma_2 > 0$	$\gamma_2 = 0$	$\gamma_2 > 0$	$\gamma_2 = 0$	$\gamma_2 > 0$	$\gamma_2 = 0$	$\gamma_2 > 0$
S&P	LR_{uc}	38.37	51.18	0.18	3.00	30.07	46.34	0.09	3.00
	LR_{ind}	1.58	1.48	0.54	0.77	0.27	2.03	0.70	0.24
	LR_{cc}	39.95	52.66	0.73	3.77	30.34	48.37	0.79	3.25
NDQ	LR_{uc}	231.48	398.43	1.86	1.86	224.71	376.44	2.50	2.50
	LR_{ind}	28.04	64.20	1.01	0.15	26.44	56.72	0.84	0.84
	LR_{cc}	259.51	462.63	2.86	2.01	251.15	433.15	3.34	3.34
DJI	LR_{uc}	8.39	12.67	0.87	0.26	8.39	8.39	1.86	0.38
	LR_{ind}	0.08	0.08	0.23	0.05	0.08	0.23	0.15	1.69
	LR_{cc}	8.47	12.75	1.10	0.30	8.47	8.62	2.01	2.07
BND(-)	LR_{uc}	0.11	0.04	2.50	6.09	0.40	0.16	0.68	2.50
	LR_{ind}	0.01	0.12	0.07	0.32	2.08	0.07	0.81	0.07
	LR_{cc}	0.12	0.16	2.56	6.41	2.48	0.23	1.50	2.56
€/$(-)	LR_{uc}	4.52	7.32	2.36	2.67	4.11	4.94	0.11	1.07
	LR_{ind}	0.45	1.18	0.96	1.68	0.52	2.85	0.62	2.60
	LR_{cc}	4.97	8.49	3.31	4.35	4.63	7.79	0.73	3.67
£/$(-)	LR_{uc}	8.39	10.74	1.29	0.18	7.32	7.32	0.00	0.01
	LR_{ind}	6.77	4.30	6.91	2.40	4.34	5.77	1.84	1.70
	LR_{cc}	15.16	15.04	8.20	2.58	11.65	13.08	1.84	1.72
BND(+)	LR_{uc}	0.01	0.26	2.50	0.26	0.00	0.26	1.57	2.50
	LR_{ind}	0.30	1.43	0.60	0.56	0.88	1.43	0.39	1.60
	LR_{cc}	0.31	1.69	3.10	0.82	0.88	1.69	1.97	4.09
€/$(+)	LR_{uc}	2.06	2.06	0.40	0.11	1.07	1.07	0.01	0.01
	LR_{ind}	0.03	0.03	0.04	0.62	0.00	0.18	0.21	0.30
	LR_{cc}	2.09	2.09	0.44	0.73	1.07	1.26	0.22	0.31
£/$(+)	LR_{uc}	3.00	4.11	0.40	0.11	2.06	3.35	0.28	0.40
	LR_{ind}	0.56	0.21	0.52	0.34	1.16	0.63	3.10	3.25
	LR_{cc}	3.57	4.32	0.92	0.45	3.22	3.99	3.38	3.65

Statistics for the unconditional coverage, independence and conditional coverage tests on the VaR predictions at a 95% confidence level produced by the different Hawkes models. The models predict the maximum of negative returns associated with a 95% confidence level the next day in the stock market, bond and exchange rate series and the maximum of positive returns associated with 95% confidence level in the bond market and exchange rate series from 1 January, 2008, to 1 July, 2015. The critical value corresponding to a 5% significance level is equal to 3.84 for the unconditional coverage test and the independence test (asymptotic $\chi^2(1)$-distribution) and 5.99 for the conditional coverage test (asymptotic $\chi^2(2)$-distribution).

Table 4.10: LR_{uc}**,** LR_{ind} **and** LR_{cc} **test statistics** $VaR_{0.975}$

		$\alpha = 0$				$\alpha > 0$			
		$\eta = 0$		$\eta > 0$		$\eta = 0$		$\eta > 0$	
		$\gamma_2 = 0$	$\gamma_2 > 0$	$\gamma_2 = 0$	$\gamma_2 > 0$	$\gamma_2 = 0$	$\gamma_2 > 0$	$\gamma_2 = 0$	$\gamma_2 > 0$
S&P	LR_{uc}	32.35	51.48	1.63	4.92	29.73	40.73	1.30	2.00
	LR_{ind}	2.63	0.89	0.84	0.31	0.73	1.70	0.32	0.43
	LR_{cc}	34.98	52.38	2.47	5.23	30.45	42.44	1.62	2.43
NDQ	LR_{uc}	91.38	258.59	2.22	1.06	85.43	235.56	2.22	1.78
	LR_{ind}	1.61	82.98	1.59	1.05	1.18	89.43	1.59	1.28
	LR_{cc}	92.99	341.56	3.80	2.10	86.61	324.98	3.80	3.06
DJI	LR_{uc}	39.28	45.22	3.31	2.00	28.45	31.03	3.31	1.00
	LR_{ind}	0.25	0.09	0.00	0.03	0.81	0.64	0.00	0.28
	LR_{cc}	39.53	45.31	3.31	2.03	29.27	31.67	3.31	1.28
BND(-)	LR_{uc}	2.40	66.73	2.40	8.99	3.81	66.73	3.81	66.73
	LR_{ind}	0.66	17.75	0.66	24.64	0.44	20.38	0.44	20.38
	LR_{cc}	3.06	84.48	3.06	33.62	4.25	87.12	4.25	87.12
€/$(-)	LR_{uc}	9.77	16.03	2.84	2.84	11.42	18.07	1.30	3.81
	LR_{ind}	1.81	0.21	3.80	0.58	3.04	0.77	2.56	0.44
	LR_{cc}	11.58	16.24	6.64	3.42	14.46	18.84	3.86	4.25
£/$(-)	LR_{uc}	40.73	75.86	6.82	4.92	24.77	49.89	1.63	2.00
	LR_{ind}	6.12	2.85	4.32	1.36	5.59	0.34	0.84	2.22
	LR_{cc}	46.85	78.71	11.14	6.28	30.37	50.23	2.47	4.21
BND(+)	LR_{uc}	1.00	0.34	0.33	1.78	0.52	0.09	0.77	2.71
	LR_{ind}	0.10	0.21	0.00	0.04	0.17	0.31	0.94	0.09
	LR_{cc}	1.10	0.55	0.33	1.82	0.69	0.40	1.71	2.79
€/$(+)	LR_{uc}	2.40	2.00	2.00	1.30	3.81	2.84	3.31	1.63
	LR_{ind}	0.49	0.43	0.43	0.32	0.68	0.55	0.61	0.05
	LR_{cc}	2.89	2.43	2.43	1.62	4.49	3.38	3.92	1.68
£/$(+)	LR_{uc}	2.40	4.92	1.30	1.30	2.40	2.84	0.74	1.00
	LR_{ind}	0.66	1.36	0.94	0.07	0.01	0.58	0.23	0.28
	LR_{cc}	3.06	6.28	2.23	1.37	2.42	3.42	0.98	1.28

Statistics for the unconditional coverage, independence and conditional coverage tests on the VaR predictions at a 97.5% confidence level produced by the different Hawkes models. The models predict the maximum of negative returns associated with a 95% confidence level the next day in the stock market, bond and exchange rate series and the maximum of positive returns associated with 95% confidence level in the bond market and exchange rate series from 1 January, 2008, to 1 July, 2015. The critical value corresponding to a 5% significance level is equal to 3.84 for the unconditional coverage test and the independence test (asymptotic $\chi^2(1)$-distribution) and 5.99 for the conditional coverage test (asymptotic $\chi^2(2)$-distribution).

the stock markets and the exchange rates, are a problem whenever the influence of the history of the event process on the sizes of subsequent events is omitted ($\eta = 0$). The unconditional and conditional coverage test statistics are particularly high for the stock market indices and the NASDAQ. The models that include the predictability of event sizes ($\eta > 0$) lead to lower statistics, which with a few exceptions are insignificant at a 5% level. In case $\eta = 0$, the Hawkes models also do not pass the independence test at a 5% level for crashes in the NASDAQ and the £/$-rate. At the the 95% confidence level, the VaR violations are independent for all other series. At the 97.5% confidence level, also the VaR predictions for investors with a long position in the US bond market, are dependent.

Both the effect of cross-excitation (the parameter γ_2) and the effect of the size of events on the triggering intensity (the parameter α), are inconclusive for most series. However, for investors with a long position in the US bond market, the incorporation of cross-excitation in the models leads to VaR forecasts, at a 97.5% confidence level, that are of lower quality, failing the tests of Christoffersen (1998) at a 5% significance level.

Besides the independence test of Christoffersen (1998), we use the Dynamic Quantile (DQ) test of Engle and Manganelli (2004) to verify whether VaR violations are predictable from the history of the event process. If the models are correct, VaR violations before time t should have no explanatory power for a VaR violation at t. The DQ test is implemented as in Berkowitz et al. (2011), that is the errors $\epsilon(t)$ from the regression

$$x_\alpha(t) = \beta_0 + \beta_1 x_\alpha(t-1) + \epsilon(t) \tag{4.20}$$

follow a logistic distribution. In this way we account for heteroscedasticity due to the binary nature of the variable $x_\alpha(t)$. The null hypothesis that $\beta_1 = 0$ is tested with a Likelihood Ratio test.

The DQ test statistics for the VaR predictions at the confidence levels 95% and 97.5% are given in Table 4.11. Almost all test statistics are insignificant. In accordance with the independence test of Christoffersen (1998), the DQ test indicates that the models in which event sizes are unpredictable fail to employ all information in the history of the event process to produce VaR predictions for investors with a long position in the NASDAQ and the £/$-rate. Also, again, the models that allow for cross-excitation produce auto-correlated forecasts

Table 4.11: DQ test statistics

		$\alpha = 0$				$\alpha > 0$			
		$\eta = 0$		$\eta > 0$		$\eta = 0$		$\eta > 0$	
		$\gamma_2 = 0$	$\gamma_2 > 0$	$\gamma_2 = 0$	$\gamma_2 > 0$	$\gamma_2 = 0$	$\gamma_2 > 0$	$\gamma_2 = 0$	$\gamma_2 > 0$
95%	S&P	1.58	1.48	0.54	0.77	0.27	2.03	0.70	0.24
	NDQ	28.04	64.20	1.01	0.15	26.44	56.72	0.84	0.84
	DJI	0.08	0.08	0.23	0.05	0.08	0.23	0.15	1.69
	BND(-)	0.01	0.12	0.07	0.32	2.08	0.07	0.81	0.07
	€/$(-)	0.45	1.18	0.96	1.68	0.52	2.85	0.62	2.60
	£/$(-)	6.77	4.30	6.91	2.40	4.34	5.77	1.84	1.70
	BND(+)	0.30	1.43	0.60	0.56	0.88	1.43	0.39	1.60
	€/$(+)	0.03	0.03	0.04	0.62	0.00	0.18	0.21	0.30
	£/$(+)	0.56	0.21	0.52	0.34	1.16	0.63	3.10	3.25
97.5%	S&P	2.63	0.89	0.84	0.31	0.73	1.70	0.32	0.43
	NDQ	1.61	82.98	1.59	1.05	1.18	89.43	1.59	1.28
	DJI	0.25	0.09	0.00	0.03	0.81	0.64	0.00	0.28
	BND(-)	0.66	17.75	0.66	24.64	0.44	20.38	0.44	20.38
	€/$(-)	1.81	0.21	3.80	0.58	3.04	0.77	2.56	0.44
	£/$(-)	6.12	2.85	4.32	1.36	5.59	0.34	0.84	2.22
	BND(+)	0.10	0.21	0.00	0.04	0.17	0.31	0.94	0.09
	€/$(+)	0.49	0.43	0.43	0.32	0.68	0.55	0.61	0.05
	£/$(+)	0.66	1.36	0.94	0.07	0.01	0.58	0.23	0.28

Statistics for the Dynamic Quantile test on the VaR predictions at a 95% and 97.5% confidence level produced by the different Hawkes models. The models predict the maximum of negative returns associated with a 95% confidence level the next day in the stock market, bond and exchange rate series and the maximum of positive returns associated with 95% confidence level in the bond market and exchange rate series from 1 January, 2008, to 1 July, 2015. The critical value corresponding to a 5% significance level is equal to ± 1.96 (asymptotic $\mathcal{N}(0,1)$-distribution).

Table 4.12: DM statistics testing cross-excitation VaR predictions

		95%				97.5%			
		$\alpha = 0$		$\alpha > 0$		$\alpha = 0$		$\alpha > 0$	
		$\eta = 0$	$\eta > 0$	$\eta = 0$	$\eta > 0$	$\eta = 0$	$\eta > 0$	$\eta = 0$	$\eta > 0$
S&P	66.52	76.87	75.30	64.66	81.39	48.84	88.39	54.47	
NDQ	-1.91	-1.16	-3.51	-5.87	7.67	8.56	5.00	1.69	
DJI	1.77	7.71	-1.51	1.78	-1.69	4.24	-5.78	-4.97	
BND(-)	-2.46	-14.66	-8.69	-10.01	13.15	-15.63	7.22	7.51	
€/$(-)	15.58	3.15	7.28	-3.11	20.34	3.01	16.12	-1.70	
£/$(-)	19.50	0.54	15.85	-0.87	18.36	-4.57	15.19	1.25	
BND(+)	-9.95	-4.72	-10.50	-3.15	-7.46	1.73	-2.29	1.45	
€/$(+)	4.92	-0.15	4.30	-0.65	5.65	3.83	3.04	0.74	
£/$(+)	22.03	4.78	8.01	5.69	16.74	7.13	12.35	5.71	

DM statistics comparing VaR predictions at a 95% and 97.5% confidence level of models with and without cross-excitation. The models predict the maximum of negative returns associated with a 95% confidence level the next day in the stock market, bond and exchange rate series and the maximum of positive returns associated with 95% confidence level in the bond market and exchange rate series from 1 January, 2008, to 1 July, 2015. The critical value corresponding to a 5% significance level is equal to ± 1.96 (asymptotic $\mathcal{N}(0,1)$-distribution).

of the Value-at-Risk at the 97.5% confidence level for investors with a long position in the US bond market.

To compare the VaR predictions made by the different models we use the DM test based on the asymmetric tick loss function of Giacomini and Komunjer (2005)

$$\mathrm{L}(e_\alpha^k(t)) = (\alpha - \mathrm{I}(e_\alpha^k(t) < 0))e_\alpha^k(t) \tag{4.21}$$

where $e_\alpha^k(t) = r(t) - (-VaR_\alpha^k(t))$ is the difference between the actual return and the predicted VaR at confidence level α and $\mathrm{I}(e_\alpha^k(t) < 0)$ is an indicator function which takes the value 1 when the negative of the return at time t is larger than the predicted VaR at confidence level α and the value 0 otherwise. As the loss function (4.21) is asymmetric, realized returns below the negative of $VaR_\alpha^k(t)$ lead to larger losses.

The DM statistics comparing VaR predictions of models with and without cross-excitation at the confidence levels 95% and 97.5% are given in Table 4.12. From the DM tests we conclude that, at the 95% confidence level, the models with cross-excitation, deliver significantly more accurate VaR predictions for investors with a long position in the S&P 500 index, the DJI index and for investors in the exchange rates. Investors with a long position in the S&P 500 index, NASDAQ and the €/$-rate and investors with a short position in the exchange

Table 4.13: DM statistics testing the influence on/of sizes VaR predictions

		$\eta\,(\alpha=0)$		$\alpha\,(\eta=0)$		$\eta\,(\alpha>0)$		$\alpha\,(\eta>0)$	
		$\gamma_2=0$	$\gamma_2>0$	$\gamma_2=0$	$\gamma_2>0$	$\gamma_2=0$	$\gamma_2>0$	$\gamma_2=0$	$\gamma_2>0$
95%	S&P	−35.72	−35.87	−3.54	−3.58	−22.91	−22.37	−3.58	−5.59
	NDQ	−49.02	−46.28	−1.05	−4.02	−49.26	−43.86	−4.02	−7.47
	DJI	−39.64	−37.22	3.25	−4.65	−30.19	−25.85	−4.65	−4.52
	BND(-)	−32.93	−35.92	14.11	−0.65	−36.98	−18.51	−0.65	27.84
	€/$(-)	−28.82	−25.90	−4.48	−12.06	−14.93	−12.87	−12.06	−6.54
	£/$(-)	−24.91	−23.28	−2.60	−4.11	−12.99	−11.67	−4.11	−8.20
	BND(+)	−35.86	−22.65	−8.42	−8.31	−32.10	−21.07	−8.31	−5.80
	€/$(+)	−35.41	−24.96	−5.62	−7.12	−20.32	−16.77	−7.12	−7.60
	£/$(+)	−33.11	−27.79	−4.58	−6.85	−8.19	−7.86	−6.85	−5.61
97.5%	S&P	−38.86	−39.10	−6.63	−10.36	−24.02	−24.92	−5.86	−6.35
	NDQ	−63.79	−53.83	0.78	−9.79	−65.14	−49.54	−2.67	−7.50
	DJI	−32.94	−33.57	−4.91	−9.77	−25.01	−21.25	−0.16	−4.83
	BND(-)	−4.42	−23.45	16.45	−5.27	−15.86	−3.84	14.44	22.63
	€/$(-)	−27.29	−23.20	0.47	−6.53	−15.87	−13.82	−5.15	−6.24
	£/$(-)	−28.83	−29.86	−3.72	−7.12	−13.62	−14.73	−8.32	−7.75
	BND(+)	−48.96	−39.87	−9.11	−7.23	−22.53	−17.46	−7.16	−7.54
	€/$(+)	−22.51	−20.74	−2.14	−4.76	−12.98	−14.73	−4.64	−7.88
	£/$(+)	−33.90	−28.27	−3.36	−9.37	−17.67	−17.84	−7.10	−9.23

DM statistics based on the asymmetric tick loss function (4.21) comparing VaR predictions at the 95-99% confidence level of models with and without influence of the the history of the events on the magnitude of subsequent events (η) and models with and without influence of the magnitude of events triggering subsequent events (α). The models predict the maximum of negative returns associated with a 95% confidence level the next day in the stock market, bond and exchange rate series and the maximum of positive returns associated with 95% confidence level in the bond market and exchange rate series from 1 January, 2008, to 1 July, 2015. The critical value corresponding to a 5% significance level is equal to ± 1.96 (asymptotic $\mathcal{N}(0,1)$-distribution).

rates should also employ the models with cross-excitation for their VaR predictions at the 97.5% confidence level. Moreover, the models with cross-excitation in which the event sizes are unpredictable and/or the sizes of events influence the rate at which events are triggered, produce significantly more accurate VaR predictions at the 97.5% confidence level than the models without cross-excitation for investors with a long position in the US bond market and the £/$-rate. Investors with a short position in the US bond market should employ models without cross-excitation to forecast their Value-at-Risk.

The DM statistics comparing probability predictions of models with and without influence of the event sizes on the conditional intensity and models with and without influence of the history of the event process on the event sizes, at the confidence levels 95% and 99% are given in Table 4.13. According to the DM tests, using models with influence of the history

of the event process on the event sizes and/or influence of the event sizes on the conditional intensity, leads to significantly less accurate VaR predictions in almost all cases.

Our key finding is that, all in all, the models that account for cross-triggering provide better VaR predictions. This is not surprising as these models are also ranked the highest in case the predictions of the probability of an extreme are evaluated. Therefore we conclude that traders, regulators of financial markets and risk management should incorporate spillovers in their models when they are employing Hawkes models to forecast extremes in financial markets.

4.6 Conclusion

We extend Hawkes models to account for cross-excitation between financial markets. We assess the contribution of including cross-excitation for forecasting extreme events and the Value-at-Risk. We apply the models to extreme negative returns in the S&P 500 index, the NASDAQ and the Dow Jones Industrial Average, and to extreme negative and positive returns in the US bond index, the euro/dollar rate and the pond/dollar rate to forecast the probability of extremes and the Value-at-Risk from 1 January, 2008, to 1 July, 2015. The forecast period includes the financial crisis of 2008. in which many crashes occurred. Our specific focus is on comparing models with and without spillover effects from the S&P 500 index, while for the S&P 500 index we investigate whether crashes in the index are provoked by crashes in the NASDAQ. On average, a Lagrange Multiplier test rejects the absence of cross-excitation except for the cross-excitation to the NASDAQ. Residual analysis confirms the adequacy of the models.

We evaluate the forecasts of the probability of an extreme and the Value-at-Risk forecasts at the 95% and 97.5% confidence level. The test of Diebold and Mariano (1995) shows that the Hawkes models with spillover effects and triggering effect that depends on the size of past event provide the most accurate forecasts of the probability of an extreme event for all series except for crashes in the S&P 500 index. The predictability of event sizes does not necessarily lead to more accurate probability predictions. The results for the VaR forecasts also indicate the added value of allowing for cross-excitation.

Our results suggest that extreme events in financial markets should not be modelled or predicted in isolation but in joint models. Although our attention is restricted to cross-

excitation from one process to another, our results provide an incentive for further research into multivariate Hawkes models.

Chapter 5

Estimation of non-affine models containing jumps using option prices

5.1 Introduction

In this Chapter I propose a framework for efficiently estimating stochastic volatility models with self-exciting jumps on stock and option prices that is very general and can easily be tailored to fit the application at hand.[1] The framework admits to analyze models with non-affine dynamics and compare both nested as non-nested models. Moreover, as my methods fit into a learning setting, models can be estimated and assessed sequentially such that models can be updated quickly when new information arrives. Within my framework, I use MCMC techniques and particle filtering methods to derive the distribution of the model parameters and the latent volatility and jump process. Utilizing the information in option prices, which can be done at a frequency different from the frequency of asset prices, I am able to derive the fractions of the equity and variance risk premium attributable to diffusive or jump risk. I pin down the sizes of the risk premia with the aid of high-frequency asset price data. The diffusive risk premium and, in particular, the jump risk premium, provide insight in the state and development of the financial market with important guidance for risk management.

The model of my main focus contains a self-exciting jump component on top of a stochastic volatility component, to accommodate the repetition of stock market crashes in a short time period, such as the ongoing sizable drops during the financial crisis of 2008. Nowadays

[1] The Chapter contains a research proposal for which the next step is to properly carry out a Monte Carlo analysis and data application.

stochastic volatility models are adapted to account for large changes in asset prices by the introduction of a jump component in the asset price and/or the volatility.[2] However many of these models cannot account for the tendency of large negative asset price movements to cluster over time. As jumps are rare events, the clustering of crashes is very unrealistic in models that prescribe jumps to arrive independently. Also model implied surfaces of option prices differ from observed surfaces if jumps have no impact on relative option prices (Eraker, 2004). Similar to Aït-Sahalia et al. (2015) and Boswijk et al. (2015), I propose a model in which the probability of a crash in the near future rises sharply after a large negative price drop as I let the jump intensity follow a self-exciting Hawkes process (Hawkes, 1971). There are quite a few studies showing that time-variation in the jump intensity, such that the jump intensity is higher in case of crises and lower during tranquil periods, improves the fit of asset and option prices.[3] Moreover, it seems that the fit of models that allow the jump intensity to suddenly increase, is superior to the fit of models in which the jump intensity can only increase gradually (Fulop et al., 2014; Aït-Sahalia et al., 2015; Boswijk et al., 2015). In contrast to studies in which the intensity is a function of the volatility,[4] I let the jump intensity change independently of the volatility over time. Therefore increments in the intensity do not have to be preceded by a increase in the volatility. Moreover, a high volatility does not necessarily imply more jumps. This is supported by the finding that innovations in volatility and the jumps intensity are not strongly correlated (Santa-Clara and Yan, 2010).

Using option prices to estimate my models, I am able to derive the different compensations investors require for taking on diffusive and jump risk. This is especially of interest as the vastly increased trading of derivatives indicates the trade-off investors face controlling their exposure to diffusive and jump risk. Santa-Clara and Yan (2010), Bollerslev and Todorov (2011), Bollerslev et al. (2015), Andersen et al. (2015b) and Boswijk et al. (2015) show that the compensation for the risk of disasters, not attributable to volatility, explains to a large extent the equity and variance risk premia, of which the last one can be seen an indication of the fear of investors. All acknowledge that the compensation for disaster risk,

[2]See among others Merton (1976), Heston (1993), Bakshi et al. (1997), Bates (2000), Chib et al. (2002), Andersen et al. (2002), Pan (2002), Chernov et al. (2003), Eraker et al. (2003), Eraker (2004), Bates (2006), Broadie et al. (2007), Li et al. (2008), Santa-Clara and Yan (2010), Kaeck and Alexander (2012), Durham (2013), Fulop et al. (2014), Aït-Sahalia et al. (2015), Boswijk et al. (2015) and Ignatieva et al. (2015)

[3]See for example Pan (2002), Eraker (2004), Bates (2006), Santa-Clara and Yan (2010), Fulop et al. (2014), Aït-Sahalia et al. (2015) and Boswijk et al. (2015)

[4]Bates (2000), Andersen et al. (2002), Pan (2002), Eraker (2004) and Bates (2006) examine jump-diffusion models in which the jump intensity is a function of the volatility of the diffusion.

and thereby the fear of investors, changes quite strongly over time. This justifies sharp market valuations. Moreover, while Santa-Clara and Yan (2010) and Bollerslev et al. (2015) find that the time-variation in the risk premia helps predict future market returns, Andersen et al. (2015b) find that options contain information about future market risk premia, which can not be identified from stock prices alone. Therefore, disentangling of volatility and jump components in risk premia using option prices provides one with important information regarding the state and development of the financial market with far-reaching implications for asset allocation, hedging, and risk management. As Andersen et al. (2015b), Bollerslev and Todorov (2011) and Bollerslev et al. (2015), I exploit the information contained in high-frequency asset prices to pin down risk premia more accurately.

I estimate my jump-diffusion model using Bayesian techniques. Bayesian analysis provides interpretable answers, a way to deal with missing data, tractable computations, and inferences that do not rely on asymptotic approximation and are exact conditional on the data. Despite the advantages of MCMC techniques, studies of models in which the jump intensity is a state variable and evolves independently from the volatility, have solely been done within the classical framework with the exception of Fulop et al. (2014). Moreover, the popular Hawkes model framework has only been estimated in continuous time with techniques that match moments in the data with moments implied by the model (Errais et al., 2010; Aït-Sahalia et al., 2014, 2015; Boswijk et al., 2015; Fan et al., 2015) or maximum likelihood (Bowsher, 2007). Before discussing my estimation procedure, I point out the three main advantages of my procedure, that is generality, flexibility and insightfullness. When discussing my procedure, it becomes clear that my methods are efficient and easy to implement.

First my framework is very general, that is I do not have to make on a lot of assumptions on both the models as well as the data. Regarding the model assumptions, my estimation framework only requires one to be able to evaluate the likeliness of the latent states given the option prices, which can be done using approximations or the integrated variance. As I do not rely on the existence of an analytical characteristic function for the price, I am able to investigate both affine as non-affine models. Even though affine model specifications are far more popular as they provide closed-form derivative prices which facilitates model calibration using option prices,[5] non-affine specifications, seems to fit and predict asset prices

[5]See for exampleMerton (1976), Heston (1993), Bakshi et al. (1997), Bates (2000), Pan (2002), Bates (2006), Broadie et al. (2007), Errais et al. (2010), Santa-Clara and Yan (2010), Aït-Sahalia et al. (2014), Fulop et al. (2014) and Boswijk et al. (2015)

considerably better.[6] Moreover, both Kaeck and Alexander (2012) and Durham (2013) show that non-affine models perform superior in fitting option prices compared to affine models, whether or not jumps are included in the models. Durham (2013) employs Monte Carlo simulation to invert the implied volatility from panels of option prices, after which he performs maximum likelihood estimation to estimate the parameters of the model under the risk neutral probability measure. A drawback of their method is that the parameters of the model under the physical probability measure are not estimated in the same estimation procedure. In contrast to Durham (2013), I estimate all parameters in one framework such that I do not have to rely on parameters derived in earlier studies. Kaeck and Alexander (2012) use the term structure of the VIX index to infer the transition from the physical to the risk neutral probability measure. I opt to estimate models from the asset prices and option prices only such that no additional measurement errors are introduced. Therefore my method is applicable to other series in addition to the index, whose risk neutral dynamics may differ from the dynamics implied by the VIX. Moreover, different from Kaeck and Alexander (2012) and Durham (2013), I do consider time-variation in the jump intensity.

Regarding the data assumptions, I accept that options are observed with error and that the information of options is utilized at a lower frequency than the information of asset prices to estimate models. Moreover, the framework allows me to exploit the intra-day information in asset prices such that I can estimate latent volatility and jump intensity states with greater precision. This improves the accuracy of the risk premia estimates. Boswijk et al. (2015) estimate a model similar to the model in this Chapter, assuming options can be observed without error. However, bid-ask spreads of options are relatively large, and therefore non-trivial. As one can solely conclude that true option values (if one believes in the existence of such), lie somewhere within the spreads, these values do not coincide with proxies of the option values such as mid-quotes (Andersen et al., 2015a). Furthermore as true option values cannot be inferred, option implied states contain errors as well. Although observed option prices contain large observation errors and my models can be estimated without, options are still very informative and bring estimation errors down even when taken in at a lower frequency than asset prices (Johannes et al., 2009).

[6] Chernov et al. (2003), Jones (2003), Christoffersen et al. (2010), Kaeck and Alexander (2012), Durham (2013) and Ignatieva et al. (2015) find non-affine models should be preferred above affine models as they are more flexible and better capable of modeling the tails of the heavy-tailed asset return distribution, while remaining equally parsimonious.

Secondly, my estimation technique is very flexible and can easily be tailored to other model specifications to fit the application at hand. For example, the jump intensity can be driven by a Brownian motion, the volatility process can contain jumps or the long-run level of the volatility can be made time-varying. Thus, it is allowed to let the dynamics under the real world measure to differ entirely from the dynamics under the risk neutral measure. Moreover, the technique can be extended to the multivariate case as it does not require direct optimization of a multidimensional integral which is a problem in several classic estimation frameworks that consider option prices. This makes the estimation technique very attractive for further investigation as jump intensities are believed to be mutually exciting (Aït-Sahalia et al., 2015; Fan et al., 2015).

Third, during my estimation procedure, I obtain statistics that give insight in the data and model fit, as a by-product. As I combine MCMC with particle filtering techniques, I am able to approximate the posterior of both the latent states and estimate parameters of the model in one framework. Furthermore, through the application of data augmentation, the distributions of the jumps and the jump sizes arise naturally from the estimation procedure. This allows me to identify the source of asset price shocks (Aït-Sahalia, 2004) and compare the models' fit of the diffusive and jump component separately. Moreover, as particle filters produce approximations of the marginal likelihood as a by product, I am able to compare models with Bayes Factors and Information Criteria.

Hereafter I explain the Bayesian estimation procedure, which is based on a Gibbs sampling algorithm. The estimation procedure consists of two steps: First a volatility, intensity and jump path is drawn given the model parameters (1). Hereafter the parameters of the models are drawn given this path (2). More specifically, I combine auxiliary particle filtering with particle Gibbs with anchestor sampling in a new estimation procedure, auxiliary particle Gibbs with anchestor sampling (APGAS). The APGAS procedure enables one to derive the latent variables, decompose unexpected shocks into diffusive and jump components, compare models with Bayes Factors and Information Criteria, and make predictions of both the latent variables as the stock and option prices. The basic auxiliary particle filter of Pitt and Shephard (1999) prescribes to incorporate the knowledge of the observations into the proposal distribution of the latent states such that particles are not moved into regions which are very unlikely given the observations. To prevent sample impoverishment even more, this APF is extended by transmitting information contained in the observations about jumps and

their sizes into the particle trajectories (Johannes et al., 2009). The steps of the APF are performed within the particle Gibbs with anchestor sampling (PGAS) method of Lindsten et al. (2014), which approximates an ideal Gibbs sampler. The PGAS method is used to estimate the invariant distribution of the latent variables and the model parameters conditional on each other. In the PGAS method as well as in the Particle Gibbs method of Andrieu et al. (2010), a reference trajectory, that is the previously sampled trajectory of the latent states, serves as a guidance through the state space. However, different from the PG method, the ancestor of the reference path is resampled. Although the particle marginal Metropolis-Hastings (PMMH) sampler of Andrieu et al. (2010) is theoretically preferred over the PGAS method, approximating an ideal marginal sampler for the parameters, the PMMH sampler is often too time-consuming to use in practice as a high number of particles and Monte Carlo iterations are required to give a somewhat accurate picture of joint distribution of the latent variables and the parameters of the model. On the contrary, in the PGAS method fast mixing is obtained with only few particles by resampling of the ancestor of a reference trajectory. Besides, with the PGAS method one samples the latent variables and parameters in separate stages of the iterative Gibbs sampling procedure, which improves mixing even further.

On top of auxiliary particle filtering and particle Gibbs with anchestor sampling, I propose to use the learning methods of Chopin et al. (2013) and Fulop and Li (2013). Using a sequential approach, new information can be processed quickly and used to update models as parameters are estimated recursively. Classical batch estimation on the other hand requires to redo the entire estimation procedure which can be time-consuming. Moreover, the methods of Chopin et al. (2013) and Fulop and Li (2013) are efficient as resample steps take place more frequently at the beginning when the data-sample is still short. I tailor in the methods Chopin et al. (2013) and Fulop and Li (2013) to the proposed auxiliary particle filter fit for jump-diffusion models, and apply particle Gibbs with anchestor sampler described above. From the particle filter sequential marginal likelihood estimates are obtained as a by product. These can be used to compute sequential Bayes factors. The sequential Bayes factor enables me to assess the performance of different models over time, something that is impossible in batch estimation.

The Chapter is organized as follows. In Section 2 the model specifications are discussed, the Hawkes process, which plays a key element in the models my main focus, is introduced, and the asset return dynamics under the physical and risk neutral measure are provided for

several affine and non-affine models. Section 3 describes my estimation method, the APF and the PGAS method are analyzed and an detailed Gibbs algorithm that includes APF and PGAS steps is proposed. A sequential estimation method based on PGAS, is presented in Section 4. Finally, Section 5 proposes a data application for my estimation framework. To avoid inconsistencies I recommend to estimate the models over a large time period (Pan, 2002; Broadie et al., 2007) instead of over a large cross-sectional option panel. A Monte Carlo analysis of the proposed framework has yet to be properly implemented. Also, applying the framework to the suggested data in Section 5 and deriving diffusive and jump risk premia, remains for further research.

5.2 Model

5.2.1 The Hawkes process

In my self-exciting jump models I use the Hawkes process (Hawkes, 1971) to describe the jump dynamics in the model. The Hawkes process is a branching process, in which each jump can trigger subsequent jumps, which in turn can trigger subsequent jumps of their own. Jumps N_t arrive according to an inhomogeneous Poisson process with a stochastic intensity λ_t. The Hawkes process is called self-exciting as the intensity gets amplified when a jump arrives, increasing the probability of future jumps, after which the intensity decays as a function of the time passed since the jump,

$$\lambda(t|\boldsymbol{\theta};\mathcal{H}_t) = \theta_\lambda + \int_{-\infty}^{t} g(t-s)\mathrm{d}N_s \tag{5.1}$$

where $\theta_\lambda > 0$ and $g(t) > 0$ whenever $N_t > 0$ and 0 elsewhere. The intensity λ_t at time t can be determined by the history of jumps prior to time t contained in $\mathcal{H}_t$, which makes the Hawkes process a path-dependent point process. The instantaneous mean jump rate conditional on $\mathcal{H}_t$ is thus given by

$$\Pr\left[N_{t+\delta} - N_t = 0|\mathcal{H}_t\right] = 1 - \lambda_t\delta + o(\delta) \tag{5.2}$$

$$\Pr\left[N_{t+\delta} - N_t = 1|\mathcal{H}_t\right] = \lambda_t\delta + o(\delta) \tag{5.3}$$

$$\Pr\left[N_{t+\delta} - N_t > 0|\mathcal{H}_t\right] = o(\delta) \tag{5.4}$$

As suggested by Hawkes (1971), I specify the triggering function $g(.)$ such that the pair (N, λ) is Markovian, and the instantaneous intensity λ_t with which jumps N_t at time t arrive, can be determined by the intensity λ_{t-1} and the number of jumps N_{t-1} at time $t-1$,

$$g(t) = \gamma_\lambda e^{-\kappa_\lambda t} \tag{5.5}$$

where $\kappa_\lambda, \gamma_\lambda > 0$. Interpreting (5.5) in combination with (5.1), the conditional intensity consists of a long-term background component given by the constant term and a short-term clustering component describing the temporal distribution of triggered jumps. The parameter θ_λ controls the steady state level to which the conditional intensity mean reverts. The impact of a jump on the conditional intensity is measured by $\gamma_\lambda e^{-\kappa_\lambda t}$, where γ_λ defines the maximum triggering intensity and the expected number of jumps directly triggered by a jump and κ_λ defines the exponential rate with which the impact of a jump decays depending on the time passed since a jump. Together, the parameters γ_λ and κ_λ determine the branching ratio of the Hawkes process n, that is the expected number of jumps triggered by a jump,

$$n = \int_0^\infty \gamma_\lambda e^{-\kappa_\lambda s} \mathrm{d}s = \frac{\gamma_\lambda}{\kappa_\lambda} \tag{5.6}$$

When $n \geq 1$, the conditional intensity, and thus the number of events arriving within a finite time interval, could grow to infinity over time. Restricting $\gamma_\lambda < \kappa_\lambda$, the branching ratio n is smaller than 1, which ensures stationarity of the Hawkes process. The compensated process $N_t - \int_{-\infty}^t \lambda_s \mathrm{d}s$ is a local martingale. The intensity process (5.1,5.5), can be verified to have the following differential form

$$\mathrm{d}\lambda_t = \kappa_\lambda(\theta_\lambda - \lambda_t)\mathrm{d}t + \gamma_\lambda \mathrm{d}N_t \tag{5.7}$$

The mean reversion to the steady state level is generated by the drift term $\kappa_\lambda(\theta_\lambda - \lambda_t)$, whereas $\gamma_\lambda \mathrm{d}N_t$ covers the self-excitation of the Hawkes process.

5.2.2 Asset return dynamics in continuous-time

Similar to the model of Bates (2000), all models consist of the following key elements for the dynamics of the log returns on assets $s_t = \log S_t$; the level of the stochastic volatility v_t,

the number of jumps N_t, the sizes of the jumps Z_t and the intensity with which these jumps arrive λ_t (which can be taken to be constant). The asset return dynamics of the models under the physical probability measure $\mathbb{P}$ are covered by the following differential equation

$$\mathrm{d}s_t = \mu_t^{\mathbb{P}}\mathrm{d}t + \sqrt{v_t}\mathrm{d}W_t^{s,\mathbb{P}} + Z_t^{\mathbb{P}}\mathrm{d}N_t \tag{5.8}$$

where $W_t^{s,\mathbb{P}}$ represents a standard Brownian motion under $\mathbb{P}$. We consider models in which the asset return dynamics are affine (A), and the variance follows the square-root process of Heston (1993),

$$\mathrm{d}v_t = \kappa_v\left(\theta_v - v_t\right)\mathrm{d}t + \sigma_v\sqrt{v_t}\mathrm{d}W_t^{v,\mathbb{P}} \tag{5.9}$$

and models in which the asset return dynamics are non-affine (NA) with the following dynamics for the log variance $h_t = \log v_t$,

$$\mathrm{d}h_t = \kappa_v\left(\theta_v - h_t\right)\mathrm{d}t + \sigma_v\mathrm{d}W_t^{v,\mathbb{P}} \tag{5.10}$$

The variance in (5.9) is bounded below by zero as we impose the Feller condition, $2\kappa_v\theta_v \geq \sigma_v^2 \geq 0$. In both models the stochastic volatility v_t is mean reverting towards a long-run mean $\bar{v}$ specified by θ_v (A) or $\exp\left(\theta_v + \frac{1}{2}\sigma_v^2\right)$ (NA). However in the affine specification the speed of the mean reversion, given by κ_v for the affine models, is slower than in the non-affine specification. Furthermore the immediate impact of the shocks on the volatility is larger in the non-affine models. Therefore the volatility moves more violently in these models than in the affine models. I express $W_t^{s,\mathbb{P}}$ as $W_t^{s,\mathbb{P}} = \rho W_t^{v,\mathbb{P}} + \sqrt{1-\rho^2}W_t^{s\backslash v,\mathbb{P}}$, such that the standard Brownian motions in (5.8) and (5.9) or (5.10) are correlated with correlation coefficient ρ, $-1 \geq \rho \geq 1$, capturing the leverage effect reported by Black (1976).

All models considered have a jump component $Z_t^{\mathbb{P}}\mathrm{d}N_t$ in the asset return dynamics (5.8) causing the stock price S_t to jump to $S_{t-}\exp\left(Z_t^{\mathbb{P}}\right)$ whenever a jump arrives indicated by N_t. I assume that the jump component is independent of the Brownian motions $W_t^{s,\mathbb{P}}, W_t^{v,\mathbb{P}}$. Furthermore, I assume the jump sizes, $Z_t^{\mathbb{P}}$, and the process indicating the arrival of jumps N_t, are mutually independent. The jump sizes $Z_t^{\mathbb{P}}$ follow a normally distribution with mean $\mu_z^{\mathbb{P}}$ and standard deviation σ_z independently of previous jump sizes. From this it follows that, on average, the relative size of stock price jumps under the physical probability measure $\mathbb{P}$ is equal to $\mu_j^{\mathbb{P}} = \mathbb{E}\left[\exp\left(Z_t^{\mathbb{P}}\right) - 1\right] = \exp\left(\mu_z^{\mathbb{P}} + \frac{1}{2}\sigma_z^2\right) - 1$.

The timing of the jumps is determined by the Poisson process N_t which has intensity λ_t. In the models denoted by (C), λ_t is constant, implying jumps should be more or less evenly distributed over the period examined. Inspired by Andersen et al. (2002), Pan (2002), Eraker (2004) and Bates (2006), I also look at models that have a state-dependent intensity, that is $\lambda_t = \alpha + \beta v_t$. These models, denoted by (S), are able to explain alternating periods in which more or less jumps occur as the intensity of jumps is able to gradually adjust in line with the volatility whenever a turmoil period with fierce stock price movements, or a tranquil period without such movements, dawns.

In the models of my main focus, the Hawkes models (H), the jump intensity λ_t follows the self-exciting Hawkes process specified in (5.7) such that the occurrence of jumps induces a higher probability of jumps in the near future. Similar to the volatility, the Hawkes jump intensity exhibits mean revering behavior to its long-run mean, $\bar{\lambda} = \frac{\theta_l \kappa_l}{\kappa_l - \gamma_l}$, with mean reversion speed κ_l. Benefits of adopting a self-exciting jump intensity compared of a state-dependent intensity, are that information within jumps and feedback effects (due to algorithmic trading, illiquidity, trade and budget deficits etc.) can be incorporated in the model.

I do not consider models with jumps in the volatility process. Models that contain jumps solely in the volatility process cannot capture the jumps observed in the price dynamics of financial assets, whereas models that contain jumps in both price as volatility dynamics are less parsimonious and not identifiable with the jump dynamics of the Hawkes process (5.7). Moreover, in the non-affine models, the log variance dynamics already allow the volatility to rapidly increase and decrease in response to shocks.

In my models contain jump size risk in addition to diffusive, volatility and jump time risk. In this setting the market is incomplete with respect to underlying stock, a money market account and a finite number of option contracts. The pricing kernel is therefore not unique. Analogous to Boswijk et al. (2015), I assume in the absence of arbitrage, the following price and volatility dynamics under an equivalent risk neutral measure $\mathbb{Q}$

$$y: \quad \mathrm{d}s_t = \mu_t^{\mathbb{Q}}\mathrm{d}t + \sqrt{v_t}\mathrm{d}W_t^{s,\mathbb{Q}} + Z_t^{\mathbb{Q}}\mathrm{d}N_t \tag{5.11}$$

$$v: \quad \mathrm{d}v_t = \kappa_v\left(\theta_v - v_t\right)\mathrm{d}t + \sigma_v\sqrt{v_t}\mathrm{d}W_t^{v,\mathbb{Q}} \quad \text{(A)} \tag{5.12}$$

$$h: \quad h_t = \log v_t \quad \text{(NA)} \tag{5.13}$$

$$h: \quad \mathrm{d}h_t = \kappa_v\left(\theta_v - h_t\right)\mathrm{d}t + \sigma_v\mathrm{d}W_t^{v,\mathbb{Q}} \quad \text{(NA)} \tag{5.14}$$

$$\lambda: \quad \lambda_t = \lambda \quad \text{(C)} \tag{5.15}$$

$$\lambda: \quad \lambda_t = \alpha + \beta v_t \quad \text{(S)} \tag{5.16}$$

$$\lambda: \quad \mathrm{d}\lambda_t = \kappa_\lambda(\theta_\lambda - \lambda_t)\mathrm{d}t + \gamma_\lambda\mathrm{d}N_t \quad \text{(H)} \tag{5.17}$$

where $W_t^{s,\mathbb{Q}}$ can be expressed as $W_t^{s,\mathbb{Q}} = \rho W_t^{v,\mathbb{Q}} + \sqrt{1-\rho^2}W_t^{s\backslash v,\mathbb{Q}}$ which induces correlation between the standard Brownian motions in the price and volatility dynamics under the risk neutral measure $\mathbb{Q}$. I assume that the timing of the jumps $\mathrm{d}N_t$ is the same under the $\mathbb{P}$ as under $\mathbb{Q}$ such that the jump intensity is unaffected by the measure change. The jump sizes $Z_t^{\mathbb{Q}}$ follow a normally distribution with mean $\mu_z^{\mathbb{Q}}$ and standard deviation σ_z independently of previous jump sizes, the jump process N_t and the Brownian motions $W_t^{s,\mathbb{Q}}, W_t^{v,\mathbb{Q}}$. Conditionally up to arrival of a jump, the mean relative size of a jump under the risk neutral probability measure is given by $\mu_j^{\mathbb{Q}} = \mathbb{E}\left[\exp\left(Z_t^{\mathbb{Q}}\right) - 1\right] = \exp\left(\mu_z^{\mathbb{Q}} + \frac{1}{2}\sigma_z^2\right) - 1$. The risk neutral dynamics differ slightly from the dynamics defined by other studies as I follow Boswijk et al. (2015), who opt set to the market price of volatility risk to zero since there is no clear empirical evidence for a diffusive volatility risk premium.

5.2.3 Options

To fully employ the information in option prices to estimate our models, I adopt the following specification for the drift term of the price process (5.8) under $\mathbb{P}$, and $\mathbb{Q}$,

$$\mu_t^{\mathbb{P}} = (\eta - \frac{1}{2})v_t + \left(\mu_j^{\mathbb{P}} - \mu_j^{\mathbb{Q}}\right)\lambda_t - \mu_j^{\mathbb{P}}\lambda_t \tag{5.18}$$

$$\mu_t^{\mathbb{Q}} = -\frac{1}{2}v_t - \mu_j^{\mathbb{Q}}\lambda_t \tag{5.19}$$

Defining (5.18) enables me to separately derive the market price of diffusive and jump risk. When I estimate the models solely using asset prices, I set $\mu = \mu_t^{\mathbb{P}}$. Adopting a rich speci-

fication for the drift of the price process like (5.18) would leave the models underidentified. Therefore, without options, I am not able to determine the different compensations investors' require for the different risk types. Using option prices, the market price of diffusive risk per unit time is equal to ηv_t. The time-variation in the diffusive risk premium arises from the stochastic time-variation in the volatility, which induce changes in investors' risk and the price investors require for bearing this risk by investing in the asset. The average level of the risk premium is determined by η. As I restrict the jump process N_t and its' intensity λ_t to be the same under the physical and risk neutral probability measures, the jump risk premium is not related to the timing of the jumps. The premium emerges entirely from perceived differences in jump size distribution under $\mathbb{P}$ and $\mathbb{Q}$. Moreover, as σ_j is the same under both measures, the deviation of the mean of the jump size distribution under $\mathbb{Q}$ from its mean under $\mathbb{P}$ solely induces the premium investors require for bearing jump risk. The jump risk premium is equal to $\left(\mu_j^{\mathbb{P}} - \mu_j^{\mathbb{Q}}\right) \lambda_t$, which makes the premium time-varying whenever λ_t is time-varying. This makes the probability of a price jump and thus investors' risk, change over time.

For the affine models (A), in which the variance dynamics are given by (5.9), I am able to derive the conditional characteristic function of the models' stochastic states using the methods of Duffie et al. (2000). Having specified this conditional characteristic function, I apply the Fourier inversion techniques based on Fourier-cosines series expansions (Fang and Oosterlee, 2008), to price options.

5.3 Estimation

To estimate the models defined in the previous section, I employ Bayesian techniques. Although Gibbs sampling schemes have been used before to estimate jump-diffusion models,[7] the estimation of a jump-diffusion model in which the the jump intensity follows a Hawkes process has only been done within the classical framework. My Bayesian estimation method offers several advantages. For one thing the distribution of the volatility, the self-exciting jump intensity, jumps times and sizes, and the model parameters can be estimated in one framework without some commonly used, but improbable, assumptions. For example, the

[7] See for example Chib et al. (2002), Eraker et al. (2003), Eraker (2004), Li et al. (2008), Kaeck and Alexander (2012) and Ignatieva et al. (2015)

estimation procedure is not based on the compliance of our models with the terms of Black-Scholes (Black and Scholes, 1973) nor on the ability to observe options without error. Moreover, as my estimation framework does not rely on the existence of an analytic characteristic function for the price of an asset, non-affine models can be estimated using my methods. Lastly, I do not require asset and option prices to be available at the same frequency to utilize all available information.

In my Gibbs sampling procedure I first draw a volatility, intensity and jump path conditional on the model parameters, after which I draw the model parameters given the sampled paths. For the first step I combine auxiliary particle filtering methods with the Particle Gibbs with anchestor sampling in a new algorithm, the auxiliary particle Gibbs with anchestor sampling algorithm (APGAS). The APGAS estimation procedure enables one to derive the latent variables of the models, disentangle stochastic volatility and jumps, compute Bayes Factors and Information Criteria, and make predictions of both the latent variables as the stock and option prices. Particle filtering techniques have been widely applied to infer latent states from observed data in state space models. Compared to the well-known sequential importance resampling (SIR) technique, the auxiliary particle filter (APF) is more capable of handling the degeneracy problem experienced with particle methods. The APF filter incorporates the information about the latent states contained in observations into the particle trajectories. This prevents particles from moving into regions that are very unlikely given the observations. Inspired by Johannes et al. (2009), I extend the basic APF of Pitt and Shephard (1999) to filter out both the latent states as the jump times and sizes from asset prices, and if available for estimation, option prices. To derive the invariant distribution of both the latent states and the model parameters conditional on each other, I make some adjustments to the filtering scheme. Using particle Gibbs with anchestor sampling (PGAS) of Lindsten et al. (2014), I am able approximate the conditional distribution of the latent states without the use of a large number of particles. This makes the method attractive in terms of computation time compared to other filtering estimation procedures such as the particle Gibbs (PG) sampler and the particle marginal Metropolis-Hastings (PMMH) sampler of Andrieu et al. (2010). Conditional on the latent states, the model parameters are drawn from conjugate posterior distributions derived in earlier studies described in Li et al. (2008), or using the Metropolis-Hastings Algorithm.

In this section I discuss my estimation methods in more detail. First I make the models suitable for my inference procedure by deriving their discrete version. Hereafter I discuss the extended version of APF and the steps of the PGAS algorithm separately. The full sampling algorithm is provided in the last subsection.

5.3.1 Asset return dynamics in discrete time

To solve the equations (5.7), (5.8), (5.9) and (5.10) at the frequency associated with the observed data, I employ the Euler scheme to derive a discrete time version of my continuous-time models. That is

$$y: \quad r_t = \mu_t^{\mathbb{P}} + \sqrt{v_t}\varepsilon_t^s + Z_t^{\mathbb{P}} N_t \tag{5.20}$$

$$v: \quad v_{t+1} = v_t + \kappa_v(\theta_v - v_t) + \sigma_v\sqrt{v_t}\varepsilon_{t+1}^v \quad \text{(A)} \tag{5.21}$$

$$h: \quad h_{t+1} = \log v_{t+1} \quad \text{(NA)} \tag{5.22}$$

$$h: \quad h_{t+1} = h_t + \kappa_v(\theta_v - h_t) + \sigma_v\varepsilon_{t+1}^v \quad \text{(NA)} \tag{5.23}$$

$$\lambda: \quad \lambda_{t+1} = \lambda \quad \text{(C)} \tag{5.24}$$

$$\lambda: \quad \lambda_{t+1} = \alpha + \beta v_t \quad \text{(S)} \tag{5.25}$$

$$\lambda: \quad \lambda_{t+1} = \lambda_t + \kappa_l(\theta_l - \lambda_t) + \gamma_l N_t \quad \text{(H)} \tag{5.26}$$

where r_t denotes the continuously compounded log return, that is $r_t = s_{t+1} - s_t$, $N_{t,j}$ is a Poisson distributed random variable with intensity $\lambda_{t,j}$, and ε_t^s and ε_t^v, are standard normally distributed random variables. Like Johannes et al. (2009), it is possible to use a finer discretization scheme which enables one to filter volatility, jumps and jump sizes more accurately from the prices of assets and/or options. However, deriving the parameters of this model specification will be a very difficult task since there is no analytic form for the joint distribution of the log returns and the volatilities (Hull and White, 1987). Moreover, in case prices are available at a daily frequency, refinement of the discretization scheme seems not too important, as the discretization bias at this frequency is negligible (Eraker et al., 2003).

5.3.2 Auxiliary particle filter

The goal of a particle filter is to approximate the nonlinear and non-analytical density function $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ for each time period t, where $\mathbf{y}_{1:t} = \{\mathbf{y}_1, \ldots, \mathbf{y}_t\}$ and $\mathbf{x}_t$ represent respectively the observations and the unobserved latent variables. The density $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ is approximated through a discrete distribution of N points or particles, denoted by $\hat{p}^N(\mathbf{x}_t|\mathbf{y}_{1:t})$. Each particle $\mathbf{x}_t^i$, $i \in \{1, \ldots, N\}$, has a certain probability mass π_t^i associated with the likeliness of values of the latent variables $\mathbf{x}_t^i$ given the observed data $\mathbf{y}_{1:t}$. Under mild regularity conditions, $\hat{p}^N(\mathbf{x}_t|\mathbf{y}_{1:t})$ converges to $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ as N increases (Del Moral, 2004). Particle filters are designed for hidden Markov models in which the observations are related to the hidden Markov states with some functional form. This funtional form specifies the conditional likelihood $p(\mathbf{y}_t|\mathbf{x}_t)$ and the transition density of the latent Markov states $p(\mathbf{x}_{t+1}|\mathbf{x}_t)$. Within this model class, particle filters have as a great advantage that they are highly adaptive and can easily be tailored to derive $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ in the application at hand.

The original auxiliary particle filter (APF) of Pitt and Shephard (1999) is build on the idea that observations contain information on the latent states, which can be processed in the proposal distribution for the latent states. In the well-known sequential importance resampling (SIR) technique, particles $\mathbf{x}_{t+1}^i$ are sampled from their transition density $p(\mathbf{x}_{t+1}|\mathbf{x}_t^i)$ after which the particles are resampled with replacement according to their likeliness $p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i)$. The APF modifies the SIR algorithm by first sampling an auxiliary variable, the particle index $z(i)$, for each particle $\mathbf{x}_t^i$, given the compatibility of the particle with the next observation

$$p(\mathbf{y}_{t+1}|\mathbf{x}_t^i) = \int p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}) p(\mathbf{x}_{t+1}|\mathbf{x}_t^i) \mathrm{d}\mathbf{x}_{t+1} \tag{5.27}$$

Through the resampling step information is transmitted from the observation $\mathbf{y}_{t+1}$ into the distribution for the particles $p(\mathbf{x}_{t+1}|\mathbf{x}_t^{z(i)})$ such that $\mathbf{x}_{t+1}^i \sim p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \mathbf{y}_{t+1})$. In case of outliers or rare events, the information in observations is especially valuable. As particle trajectories are preselected based on their predictive likelihood, only the particles trajectories with a high likelihood survive. This is of great importance as states are generally very persistent, and thereby circumvents sample impoverishment. Density (5.27) is generally not available in closed form and approximated by $\hat{p}(\mathbf{y}_{t+1}|\hat{\mathbf{x}}_{t+1}^i)$, where $\hat{\mathbf{x}}_{t+1}^i$ is a value with high probability given $p(\mathbf{x}_{t+1}|\mathbf{x}_t^i)$ such as its mean or mode. Hereafter the particle histories $\mathbf{x}_{1:t}^i$, $i = 1, \ldots, N$, are set to $\mathbf{x}_{1:t}^{z(i)}$. Whenever the predictive density (5.27) cannot be derived

exactly, the particles are reweighted in a last step according to

$$w_{t+1}^i \propto \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i)}{\hat{p}(\mathbf{y}_{t+1}|\hat{\mathbf{x}}_{t+1}^i)} \qquad i = 1, \ldots, N \tag{5.28}$$

In case of full adaption, that is the predictive density (5.27) can be derived in closed form, w_{t+1}^i is equal to 1.

Setting $w_{t|t+1}^i = \hat{p}(\mathbf{y}_{t+1}|\hat{\mathbf{x}}_{t+1}^i)$ an estimator of the likelihood of observation $\mathbf{y}_{t+1}$ conditional on $\mathbf{y}_{1:t}$ is given by

$$\hat{p}^N(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}) = \left\{ \frac{1}{N} \sum_{i=1}^N w_{t+1}^i \right\} \left\{ \sum_{i=1}^N w_{t|t+1}^i \right\} \tag{5.29}$$

Using (5.29) the following estimator of the likelihood can be constructed

$$\hat{p}^N(\mathbf{y}_{1:t}) = \hat{p}^N(\mathbf{y}_1) \prod_{t=1}^{T-1} \hat{p}^N(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}) \tag{5.30}$$

Del Moral (2004) and Pitt et al. (2012) show the likelihood estimator (5.30) is unbiased with respect to all random quantities used in the particle filter.

The extended APF of Johannes et al. (2009) transmits the knowledge observations have on the latent variables also into the distribution of the non-persistent latent jump variables. This affects the proposal distribution of the persistent latent Markov states, which can be inferred more accurately than with the basic APF. Given that the jump times are conditionally Poisson and jump sizes are normally distributed, the non-persistent latent variables pertaining the jump component in the models can be integrated from the predictive density (5.27). To integrate out the persistent latent states, these states are, as in the original APF, replaced by some function $\hat{\mathbf{x}}_{t+1}^i$ of $\mathbf{x}_t^i$ with high probability $p(\hat{\mathbf{x}}_{t+1}^i|\mathbf{x}_t^i)$.

Similar to Johannes et al. (2009), I let y_t consist of the continuously compounded log-returns on a stock price $r_t = S_{t+1} - S_t$, and, depending on the timing of the observation, r_t can also contain the price of an option C_t. The volatility v_t, the jump time N_t and jump size Z_t are latent variables in my models. Different from Johannes et al. (2009), my observations also include high frequency measures of the daily integrated diffusive volatility IV_t. Moreover, I substitute the persistent latent jump intensities on top of the volatilities. That is $\mathbf{x}_t$ also contains the jump intensity λ_t besides the volatility v_t, in models in which the jump

intensity is a state variable. I make another adjustment to the filtering scheme of Johannes et al. (2009) as I follow Creal (2012) and Pitt et al. (2012), and combine the second stage weights of the previous time period with the first stage weights to select the relevant particles more efficiently.

Furthermore I follow Yu (2005) by modeling the leverage effect through a contemporaneous dependence between return and volatility innovations. That is $\mathrm{corr}(\varepsilon_t^s, \varepsilon_{t+1}^v) = \rho$, instead of the inter-temporal dependence $\mathrm{corr}(\varepsilon_t^s, \varepsilon_t^v) = \rho$ modeled by Johannes et al. (2009). Yu (2005) find that the contemporaneous specification is superior to the inter-temporal specification from both theoretical and empirical perspectives. Most importantly, stochastic volatility models with inter-temporal dependence between the return and volatility innovations do not constitute a martingale difference series. The implication is that these models are not consistent with the efficient market hypothesis.

Similar to the basic APF, the extended APF prescribes to sample the auxiliary particle indices $z(i)$ according to the compatibility of the particles with the coming observation (5.27). Assuming the occurrence of jumps is rare such that the probability of more than $\bar{K}$ jumps is negligible, the predictive density (5.27) can be approximated as

$$\begin{aligned}\hat{p}(\mathbf{y}_{t+1}|\hat{\mathbf{x}}_{t+1}^i) \propto & p(C_{t+1}|S_{t+1}, \hat{v}_{t+1}^i, \hat{\lambda}_{t+1}^i) p(IV_{t+1}|\hat{v}_{t+1}^i) \\ & \sum_{k=0}^{\bar{K}} \phi(r_{t+1}|\hat{\mu}_{t+1}^{i,\mathbb{P}} + k\mu_z, \hat{v}_{t+1}^i + k\sigma_z^2) \mathcal{P}(k|\lambda_t^i)\end{aligned} \tag{5.31}$$

where $\phi(\cdot|\mu, \sigma^2)$ and $\mathcal{P}(\cdot|\lambda)$ denote respectively, the normal probability density function with mean μ and variance σ^2, and the Poisson density function with rate λ. As Creal (2012) and Pitt et al. (2012), I combine predictive density, $w_{t|t+1}^i = \hat{p}(\mathbf{y}_{t+1}|\hat{\mathbf{x}}_{t+1}^i, J_{t+1}, Z_{t+1})$, with the likelihood of the particle trajectories $\mathbf{x}_{1:t}^i$ till time t given all past information $\mathbf{y}_{1:t}$, w_t^i, to select the particles that are most congruent with both the coming observation and the past observations. That is I resample $z(i)$ according to

$$w_{t:t+1}^i \propto \pi_t \hat{p}(\mathbf{y}_{t+1}|\hat{\mathbf{x}}_{t+1}^i) \tag{5.32}$$

Hereafter the particle trajectories $\{\mathbf{x}_{1:t}^i, \mathbf{J}_{1:t}^i, \mathbf{Z}_{1:t}^i\}$, $i = 1, \ldots, N$, are set to their resampled counterparts $\left\{\mathbf{x}_{1:t}^{z(i)}, \mathbf{J}_{1:t}^{z(i)}, \mathbf{Z}_{1:t}^{z(i)}\right\}$. In the extended APF the information in $\mathbf{y}_{t+1}$ is exploited to sample the number of jumps J_{t+1}^i and the jump sizes Z_{t+1}^i for each particle i as follows. The

number of jumps J^i_{t+1}, which is allowed to range from 0 to $\bar{K}$, is simulated using weights

$$\hat{p}(\mathbf{y}_{t+1}|\hat{\mathbf{x}}^i_t, k) \propto \phi(r_{t+1}|\hat{\mu}^{i,\mathbb{P}}_{t+1} + k\mu_z, \hat{v}^i_{t+1} + k\sigma^2_z)\mathcal{P}(k|\lambda^i_t) \tag{5.33}$$

whereas the jump sizes Z^i_{t+1} are be sampled from a normal distribution with mean and variance

$$\mu^i_{z|r} = J^i_{t+1}\mu_z + \frac{J^i_{t+1}\sigma^2_z}{\hat{v}^i_{t+1} + J^i_{t+1}\sigma^2_z}\left(r_{t+1} - \hat{\mu}^{i,\mathbb{P}}_{t+1} - J^i_{t+1}\mu_z\right) \tag{5.34}$$

$$\sigma^{2,i}_{z|r} = J^i_{t+1}\sigma^2_z\left(1 - \frac{J^i_{t+1}\sigma^2_z}{\hat{v}^i_{t+1} + J^i_{t+1}\sigma^2_z}\right) \tag{5.35}$$

Thereby we assume the distribution of the returns and the jump sizes is bivariate normal conditional on the number of jumps and the proxies of the persistent states. I proceed drawing the persistent states from $p(\mathbf{x}_{t+1}|\mathbf{x}^i_t, J_{t+1}, Z_{t+1}, y_{t+1})$. As a last step I compute the normalized weight of the particle trajectories till time t, $\pi_{t+1} = \frac{w^i_{t+1}}{\sum_{i=1}^N w^i_{t+1}}$, where w^i_{t+1} is equal to

$$w^i_{t+1} \propto \frac{p(\mathbf{y}_{t+1}|\mathbf{x}^i_{t+1}, J^i_{t+1}, Z^i_{t+1})\mathcal{P}(J^i_{t+1}|\lambda^i_t)\phi(Z^i_{t+1}|J^i_{t+1}\mu_z, J^i_{t+1}\sigma^2_z)}{\hat{p}(\mathbf{y}_{t+1}|\hat{\mathbf{x}}^i_{t+1})\hat{p}(\mathbf{y}_{t+1}|\hat{\mathbf{x}}^i_{t+1}, J^i_{t+1})\phi(Z^i_{t+1}|\mu^i_{z|r}, \sigma^{2,i}_{z|r})} \qquad i = 1, \ldots, N \tag{5.36}$$

5.3.3 Particle Gibbs with anchestor sampling

Particle Gibbs with anchestor sampling (PGAS) (Lindsten et al., 2014) fits into the Particle Markov chain Monte Carlo techniques (Andrieu et al., 2010) combining particle filtering and Markov chain Monte Carlo. PMCMC samplers approximate the ideal Gibbs sampler using particle filtering to efficiently construct high-dimensional MCMC kernels, which can be used in inference strategies. The PGAS method is inspired by the Particle Gibbs sampling (PG) method of Andrieu et al. (2010). In the PG method one of the particle trajectories is fixed in advance, the reference trajectory. The reference trajectory serves as a guide for the particles such that the particles remain in relevant section of the state space. After each evaluation of the particle filter, a new particle path is drawn using the importance weights that are derived in the filter. By setting of one of the particle trajectories equal to the reference trajectory (given by the previously sampled particle path), the obtained Markov kernel does not affect the target distribution regardless of the number of particles employed in the filter.

A big disadvantage of the PG method is that the Markov kernel mixes very poorly when the particle filter suffers from degeneracy. As in high-dimensional models degeneracy is unavoidable, an extremely large number of particles is needed to utilize the PG method to estimate these models. A solution to overcome this drawback of the PG sampler, is to add a backward simulation step to the sampler. The backward simulation step improves mixing of the Markov kernel. However, this step cannot be defined for non-Markovian latent variable models. In contrast to PG with backward simulation, Particle Gibbs with anchestor sampling averts the need of an explicit backward simulation step by the insertion of an ancestor resampling step in the PG sampler. Therefore this method can be applied to non-Markovian latent variable models. Moreover, the ancestor sampling step enables movement around the reference trajectory, making the PGAS sampler less sticky than the PG sampler.

As mentioned before, the PGAS method does not leave the reference trajectory $\mathbf{x}'_{1:T}$ intact through the particle filtering procedure. At time t, the history of part of the reference trajectory $\mathbf{x}'_{t+1:T}$ is reassigned in the ancestor resampling step. To link the partial reference path to the particles, the probability that the path $\mathbf{x}'_{t+1:T}$ is preceded by particle $\mathbf{x}^i_{1:t}$ for $i \in \{1, \ldots, N\}$ is evaluated as

$$\gamma^i_{t+1|T} = \frac{p(\mathbf{y}_{1:T}, (\mathbf{x}^i_{1:t}, \mathbf{x}'_{t+1:T}))}{p(\mathbf{y}_{1:t}, \mathbf{x}^i_{1:t})} = \prod_{s=t+1}^{T} p(\mathbf{y}_s|\mathbf{x}^i_{1:t}, \mathbf{x}'_{t+1:s}) p(\mathbf{x}'_s|\mathbf{x}^i_{1:t}, \mathbf{x}'_{t+1:s}) \tag{5.37}$$

where $(\mathbf{x}^i_{1:t}, \mathbf{x}'_{t+1:T})$ denotes the particle path formed when concatenating the two partial paths $\mathbf{x}^i_{1:t}$ and $\mathbf{x}'_{t+1:T}$. The history of partial reference path $\mathbf{x}'_{t+1:T}$ is resampled drawing a particle index z' from a distribution with weights given by the multiplication of the probabilities (5.37) and the prior weights of the particle trajectories $\mathbf{x}^i_{1:t}$. The reference path is redefined thereafter as $\mathbf{x}'_{1:T} = (\mathbf{x}^{z'}_{1:t}, \mathbf{x}'_{t+1:T})$. For state space models the relevant part of the resampling weight that indicates the likeliness the partial reference path $\mathbf{x}'_{t+1:T}$, is preceded by particle $\mathbf{x}^i_{1:t}$, reduces to $p(\mathbf{x}'_{t+1}|\mathbf{x}^i_t)$.

Lindsten et al. (2014) prove that the target distribution remains invariant when the ancestor resampling step is used in the PG sampler to construct the MCMC kernel. Lindsten et al. (2014) also prove uniform ergodicity of the PGAS method, under the condition that the importance weights derived in the particle filter are bounded by some constant M, $M < \infty$, for all parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and time periods $t \in \{1, \ldots, T\}$. Characterizing the support of the

proposal and the target density $\bar{\gamma}_{\boldsymbol{\theta}}$,

$$\mathcal{S}_t^{\boldsymbol{\theta}} = \left\{ \mathbf{x}_{1:t} \in \mathbf{x}^t : \bar{\gamma}_{\boldsymbol{\theta}}(\mathbf{y}_{1:t}, \mathbf{x}_{1:t}) > 0 \right\} \tag{5.38}$$

$$\mathcal{Q}_t^{\boldsymbol{\theta}} = \left\{ \mathbf{x}_{1:t} \in \mathbf{x}^t : p_{\boldsymbol{\theta}}\left(\mathbf{x}_t | \mathbf{x}_{1:t-1}\right) \bar{\gamma}_{\boldsymbol{\theta}}(\mathbf{y}_{1:t-1}, \mathbf{x}_{1:t-1}) > 0 \right\} \tag{5.39}$$

somewhat weaker ergodicity results can be established for the PGAS method under the assumption that $\mathcal{S}_t^{\boldsymbol{\theta}} \subseteq \mathcal{Q}_t^{\boldsymbol{\theta}}$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $t \in \{1, \ldots, T\}$. The proof is similar to that for the ergodicity of the PG method (Andrieu et al., 2010) and requires that the number of iterations grows to infinity.

As mentioned before, the PGAS method is very appropriate for inference in non-Markovian latent variable models in contrast to the PG method with backward simulation. Although the Hawkes model is Markovian given the specification of the intensity in (5.20), the Markovian property of the model cannot be used to sample the intensity backwards. The reason is that the intensity in combination with the parameters of the model can be expressed as a function of the jump history $\mathbf{N}_{1:t}$, implying a certain jump path. Considering the number of jumps instead of the intensity as a persistent latent state variable, the transition density depends on the entire history of the latent process. Therefore the latent process is not conditional independent anymore. However, particle Gibbs with the ancestor sampling can be used for inference in this specification. Moreover, Lindsten et al. (2014) show that for non-Markovian models the probability that the partial reference path $\mathbf{x}'_{t+1:T}$ is preceded by particle $\mathbf{x}^i_{1:t}$ (5.37) can be approximated by

$$\begin{aligned} \hat{\gamma}^i_{t+1|T} &= \frac{p(\mathbf{y}_{1:T}, (\mathbf{x}^i_{1:t}, \mathbf{x}'_{t+1:T}))}{p(\mathbf{y}_{1:t}, \mathbf{x}^i_{1:t})} = \prod_{s=t+1}^{T} p(\mathbf{y}_s | \mathbf{x}^i_{1:t}, \mathbf{x}'_{t+1:s}) p(\mathbf{x}'_s | \mathbf{x}^i_{1:t}, \mathbf{x}'_{t+1:s}) \\ &\approx \frac{p(\mathbf{y}_{1:T}, (\mathbf{x}^i_{1:t}, \mathbf{x}'_{t+1:t+l})}{p(\mathbf{y}_{1:t}, \mathbf{x}^i_{1:t})} = \prod_{s=t+l}^{T} p(\mathbf{y}_s | \mathbf{x}^i_{1:t}, \mathbf{x}'_{t+1:s}) p(\mathbf{x}'_s | \mathbf{x}^i_{1:t}, \mathbf{x}'_{t+1:s}) \end{aligned} \tag{5.40}$$

whenever the maximum of the ratios of the individual terms in the product of (5.37) is bounded by some constant M, $M < \infty$.

In contrast to the PGAS method, which approximates an ideal Gibbs sampler, particle marginal Metropolis-Hastings (Andrieu et al., 2010) approximates an ideal marginal Metropolis-Hastings sampler. Although the PMMH sampler is theoretically preferred over the PGAS method, the sampler is often too time-consuming to use in practice as a high

number of particles and Monte Carlo iterations are required to give a somewhat accurate picture of the joint distribution of the latent variables and the parameters of the model. On the contrary, in the PGAS method fast mixing is obtained with only a few particles through the ancestor resampling step. Moreover, with the PGAS method I sample the latent variables and parameters in separate stages of our iterative Gibbs sampling procedure, which improves mixing even further.

5.3.4 Auxiliary Particle Gibbs with anchestor sampling

Combining the auxiliary particle filter described in Subsection 5.3.2 with the particle Gibbs with anchestor resampling method described in Subsection 5.3.3, the advantages of both can be exploited to infer latent variables of nonlinear models containing non-analytical and non-Markovian density functions. The auxiliary Particle Gibbs with anchestor sampling is more capable of handling the degeneracy problem compared to PGAS, as information that observations carry about the latent variables, is incorporated in their proposal distributions. Integrating the APF into the PGAS algorithm instead of the Particle Gibbs or the particle marginal Metropolis-Hastings sampler, the latent variables can be approximated without using a large number of particles saving computation time. The sampling procedure of the APGAS algorithm is outlined in Algorithm 1.

Algorithm 1: APGAS for non-Markovian latent variable models

Input: $\mathbf{y}_{1:T}, \mathbf{x}'_{1:T}, \boldsymbol{\theta}$

Output: $\hat{p}^N(\mathbf{x}_{1:T}|\mathbf{y}_{1:T}), \hat{p}^N(\mathbf{y}_{1:T})$

First-stage resampling

$t = 1$

1. Set $\mathbf{x}_0^i = \hat{\mathbf{x}}_1^i = \mathbb{E}[\mathbf{x}_t]$, $w_{0:1}^i = \frac{1}{N}$, and go to 7

$t = 2, \ldots, T-1$

2. Compute $\hat{\mathbf{x}}_{t+1}^i = \mathbb{E}[\mathbf{x}_{t+1}|\mathbf{x}_t]$, $w_{t:t+1}^i = \pi_t^i \hat{p}(\mathbf{y}_{t+1}|\hat{\mathbf{x}}_{t+1}^i)$
3. For $i \neq N$ sample $z(i) \propto w_{t:t+1}^i$
4. Compute $\tilde{w}_{t:t+1}^i = \pi_t^i \hat{\gamma}_{t+1:T}^i$ where $\hat{\gamma}_{t+1:T}^i$ is defined in (5.40)
5. Sample $z(N) \propto \tilde{w}_{t:t+1}^i$
6. Set $\mathbf{x}_{1:t}^i = \mathbf{x}_{1:t}^{z(i)}$

Particle propagation

7. For $i \neq N$ sample $\mathbf{x}_{t+1}^i \propto p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \mathbf{y}_{t+1})$
8. Set $\mathbf{x}_{t+1}^N = \mathbf{x}'_{t+1}$

Second-stage reweighting

9. Compute w_{t+1}^i as defined in (5.32) and $\pi_{t+1}^i = \frac{w_{t+1}^i}{\sum_{i=1}^N w_{t+1}^i}$
10. Compute $\hat{p}_{\boldsymbol{\theta}}^N(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}) = \left\{\frac{1}{N}\sum_{i=1}^N w_{t+1}^i\right\}\left\{\sum_{i=1}^N w_{t:t+1}^i\right\}$, and if $t < T$ go to 2

$t = T$

11. $\hat{p}^N(\mathbf{x}_{1:T}|\mathbf{y}_{1:T}) = \sum_{i=1}^N \pi_T^i \delta_{\boldsymbol{x}_{1:T}^i}$
12. $\hat{p}^N(\mathbf{y}_{1:t}) = \hat{p}^N(\mathbf{y}_1)\prod_{t=1}^{T-1}\hat{p}^N(\mathbf{y}_{t+1}|\mathbf{y}_{1:t})$

Moreover, incorporating the extended auxiliary particle filter for jump-diffusion models in the PGAS procedure, the information of the observations is also transfered into the proposal distribution of the non-persistent jump times and sizes. This way the accuracy of the inference on the persistent latent variables is improved. In Algorithm 2 the particle propagation step of Algorithm 1 is modified to account for the observational information about jumps.

Algorithm 2: APGAS for non-Markovian latent variable models containing jumps

Input: $\mathbf{y}_{1:T}, \mathbf{x}'_{1:T}, \mathbf{N}'_{1:T}, \mathbf{Z}'_{1:T}, \boldsymbol{\theta}$

Output: $\hat{p}^N(\mathbf{x}_{1:T}, \mathbf{J}_{1:T}, \mathbf{Z}_{1:T}|\mathbf{y}_{1:T}), \hat{p}^N(\mathbf{y}_{1:T})$

First-stage resampling

$t = 0$

1. Set $\mathbf{x}_0^i = \hat{\mathbf{x}}_1^i = \mathbb{E}[\mathbf{x}_t]$, $w_{0:1}^i = \frac{1}{N}$, and go to 7

$t = 1, \ldots, T-1$

2. Compute $\hat{\mathbf{x}}_{t+1}^i = \mathbb{E}[\mathbf{x}_{t+1}|\mathbf{x}_t]$, $w_{t:t+1}^i = \pi_t^i \hat{p}(\mathbf{y}_{t+1}|\hat{\mathbf{x}}_{t+1}^i)$ using (5.31)
3. For $i \neq N$ sample $z(i) \propto w_{t:t+1}^i$
4. Compute $\tilde{w}_{t:t+1}^i = \pi_t^i \hat{\gamma}_{t+1:T}^i$ where $\hat{\gamma}_{t+1:T}^i$ is defined in (5.40)
5. Sample $z(N) \propto \tilde{w}_{t:t+1}^i$
6. Set $\mathbf{x}_{1:t}^i = \mathbf{x}_{1:t}^{z(i)}$

Particle propagation

7. For $i \neq N$ sample $J_{t+1}^i \propto \hat{p}(\mathbf{y}_{t+1}|\hat{\mathbf{x}}_{t+1}^i, k)$ using (5.33)
8. For $i \neq N$ sample $Z_{t+1}^i \propto \phi(\mu_{z|r}^i, \sigma_{z|r}^{2,i})$ using (5.34)
9. For $i \neq N$ sample $\mathbf{x}_{t+1}^i \propto p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, J_{t+1}^i, Z_{t+1}^i, \mathbf{y}_{t+1})$
10. Set $\mathbf{x}_{t+1}^N = \mathbf{x}'_{t+1}$

Second-stage reweighting

11. Compute w_{t+1}^i as defined in (5.36) and $\pi_{t+1}^i = \frac{w_{t+1}^i}{\sum_{i=1}^N w_{t+1}^i}$
12. Compute $\hat{p}_{\boldsymbol{\theta}}^N(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}) = \left\{\frac{1}{N}\sum_{i=1}^N w_{t+1}^i\right\}\left\{\sum_{i=1}^N w_{t:t+1}^i\right\}$, and if $t < T$ go to 2

$t = T$

13. $\hat{p}^N(\{\mathbf{x}_{1:T}, \mathbf{N}_{1:T}, \mathbf{Z}_{1:T}\}|\mathbf{y}_{1:T}) = \sum_{i=1}^N \pi_T^i \delta_{\{\boldsymbol{x}_{1:T}^i, \mathbf{J}_{1:T}^i, \mathbf{Z}_{1:T}^i\}}$
14. $\hat{p}^N(\mathbf{y}_{1:t}) = \hat{p}^N(\mathbf{y}_1)\prod_{t=1}^{T-1}\hat{p}^N(\mathbf{y}_{t+1}|\mathbf{y}_{1:t})$

In both algorithms $\hat{\gamma}_{t+1:T}^i$ reduces to $p(\mathbf{x}_{t+1}^N|\mathbf{x}_t^i)$ in case of Markovian models.

5.3.5 Sampling of $\boldsymbol{\theta}$

Within the Gibbs sampling algorithm, the model parameters contained in the vector $\boldsymbol{\theta}$ are sampled conditional on the latent states.

In case models are estimated without options, parameters are drawn from their conjugate posterior distributions derived in earlier studies (Li et al., 2008). If no conjugate distribution

is available I employ a random walk Metropolis-Hastings algorithm to draw the parameters. In this algorithm a normal proposal is used, of which the proposal variance is adapted during the burn-in period of the Gibbs sampler (Roberts and Rosenthal, 2009).

When the models are estimated with options I employ the Adaptive Metropolis algorithm of Haario et al. (2001) in combination with the Delayed Rejection sampling (Green and Mira, 2001; Haario et al., 2006) to efficiently draw all parameters in the vector $\boldsymbol{\theta}$ at once.

The complete Gibbs algorithm in which the APGAS algorithm is used to infer the latent variables is outlined below.

Algorithm 3: Approximating an ideal Gibbs sampler using APGAS

$s = 0$

1. Set $\boldsymbol{\theta}[0]$ and $\mathbf{x}_{1:T}[0]$

$s = 0 : S - 1$

2. Draw $\boldsymbol{\theta}(s+1) \sim p(\cdot|\mathbf{y}_{1:T}, \mathbf{x}_{1:T}(s))$
3. Draw $\mathbf{x}_{1:T}(s+1) \sim \hat{p}^N_{\boldsymbol{\theta}(s+1)}(\cdot|\mathbf{y}_{1:T})$ by running Algorithm 1

The first reference trajectory can be obtained by carrying out step 3, in which the N-th particle trajectory is sampled in the same way as the trajectories of the other particles.

The APGAS algorithm can also be used to approximate the PMCMC sampler of Andrieu et al. (2010) as follows.

Algorithm 4: Approximating an ideal marginal MH sampler using APGAS

$s = 0$

1. Set $\boldsymbol{\theta}(0)$, $\mathbf{x}_{1:T}(0)$ and $\mathbf{x}'_{1:T} = \mathbf{x}_{1:T}(-1)$ and run Algorithm 1 to obtain $\hat{p}^N_{\boldsymbol{\theta}(0)}(\mathbf{y}_{1:T})$, $\hat{p}^N_{\boldsymbol{\theta}(0)}(\cdot|\mathbf{y}_{1:T})$ given $\mathbf{x}'_{1:T}$

$s = 1 : S-1$

2. Draw $\boldsymbol{\theta}(*)$ from $q\left(\boldsymbol{\theta}|\mathcal{H}_{\boldsymbol{\theta}(s)}\right)$
3. Obtain $\hat{p}^N_{\boldsymbol{\theta}(*)}(\mathbf{y}_{1:T})$ and $\hat{p}^N_{\boldsymbol{\theta}(*)}(\cdot|\mathbf{y}_{1:T})$ by running Algorithm 1 given $\mathbf{x}'_{1:T}$
4. Sample $u \sim \mathcal{U}[0,1]$ and compute the acceptance probability

$$p_a = \min\left\{\frac{p(\boldsymbol{\theta}(*))\hat{p}^N_{\boldsymbol{\theta}(*)}(\mathbf{y}_{1:T})q\left(\boldsymbol{\theta}(s)|\mathcal{H}_{\boldsymbol{\theta}(s)}\right)}{p(\boldsymbol{\theta}(s))\hat{p}^N_{\boldsymbol{\theta}(s)}(\mathbf{y}_{1:T})q\left(\boldsymbol{\theta}(*)|\mathcal{H}_{\boldsymbol{\theta}(s)}\right)}, 1\right\} \tag{5.41}$$

5. If $u \leq p_a$ draw $\mathbf{x}^*_{1:T} \sim \hat{p}^N_{\boldsymbol{\theta}^*}(\cdot|\mathbf{y}_{1:T})$ and set $\mathbf{x}'_{1:T} = \mathbf{x}_{1:T}(s-1)$, $\left\{\boldsymbol{\theta}(s+1), \mathbf{x}_{1:T}(s+1), \hat{p}^N_{\boldsymbol{\theta}(s+1)}(\mathbf{y}_{1:T})\right\} = \left\{\boldsymbol{\theta}(*), \mathbf{x}_{1:T}(*), \hat{p}^N_{\boldsymbol{\theta}(*)}(\mathbf{y}_{1:T})\right\}$, else set $\left\{\boldsymbol{\theta}(s+1), \mathbf{x}_{1:T}(s+1), \hat{p}^N_{\boldsymbol{\theta}(s+1)}(\mathbf{y}_{1:T})\right\} = \left\{\boldsymbol{\theta}(s), \mathbf{x}_{1:T}(s), \hat{p}^N_{\boldsymbol{\theta}(s)}(\mathbf{y}_{1:T})\right\}$

The extension of Algorithm 3 and 4 to account for the information of observations about jumps when estimating jump models, is straightforward.

5.3.6 Comparison with Bayes Factors and Deviance Information Criteria

Particle filters are very suitable for model comparison as the marginal likelihood conditional on the model parameters $p_{\boldsymbol{\theta}(s)}(\mathbf{y}_{1:T})$ is produced as a by product when filtering the latent states from the observed data (Pitt et al., 2012). To derive the marginal likelihood of the model (5.30), $p_{\boldsymbol{\theta}(s)}(\mathbf{y}_{1:T})$ has to be integrated over the parameter space $\boldsymbol{\Theta}$,

$$p(\mathbf{y}_{1:T}) = \int_{\boldsymbol{\Theta}} p_{\boldsymbol{\theta}(s)}(\mathbf{y}_{1:T})p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} \tag{5.42}$$

To avoid the computation of this integral I utilize the proof of Gelfand and Dey (1994). Gelfand and Dey (1994) who show that when using a proper function $g(\boldsymbol{\theta})$ the marginal

likelihood can be consistently approximated by

$$\hat{p}(\mathbf{y}_{1:T}) = \left\{ \frac{1}{S} \sum_{s=1}^{S} \frac{g(\boldsymbol{\theta}(s))}{p_{\boldsymbol{\theta}(s)}(\mathbf{y}_{1:T}) p(\boldsymbol{\theta}(s))} \right\}^{-1} \tag{5.43}$$

Calculating the marginal likelihood from (5.43), I employ for $g(\boldsymbol{\theta})$ a truncated multivariate normal distribution with mean and variance given by the first two sample moments of $\boldsymbol{\theta}$, $\overline{\boldsymbol{\theta}}$ and $\overline{\boldsymbol{\Sigma}}$. The bounds of the distribution are given by

$$\boldsymbol{\Theta} = \left\{ \boldsymbol{\theta} : \left(\boldsymbol{\theta}(s) - \overline{\boldsymbol{\theta}}\right)' \left(\overline{\boldsymbol{\Sigma}}\right)^{-1} \left(\boldsymbol{\theta}(s) - \overline{\boldsymbol{\theta}}\right) \leq \chi^2_{\alpha}(d) \right\} \tag{5.44}$$

where d is the dimension of the parameter vector and α is the percentile of the χ^2-distribution with d degrees of freedom.

The marginal likelihood can be used to compute Bayes factors and the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002). These goodness-of-fit measures are particularly appropriate for this setting as not all of the models are nested and both nested and non-nested models can be compared with the measures. The Bayes factor indicating the relative evidence for model $\mathcal{M}_a$ against model $\mathcal{M}_b$ given the observed data, is specified as $\mathcal{BF}_{ab} = p(\mathbf{y}_{1:T}|\mathcal{M}_a)/p(\mathbf{y}_{1:T}|\mathcal{M}_b)$. The DIC quantifies the ability of the posterior to predict future data generated by the estimated model. Defining $\overline{D(\boldsymbol{\theta})} = \frac{1}{S}\sum_{s=1}^{S} -2 \log p_{\boldsymbol{\theta}(s)}(\mathbf{y}_{1:T})$ and $D(\overline{\boldsymbol{\theta}}) = -2 \log p_{\overline{\boldsymbol{\theta}}}(\mathbf{y}_{1:T})$, the DIC is computed as $\mathrm{DIC} = D(\overline{\boldsymbol{\theta}}) + 2(\overline{D(\boldsymbol{\theta})} - D(\overline{\boldsymbol{\theta}}))$. $D(\overline{\boldsymbol{\theta}})$ can be approximated running the APAS algorithm using the mean of the posterior parameter draws $\overline{\boldsymbol{\theta}}$. The DIC corrects for the complexity of models by the inclusion of the penalty term, $p_D = \overline{D(\boldsymbol{\theta})} - D(\overline{\boldsymbol{\theta}})$.

5.4 Learning

In non-analytic latent variable models, the estimation of static parameters is not straightforward. In a batch context Andrieu et al. (2010) provide solutions to the estimation problem using MCMC techniques, for which they prove convergence to the true posterior distribution of the parameters. Lindsten et al. (2014) among many others, set forward on the work of Andrieu et al. (2010) making their samplers more efficient and less vulnerable to sample impoverishment. Chopin et al. (2013) and Fulop and Li (2013) show that the methods of

Andrieu et al. (2010) can also be used in a sequential context. In applications, for example in finance where new information needs to be processed quickly, it can be more appropriate to estimate parameters recursively. In a sequential approach, information can be exploited to update models while moving forward. On the contrary, batch estimation requires to redo the entire estimation procedure, which can be very time-consuming.

Both Chopin et al. (2013) and Fulop and Li (2013) use a particle filter to marginalize out the latent states and obtain an estimate of the likelihood for a given parameter vector. Assuming the target distribution $p(\boldsymbol{\theta}, \mathbf{x}_{1:t}|\mathbf{y}_{1:t})$ at t can be represented by the following set of weighted samples: $\left\{\boldsymbol{\theta}^j, \left\{\mathbf{x}_t^{i,j}, \pi_t^{i,j}\right\}_{i=1,...,N}, \hat{p}_{\boldsymbol{\theta}^j}^N(\mathbf{y}_{1:t})\right\}$ with weight s_t^j, $j = 1, ..., M$. The target distribution $p(\boldsymbol{\theta}, \mathbf{x}_{1:t+1}|\mathbf{y}_{1:t+1})$ at time $t+1$ is computed using the following recursion

$$p(\boldsymbol{\theta}, \mathbf{x}_{1:t+1}|\mathbf{y}_{1:t+1}) \propto p_{\boldsymbol{\theta}}(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}) p_{\boldsymbol{\theta}}(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{y}_{t+1}) p(\boldsymbol{\theta}, \mathbf{x}_{1:t}|\mathbf{y}_{1:t}) \tag{5.45}$$

As the target distribution is changing over time, the set of parameters has to be enriched. Otherwise the weighted sample representation of the target will be less and less accurate. To enrich the set of parameters a resample-move algorithm in the spirit of Gilks and Berzuini (2001) is applied. The parameter vectors and the particle distribution associated with these vectors are resampled using the weights, such that the weighted sample consists of high likely particles. Hereafter the samples are passing through a Metropolis-Hastings kernel that leaves the target distribution invariant. For any fixed number of state particles, N, the methods of Chopin et al. (2013) and Fulop and Li (2013) produces a sample from the true posterior distribution as the number of parameter particles, M, goes tot infinity.

The parameter learning methods of Chopin et al. (2013) and Fulop and Li (2013) are very generic as they only require to tailor a particle filter to fit the model at hand. Moreover, as particle filters are used, sequential marginal likelihood estimates are obtained as a by product. Futhermore, the methods of Chopin et al. (2013) and Fulop and Li (2013) are efficient as the resample-move steps take place more frequently at the beginning when only a short data sample has to be explored.

Besides Chopin et al. (2013) and Fulop and Li (2013), Gilks and Berzuini (2001), Liu and West (2001), Chopin (2002) and Carvalho et al. (2010) produced notable work in the area of parameter learning in general latent variable models. However, the methods of Liu and West (2001) and Carvalho et al. (2010) seem to be less accurate, while Gilks and Berzuini (2001)

and Chopin (2002) are not easy to implement. Liu and West (2001) add uncertainty to the posterior of the parameter by artificially exposing the parameters to stochastic evolution to approximate their posterior. Carvalho et al. (2010) use sufficient statistics that describe the dependence between the latent states and the parameters to approximate the distribution of the parameters. As the path of the latent states does not have to be explored again, the algorithm is fast but not so reliable when the sample size increases. Gilks and Berzuini (2001) propose to reuse particles by sampling from MCMC kernels, that leave targeted distribution of both parameters and particles unchanged. This requires to specify distribution of the parameters given the states and the other way around, which is not straightforward. To apply the methods of Chopin (2002), both a closed-form likelihood function as well as the MCMC kernels of the parameters given the states, need to be available.

To improve efficiency and reduce computation time I adjust the methods of Chopin et al. (2013) and Fulop and Li (2013) as I obtain an estimate of the sequential marginal likelihood applying the particle Gibbs methods of Lindsten et al. (2014). The following algorithm outlines how I incorporate anchestor resampling to estimate the parameter distribution over time.

Algorithm 5: Parameter learning using APGAS

Augmentation

$t = 0$

1. Set $\boldsymbol{\theta}^j$, $\left\{\mathbf{x}_0^{i,j}, \pi_0^{i,j}\right\}_{i=1,\ldots,N}$, s_0^j, $j = 1, \ldots, M$, and use step 7 and 8 to set t^*, $\hat{\boldsymbol{\theta}}$, $\mathbf{x}'_{1:t^*}$

$t = 1 : T-1$

2. For each $\boldsymbol{\theta}^j$, $j = 1, \ldots, M$, obtain $\hat{p}^N_{\boldsymbol{\theta}^j}(\mathbf{y}_{t+1}|\mathbf{y}_{1:t})$ and $\left\{\mathbf{x}_{t+1}^{i,j}, \pi_{t+1}^{i,j}\right\}_{i=1,\ldots,N}$ by cycling once through step 2 to step 12 of Algorithm 1 given $\mathbf{y}_{t+1}$, $\mathbf{x}'_{t+1}$, $\left\{\mathbf{x}_t^{i,j}, \pi_t^{i,j}\right\}_{i=1,\ldots,N}$

Reweighting

3. Update the weights of the parameter particles and the likelihood

$$s_{t+1}^j = s_t^j \times \hat{p}^N_{\boldsymbol{\theta}^j}(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}) \tag{5.46}$$

$$\hat{p}^N_{\boldsymbol{\theta}^j}(\mathbf{y}_{1:t+1}) = \hat{p}^N_{\boldsymbol{\theta}^j}(\mathbf{y}_{1:t}) \times \hat{p}^N_{\boldsymbol{\theta}^j}(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}) \tag{5.47}$$

4. Compute the effective sample size $ESS_{t+1} = \frac{1}{\sum_{j=1}^M \left(\tau_{t+1}^j\right)^2}$ where $\tau_{t+1}^j = \frac{s_{t+1}^j}{\sum_{j=1}^M s_{t+1}^j}$

Resample-move

5. If $ESS_{t+1} > B_1$ go to step 8

   else for each $j = 1, \ldots, M$ sample the particle index $z(j) \propto \tau_t^j$ and set $s_{t+1}^j = 1$,
   $\left\{\boldsymbol{\theta}^j, \left\{\mathbf{x}_{t+1}^{i,j}, \pi_{t+1}^{i,j}\right\}_{i=1,\ldots,N}, \hat{p}^N_{\boldsymbol{\theta}^j}(\mathbf{y}_{1:t+1})\right\} = \left\{\boldsymbol{\theta}^{z(j)}, \left\{\mathbf{x}_{t+1}^{i,z(j)}, \pi_{t+1}^{i,z(j)}\right\}_{i=1,\ldots,N}, \hat{p}^N_{\boldsymbol{\theta}^{z(j)}}(\mathbf{y}_{1:t+1})\right\}$

6. If the number of unique particles is bigger than B_2 go to step 8

   else for each $j = 1, \ldots, M$

   (a) Draw $\boldsymbol{\theta}^*$ from $q\left(\boldsymbol{\theta}|\mathcal{H}_{\boldsymbol{\theta}^j}\right)$

   (b) Obtain $\hat{p}^N_{\boldsymbol{\theta}^*}(\mathbf{y}_{1:t+1})$ and $\hat{p}^N_{\boldsymbol{\theta}^*}(\cdot|\mathbf{y}_{1:t+1})$ by running Algorithm 1 given $\mathbf{x}'_{1:t+1}$

   (c) Sample $u \sim \mathcal{U}[0,1]$ and compute the acceptance probability

$$p_a = \min\left\{\frac{p(\boldsymbol{\theta}^*)\hat{p}^N_{\boldsymbol{\theta}^*}(\mathbf{y}_{1:t+1})q\left(\boldsymbol{\theta}^j|\mathcal{H}_{\boldsymbol{\theta}^j}\right)}{p(\boldsymbol{\theta}^j)\hat{p}^N_{\boldsymbol{\theta}^j}(\mathbf{y}_{1:t+1})q\left(\boldsymbol{\theta}^*|\mathcal{H}_{\boldsymbol{\theta}^j}\right)}, 1\right\} \tag{5.48}$$

   (d) If $u \leq p_a$ set
   $\left\{\boldsymbol{\theta}^j, \left\{\mathbf{x}_{t+1}^{i,j}, \pi_{t+1}^{i,j}\right\}_{i=1,\ldots,N}, \hat{p}^N_{\boldsymbol{\theta}^j}(\mathbf{y}_{1:t+1})\right\} = \left\{\boldsymbol{\theta}^*, \left\{\mathbf{x}_{t+1}^{i,*}, \pi_{t+1}^{i,*}\right\}_{i=1,\ldots,N}, \hat{p}^N_{\boldsymbol{\theta}^*}(\mathbf{y}_{1:t+1})\right\}$

7. Compute some high likely value of the parameter vector $\hat{\boldsymbol{\theta}}$ such as the mean or median of $\left\{\boldsymbol{\theta}^j\right\}_{j=1,\ldots,M}$ and set $t^* = t+1$

8. If $t+1 = t^*$ set $t^* = t + t_{update}$ and update the reference trajectory $\mathbf{x}'_{1:t^*}$ given $\hat{\boldsymbol{\theta}}$ by running Algorithm 1, in which the N-th particle trajectory is sampled in the same way as the trajectories of the other particles

9. If $t+1 < T$ go to step 2

The marginal likelihood of each new observation

$$p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}) = \int p_{\boldsymbol{\theta}}(\mathbf{y}_{t+1}|\mathbf{y}_{1:t})p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} \tag{5.49}$$

can be approximated by

$$\hat{p}^{N,M}(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}) = \sum_{j=1}^{M} \tau_{t+1}^{j} \hat{p}_{\boldsymbol{\theta}^j}^{N}(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}) \tag{5.50}$$

Using the marginal likelihood, the sequential Bayes factor of two models can be constructed as follows

$$\mathcal{BF}_{t+1} = \frac{p(\mathbf{y}_{1:t+1}|\mathcal{M}_2)}{p(\mathbf{y}_{1:t+1}|\mathcal{M}_1)} = \frac{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t},\mathcal{M}_2)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t},\mathcal{M}_1)}\mathcal{BF}_t \tag{5.51}$$

The sequential Bayes factor enables us to assess the performance of different models over time, something that is impossible in batch estimation.

5.5 Application

To avoid inconsistencies I recommend to estimate the models over a large time period (Broadie et al., 2007) instead of over a large cross-sectional option panel. The cross-section of option prices and changes in the volatility smirk over time seem to be well explained by using just one option (Pan, 2002). I opt to apply my estimation methods to stock and option data on the S&P 500 index from January 1, 2000, to December 31, 2015. Depending on the method used, I suggests to take the first ten years, that is January 1, 2000, till December 31, 2010, as the in-sample period and the last 5 years, January 1, 2011, till December 31, 2015, as the out-of-sample period. Daily stock market data and a high frequency measure of the diffusive part of the daily quadratic variation, the Median Truncated Realized variance, can be downloaded from Research Data Oxford.

An option dataset containing closing bid and ask prices of European style calls can be downloaded from Option metrics. As it would be too computationally demanding to calculate option prices at a daily frequency, I suggest to follow the literature and use weekly prices for which to select Wednesday as the day of the week to include. Furthermore interest rate and dividend yield data should be downloaded for each day options are sampled. The interest rate dataset can be expanded by interpolating the yields for the maturities of the options

traded on these days assuming the dividend yields to be the same across maturities. When using the put-call parity to back out Forward prices, there is no need for dividend prices. However, as both stock and option prices are modeled and the empirical relation between the two is of interest, the risk-neutral density for calls should be derived. This density can differ from the risk-neutral density implied by the put-call parity.

The option dataset should be filtered such that options prices that are not reliable as they contain large errors, are removed. I suggest to delete options with a maturity shorter than 20 days and longer than 252 working days, an implied volatility above 70% and options of which the implied volatility of the delta is missing. Furthermore, I suggest to use a minimum price grid like 0.01 dollar below which to exclude options. Following Andersen et al. (2015a), I recommend to consider out-of-the-money calls, as in-the-money options are less actively traded and are therefore priced less accurately. For the same reason I recommend also to not consider calls that are very deep out-of-the-money. This leaves one with a set options with moneyness level above 0.85 and below 1.05.

Similar to Barone-Adesi et al. (2008) and Bollen and Whaley (2004), I opt to create a representative set of options that spans the cross-section of different maturities and moneyness levels of the options traded over the sample period. First, it is needed to divide the options into groups based on maturity and moneyness separating options by the maturities of a quarter and half a business year, and the moneyness levels of 0.9 and 0.95. This provides one with 9 groups which contain a subset of the options traded on each sample day. For each sample day and each group I recommend to select the option that is closest to the midpoint of the group to be the representative of that group. If the group is empty, I recommend to consider the option that is closest among all groups as a representitive.

Nederlandse Samenvatting

(Summary in Dutch)

Het identificeren en voorspellen van crashes op financiële markten is zeer belangrijk voor handelaren, toezichthouders op financiële markten en risicomanagement. Een reeks sterke prijsdalingen op de financiële markten gedurende een korte periode kan ernstige gevolgen hebben. Op Black Monday (19 oktober 1987) registreerde de S&P 500-index bijvoorbeeld zijn hoogste dagelijkse procentuele verlies van 20,5 %. Tijdens de recente kredietcrisis (2008) is de S&P 500-index gedurende vele dagen dramatisch gedaald, waardoor het hoogste jaarlijkse procentuele verlies van 38,5 % werd leden. Helaas zijn crashes niet eenvoudig te voorspellen, en is er een grote behoefte aan een model dat hierin kan voorzien.

Om een model voor beurscrashes te ontwikkelen, is het belangrijk om te begrijpen wat mogelijke oorzaken van dergelijke crashes zijn. Computerhandel, de toegenomen handel in afgeleide effecten, illiquiditeit, handels- en begrotingstekorten en overwaardering kunnen sterke opeenvolgende prijsdalingen veroorzaken. Belangrijker is te begrijpen dat speculatieve bubbels die leiden tot financiële crashes, het gevolg zijn van kuddegedrag van beleggers. Wanneer een groep beleggers zijn aandelen verkoopt, zorgt dit ervoor dat ook andere beleggers op dezelfde markt hun aandelen gaan verkopen. Zo zorgt het kuddegedrag van beleggers dat crashes zich lokaal versterken. Een model voor beurscrashes moet deze zelfexcitatie omvatten. Een dergelijke zelfexcitatie wordt ook waargenomen in het seismisch gedrag rondom aardbevingsreeksen. Daarbij genereert een aardbeving meestal naschokken die op hun beurt nieuwe naschokken kunnen genereren, enzovoort. Aardbevingen en aandelenrendementen hebben vergelijkbare kenmerken als de clustering van schokken in rendementen dan wel seismische activiteit, en de afhankelijkheid van deze schokken in de tijd.

Dit proefschrift richt zich op de identificatie en voorspelling van crashes op de financiële markt met Hawkes processen (hoofdstuk 2 en 4), het toetsen van deze Hawkes processen op correcte specificatie (hoofdstuk 3) en het schatten van niet-affiene Hawkes processen in continue tijd met behulp van optieprijzen (hoofdstuk 5). Hawkes processen matchen het zelfexciterend gedrag van aandelenrendementen rond een crash op een financiële markt. Dit gedrag is vergelijkbaar met de seismische activiteit rond aardbevingen. In de processen neemt de snelheid waarmee schokken optreden toe wanneer er net een schok heeft plaatsgevonden. Vervolgens daalt de frequentie waarmee schokken arriveren, als functie van de tijd die is verstreken sinds de voorgaande schok. Omdat de kans op schokken toeneemt nadat een schok heeft plaatsgevonden, worden Hawkes processen zelfexciterend genoemd.

Als eerder benoemd, vertonen schokken in seismische activiteit rondom aardbevingen en aandelenrendementen clustergedrag in de tijd. Aardbevingen vinden daarbovenop in de buurt van elkaar plaats, en vertonen zo clustergedrag in de ruimte. Ook financiële schokken lijken te clusteren in een andere dimensie dan de tijdsdimensie. De financiële crisis van 2008 (tevens genoemd in de eerste alinea van deze inleiding) laat bijvoorbeeld overlappende periodes zien waarin financiële markten onderhevig zijn aan spanning met extreme prijsbewegingen als gevolg. Zo leden de S&P 500, de Dow Jones Industrial Average (DJI) en de NASDAQ allemaal top 20 procent verliezen op 29 september, 15 oktober en 1 december 2008. Bovendien daalden op 29 september zowel de koers van de euro/dollar als de koers van het pond/dollar aanzienlijk, terwijl de Amerikaanse obligatiekoers fors steeg. Op 16 oktober, slechts een dag nadat de grote Amerikaanse aandelenmarkten crashten, en op 1 december daalden beide valuta's opnieuw scherp. Bovendien vertoonde de Amerikaanse obligatiekoers 4 dagen na deze data een opmerkelijke stijging. Om de afhankelijkheid tussen financiële markten op het gebied van crashes te modelleren breiden we het univariate Hawkes modelleringskader uit. Op die manier zorgen we ervoor dat extreme prijsveranderingen op de ene financiële markt het optreden en/of de omvang van extreme prijsveranderingen op andere markten kan uitlokken.

Naast historische aandelenrendementen, bevatten ook optieprijzen informatie over toekomstige aandelenrendementen. De informatie in optieprijzen kan dan ook worden gebruikt om Hawkes modellen voor aandelenrendementen nauwkeuriger te schatten. De zogenaamde affiene modelspecificaties bevatten optieprijzen in gesloten vorm. Hoewel dit modelkalibratie vergemakkelijkt, lijken niet-affiene specificaties aandelenkoersen beter te beschrijven.

Markov chain Monte Carlo, particle filtering en learning methoden bieden uitkomst bij het schatten van dergelijke modellen. Gebruikmakend van informatie uit optieprijzen, kan de compensatie die beleggers ontvangen voor volaliteit risico en het risico op schokken worden afgeleid. De compensatie voor het risico op schokken, is tevens een indicatie is van de angst van beleggers. Daarom biedt het berekenen van volaliteits- en schokcomponenten in risicopremies met behulp van optieprijzen, belangrijke informatie over de toestand en ontwikkeling van de financile markt met verstrekkende implicaties voor spreiding van investering, afdekking en risicomanagement.

Hoofdstuk 2 is gebaseerd op Gresnigt et al. (2015). In dit hoofdstuk gebruiken we het Hawkes modelleringskader om waarschijnlijkheidsvoorspellingen te maken voor een aankomende crash op de financiële markt op de middellange termijn (bijvoorbeeld ergens binnen de komende vijf dagen). Daarbij interpreteren we financiële crashes als aardbevingen op de financiële markt. Dit stelt ons in staat een waarschuwingssyteem (Early Warning System) voor crashdagen te ontwikkelen. Wanneer we dit waarschuwingssyteem toetsen op S&P 500 data tijdens de recente financiële crisis, vinden we dat het Hawkes model in staat is crashes te voorspellen. Bovendien is het model in staat informatie uit de aandelenrendementen te benutten die niet wordt gevat door bekende en veelgebruikte volatiliteitsmodellen. Het waarschuwingssyteem op basis van Hawkes modellen overtreft het waarschuwingssysteem op basis van de volatiliteitsmodellen die extreme prijsbewegingen voorspellen, terwijl voorspellen veel minder tijdrovend is.

Om nauwkeurig het optreden van extreme prijsbewegingen op financiële markten met behulp van Hawkes modellen te identificeren en te voorspellen, is het nodig dat deze modellen aantoonbaar correct gespecificeerd zijn. Hoofdstuk 3 voorziet hierin. In dit hoofdstuk, gebaseerd op Gresnigt et al. (2016a), ontwikkelen we verschillende specificatietoetsen voor Hawkes modellen op basis van het Lagrange Multiplier (LM) principe. De toetsfocus ligt op het uitbreiden van een univariate model naar een multivariate model. We onderzoeken zo of er een voorwaardelijke afhankelijkheid is tussen reeksen extreme prijsveranderingen in (verschillende) financiële markten. Simulaties tonen aan dat de toetsen goed presteren voor steekproefgroottes die in de praktijk meestal worden onderzocht. Door de specificatietoetsen op Amerikaanse aandelen, obligaties en wisselkoersen toe te passen, vinden we sterk bewijs voor de afhankelijkheid tussen de schokken op verschillende financiële markten. Daarom ra-

den we aan met deze afhankelijkheid in de modelspecificatie van Hawkes modellen rekening mee te houden.

Hoofdstuk 4 is gebaseerd op Gresnigt et al. (2016b). In dit hoofdstuk onderzoeken we of het opnemen van de afhankelijkheid op het gebied van crashes tussen financiële markten, voorspellingen van het optreden van deze crashes verbeteren. We volgen de aanbeveling van hoofdstuk 3 (Gresnigt et al. (2016a)), en gebruiken Hawkes modellen waarin schokken kunnen worden uitgelokt door zowel zelfexcitatie als crossexcitatie om onze voorspellingen te maken. De modellen worden toegepast op Amerikaanse aandelen, obligaties en wisselkoersen. Wanneer we crashes over de periode van de financiële crisis voorspellen, vinden we dat de Hawkes modellen waarin crossexcitatie effecten zijn gevat, aanzienlijk nauwkeuriger voorspelen dan modellen zonder deze effecten.

Hoofdstuk 5 bevat een onderzoeksvoorstel. Het hoofdstuk stelt een raamwerk voor waarin optieprijzen worden gebruikt om Hawkes modelspecificaties voor aandelenrendementen in continue tijd, nauwkeuriger te schatten. Het gebruiken van optieprijzen bij de schatting van Hawkes modellen verhoogt niet alleen de nauwkeurigheid, maar maakt het ook mogelijk compensaties die beleggers verlangen voor het risico op volatiliteit en schokken af te leiden. Dit biedt inzicht in de toestand en ontwikkeling van de financile markt met belangrijke richtlijnen voor risicomanagement. Met het raamwerk is het mogelijk modellen te schatten waarin aandelenprijzen geen analytische karakteristieke functie hebben (modellen van het niet-affiene type). Met Markov chain Monte Carlo, particle filtering en learning methoden worden modellen efficint geschat en achtereenvolgens beoordeeld, zodat modellen snel kunnen worden bijgewerkt wanneer nieuwe informatie binnenkomt. Inzichtgevende latente processen van volatiliteit en schokken worden geproduceerd als bijproduct van de schattingsprocedure. Het schattingsraamwerk kan worden aangepast aan de betreffende applicatie. Het raamwerk kan bijvoorbeeld worden uitgebreid om multivariate Hawkes modellen te schatten omdat dit in het raamwerk, anders dan in veel klassieke raamwerken, geen directe optimalisatie van een multidimensionale integraal vereist. Aangezien hoofdstuk 3 en 4 laten zien dat sprongintensiteiten wederzijds opwindend zijn, maakt dit de schattingsmethode zeer aantrekkelijk voor verder onderzoek. Bovendien maakt het raamwerk het mogelijk om informatie uit opties op een lagere frequentie te gebruiken dan de informatie uit aandelen.

Bibliography

Aït-Sahalia, Y., 2004. Disentangling diffusion from jumps. Journal of Financial Economics 74 (3), 487–528.

Aït-Sahalia, Y., Cacho-Diaz, J., Laeven, R. J., 2015. Modeling financial contagion using mutually exciting jump processes. Journal of Financial Economics 117 (3), 585–606.

Aït-Sahalia, Y., Laeven, R. J., Pelizzon, L., 2014. Mutual excitation in eurozone sovereign cds. Journal of Econometrics 183 (2), 151–167.

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

Andersen, T. G., Benzoni, L., Lund, J., 2002. An empirical investigation of continuous-time equity return models. The Journal of Finance 57 (3), 1239–1284.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., Diebold, F. X., 2006. Volatility and correlation forecasting. Elsevier, pp. 777–878.

Andersen, T. G., Fusari, N., Todorov, V., 2015a. Parametric inference and dynamic state recovery from option panels. Econometrica 83 (3), 1081–1145.

Andersen, T. G., Fusari, N., Todorov, V., 2015b. The risk premia embedded in index options. Journal of Financial Economics 117 (3), 558–584.

Andrews, D. W., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61 (4), 821–856.

Andrews, D. W., 2001. Testing when a parameter is on the boundary of the maintained hypothesis. Econometrica 69 (3), 683–734.

Andrieu, C., Doucet, A., Holenstein, R., 2010. Particle markov chain monte carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72 (3), 269–342.

Bacry, E., Dayri, K., Muzy, J.-F., 2012. Non-parametric kernel estimation for symmetric hawkes processes: Application to high frequency financial data. European Physical Journal B 85 (5), 1–12.

Bacry, E., Mastromatteo, I., Muzy, J.-F., 2015. Hawkes processes in finance. Market Microstructure and Liquidity 1 (1), 1550005.

Bae, K.-H., Karolyi, G. A., 1994. Good news, bad news and international spillovers of stock return volatility between japan and the us. Pacific-Basin Finance Journal 2 (4), 405–438.

Bae, K.-H., Karolyi, G. A., Stulz, R. M., 2003. A new approach to measuring financial contagion. Review of Financial studies 16 (3), 717–763.

Baker, M., Stein, J. C., 2004. Market liquidity as a sentiment indicator. Journal of Financial Markets 7 (3), 271–299.

Bakshi, G., Cao, C., Chen, Z., 1997. Empirical performance of alternative option pricing models. The Journal of Finance 52 (5), 2003–2049.

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

Baldovin, F., Bovina, D., Camana, F., Stella, A. L., 2011. Modeling the non-markovian, non-stationary scaling dynamics of financial markets. In: Econophysics of order-driven markets. Springer, pp. 239–252.

Baldovin, F., Camana, F., Caporin, M., Caraglio, M., Stella, A. L., 2015. Ensemble properties of high-frequency data and intraday trading rules. Quantitative Finance 15 (2), 231–245.

Baldovin, F., Camana, F., Caraglio, M., Stella, A. L., Zamparo, M., 2013. Aftershock prediction for high-frequency financial markets dynamics. In: Econophysics of Systemic Risk and Network Dynamics. Springer, pp. 49–58.

Barone-Adesi, G., Engle, R. F., Mancini, L., 2008. A garch option pricing model with filtered historical simulation. The review of financial studies 21 (3), 1223–1258.

Bates, D. S., 2000. Post-'87 crash fears in the s&p 500 futures option market. Journal of Econometrics 94 (1), 181–238.

Bates, D. S., 2006. Maximum likelihood estimation of latent affine processes. Review of Financial Studies 19 (3), 909–965.

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

Bauwens, L., Hautsch, N., 2009. Modelling financial high frequency data using point processes. Springer Berlin Heidelberg.

Beine, M., Cosma, A., Vermeulen, R., 2010. The dark side of global integration: Increasing tail dependence. Journal of Banking & Finance 34 (1), 184–192.

Bekaert, G., Ehrmann, M., Fratzscher, M., Mehl, A., 2010. Global crises and equity market contagion, working paper, NBER.

Berkowitz, J., Christoffersen, P., Pelletier, D., 2011. Evaluating value-at-risk models with desk-level data. Management Science 57 (12), 2213–2227.

Black, F., 1976. Studies of stock price volatility changes. In: Proceedings of the 1976 Meetings of the American Statistical Association. pp. 171–181.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. The Journal of Political Economy 81 (3), 637–654.

Bollen, N. P., Whaley, R. E., 2004. Does net buying pressure affect the shape of implied volatility functions? The Journal of Finance 59 (2), 711–753.

Bollerslev, T., Todorov, V., 2011. Tails, fears, and risk premia. The Journal of Finance 66 (6), 2165–2211.

Bollerslev, T., Todorov, V., Xu, L., 2015. Tail risk premia and return predictability. Journal of Financial Economics 118 (1), 113–134.

Booth, G. G., Martikainen, T., Tse, Y., 1997. Price and volatility spillovers in scandinavian stock markets. Journal of Banking & Finance 21 (6), 811–823.

Bormetti, G., Calcagnile, L. M., Treccani, M., Corsi, F., Marmi, S., Lillo, F., 2015. Modelling systemic price cojumps with hawkes factor models. Quantitative Finance 15 (7), 1–20.

Boswijk, H. P., Laeven, R. J., Lalu, A., 2015. Asset returns with self-exciting jumps: Option pricing and estimation with a continuum of moments, working paper, University of Amsterdam.

Bowsher, C. G., 2002. Modelling security market events in continuous time: Intensity based, multivariate point process models, working paper, Oxford University.

Bowsher, C. G., 2007. Modelling security market events in continuous time: Intensity based, multivariate point process models. Journal of Econometrics 141 (2), 876–912.

Breusch, T. S., Pagan, A. R., 1980. The lagrange multiplier test and its applications to model specification in econometrics. The Review of Economic Studies 47 (1), 239–253.

Broadie, M., Chernov, M., Johannes, M., 2007. Model specification and risk premia: Evidence from futures options. The Journal of Finance 62 (3), 1453–1490.

Byström, H. N., 2004. Managing extreme risks in tranquil and volatile markets using conditional extreme value theory. International Review of Financial Analysis 13 (2), 133–152.

Campbell, J. Y., Thompson, S. B., 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? Review of Financial Studies 21 (4), 1509–1531.

Candelon, B., Dumitrescu, E.-I., Hurlin, C., 2012. How to evaluate an early-warning system: Toward a unified statistical framework for assessing financial crises forecasting methods. IMF Economic Review 60 (1), 75–113.

Carvalho, C. M., Johannes, M. S., Lopes, H. F., Polson, N. G., et al., 2010. Particle learning and smoothing. Statistical Science 25 (1), 88–106.

Chavez-Demoulin, V., Davison, A. C., McNeil, A. J., 2005. Estimating value-at-risk: A point process approach. Quantitative Finance 5 (2), 227–234.

Chavez-Demoulin, V., McGill, J., 2012. High-frequency financial data modeling using hawkes processes. Journal of Banking & Finance 36 (12), 3415–3426.

Chernov, M., Gallant, A. R., Ghysels, E., Tauchen, G., 2003. Alternative models for stock price dynamics. Journal of Econometrics 116 (1), 225–257.

Chib, S., Nardari, F., Shephard, N., 2002. Markov chain monte carlo methods for stochastic volatility models. Journal of Econometrics 108 (2), 281–316.

Chopin, N., 2002. A sequential particle filter method for static models. Biometrika 89 (3), 539–552.

Chopin, N., Jacob, P. E., Papaspiliopoulos, O., 2013. Smc2: an efficient algorithm for sequential analysis of state space models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75 (3), 397–426.

Christoffersen, P., Jacobs, K., Mimouni, K., 2010. Volatility dynamics for the s&p500: Evidence from realized volatility, daily returns, and option prices. Review of Financial Studies 23 (8), 3141–3189.

Christoffersen, P. F., 1998. Evaluating interval forecasts. International Economic Review 39 (4), 841–862.

Chung, S.-L., Tsai, W.-C., Wang, Y.-H., Weng, P.-S., 2011. The information content of the s&p 500 index and vix options on the dynamics of the s&p 500 index. Journal of Futures Markets 31 (12), 1170–1201.

Cicerone, R. D., Ebel, J. E., Britton, J., 2009. A systematic compilation of earthquake precursors. Tectonophysics 476 (3), 371–396.

Clark, T. E., West, K. D., 2007. Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138 (1), 291–311.

Clements, A., Liao, Y., et al., 2013. The dynamics of co-jumps, volatility and correlation, working paper, NCER.

Connolly, R. A., Wang, F. A., 2003. International equity market comovements: Economic fundamentals or contagion? Pacific-Basin Finance Journal 11 (1), 23–43.

Cont, R., 2001. Empirical properties of asset returns: Stylized facts and statistical issues. Quantitative Finance 1 (1), 223–236.

Creal, D., 2012. A survey of sequential monte carlo methods for economics and finance. Econometric reviews 31 (3), 245–296.

Daley, D. J., Vere-Jones, D., 2005. An introduction to the theory of point processes: Volume I: Elementary Theory and Methods. Springer.

Davidson, R., MacKinnon, J. G., 2004. Econometric theory and methods. Vol. 5. Oxford University Press New York.

Del Moral, P., 2004. Feynman-Kac formulae: Genealogical and interacting particle systems with applications. Springer New York.

Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics 13 (3), 253–265.

Dijk, D., Franses, P. H., 2003. Selecting a nonlinear time series model using weighted tests of equal forecast accuracy. Oxford Bulletin of Economics and Statistics 65 (s1), 727–744.

Duffie, D., Pan, J., Singleton, K., 2000. Transform analysis and asset pricing for affine jump-diffusions. Econometrica 68 (6), 1343–1376.

Durham, G. B., 2013. Risk-neutral modeling with affine and nonaffine models. Journal of Financial Econometrics 11 (4), 650–681.

Embrechts, P., Liniger, T., Lin, L., et al., 2011. Multivariate hawkes processes: An application to financial data. Journal of Applied Probability 48 (A), 367–378.

Engle, R. F., 1982. A general approach to lagrange multiplier model diagnostics. Journal of Econometrics 20 (1), 83–104.

Engle, R. F., Manganelli, S., 2004. Caviar: Conditional autoregressive value at risk by regression quantiles. Journal of Business & Economic Statistics 22 (4), 367–381.

Engle, R. F., Russell, J. R., 1998. Autoregressive conditional duration: A new model for irregularly spaced transaction data. Econometrica 66 (5), 1127–1162.

Eraker, B., 2004. Do stock prices and volatility jump? reconciling evidence from spot and option prices. The Journal of Finance 59 (3), 1367–1403.

Eraker, B., Johannes, M., Polson, N., 2003. The impact of jumps in volatility and returns. The Journal of Finance 58 (3), 1269–1300.

Errais, E., Giesecke, K., Goldberg, L. R., 2010. Affine point processes and portfolio credit risk. SIAM Journal on Financial Mathematics 1 (1), 642–665.

Eun, C. S., Shim, S., 1989. International transmission of stock market movements. Journal of Financial and Quantitative Analysis 24 (2), 241–256.

Fan, Z., Laeven, R. J., van den Goorbergh, R., 2015. Asymmetric excitation and the us bias in portfolio choice, working paper, University of Amsterdam.

Fang, F., Oosterlee, C. W., 2008. A novel pricing method for european options based on fourier-cosine series expansions. SIAM Journal on Scientific Computing 31 (2), 826–848.

Filimonov, V., Sornette, D., 2015. Apparent criticality and calibration issues in the hawkes self-excited point process model: Application to high-frequency financial data. Quantitative Finance 15 (8), 1293–1314.

Fischer, K. P., Palasvirta, A., 1990. High road to a global marketplace: The international transmission of stock market fluctuations. Financial Review 25 (3), 371–394.

Fulop, A., Li, J., 2013. Efficient learning via simulation: A marginalized resample-move approach. Journal of Econometrics 176 (2), 146–161.

Fulop, A., Li, J., Yu, J., 2014. Self-exciting jumps, learning, and asset pricing implications. Review of Financial Studies 28 (3), 876–912.

Gelfand, A. E., Dey, D. K., 1994. Bayesian model choice: asymptotics and exact calculations. Journal of the Royal Statistical Society: Series B (Methodological), 501–514.

Giacomini, R., Komunjer, I., 2005. Evaluation and combination of conditional quantile forecasts. Journal of Business & Economic Statistics 23 (4), 416–431.

Giacomini, R., White, H., 2006. Tests of conditional predictive ability. Econometrica 74 (6), 1545–1578.

Gilks, W. R., Berzuini, C., 2001. Following a moving targetmonte carlo inference for dynamic bayesian models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63 (1), 127–146.

Glosten, L. R., Jagannathan, R., Runkle, D. E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. The Journal of Finance 48 (5), 1779–1801.

Gonzalo, J., Olmo, J., 2005. Contagion versus flight to quality in financial markets, working paper, Universidad Carlos III Madrid.

Gourieroux, C., Monfort, A., Renne, J.-P., 2014. Pricing default events: Surprise, exogeneity and contagion. Journal of Econometrics 182 (2), 397–411.

Green, P. J., Mira, A., 2001. Delayed rejection in reversible jump metropolis–hastings. Biometrika 88 (4), 1035–1053.

Gresnigt, F., Kole, E., Franses, P. H., 2015. Interpreting financial market crashes as earthquakes: A new early warning system for medium term crashes. Journal of Banking & Finance 56 (1), 123–139.

Gresnigt, F., Kole, E., Franses, P. H., 2016a. Specification testing in hawkes models. Journal of Financial Econometrics 15 (1), 139–171.

Gresnigt, F., Kole, E., Franses, P. H., 2016b. Specification testing in hawkes models. Journal of Forecasting.

Grothe, O., Korniichuk, V., Manner, H., 2014. Modeling multivariate extreme events using self-exciting point processes. Journal of Econometrics 182 (2), 269–289.

Haario, H., Laine, M., Mira, A., Saksman, E., 2006. Dram: efficient adaptive mcmc. Statistics and Computing 16 (4), 339–354.

Haario, H., Saksman, E., Tamminen, J., et al., 2001. An adaptive metropolis algorithm. Bernoulli 7 (2), 223–242.

Hamao, Y., Masulis, R. W., Ng, V., 1990. Correlations in price changes and volatility across international stock markets. Review of Financial studies 3 (2), 281–307.

Hamilton, J. D., 1996. Specification testing in markov-switching time-series models. Journal of Econometrics 70 (1), 127–157.

Hansen, B. E., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica 64 (2), 413–430.

Hansen, P. R., Lunde, A., 2005. A forecast comparison of volatility models: Does anything beat a garch (1, 1)? Journal of Applied Econometrics 20 (7), 873–889.

Hardebeck, J. L., Felzer, K. R., Michael, A. J., 2008. Improved tests reveal that the accelerating moment release hypothesis is statistically insignificant. Journal of Geophysical Research 113 (B8).

Hardiman, S., Bercot, N., Bouchaud, J.-P., 2013. Critical reflexivity in financial markets: A hawkes process analysis. Physical Review E 85 (5), 056108.

Hartmann, P., Straetmans, S., De Vries, C. G., 2004. Asset market linkages in crisis periods. Review of Economics and Statistics 86 (1), 313–326.

Hawkes, A. G., 1971. Point spectra of some mutually exciting point processes. Journal of the Royal Statistical Society: Series B (Methodological), 438–443.

Helmstetter, A., Sornette, D., 2002. Subcritical and supercritical regimes in epidemic models of earthquake aftershocks. Journal of Geophysical Research 107 (B10).

Herrera, R., Schipp, B., 2009. Self-exciting extreme value models for stock market crashes. In: Statistical inference, econometric analysis and matrix algebra. Physica-Verlag HD, pp. 209–231.

Heston, S. L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. Review of Financial Studies 6 (2), 327–343.

Hewlett, P., 2006. Clustering of order arrivals, price impact and trade path optimisation. In: Workshop on Financial Modeling with Jump processes, Ecole Polytechnique. pp. 6–8.

Hu, L., 2006. Dependence patterns across financial markets: A mixed copula approach. Applied Financial Economics 16 (10), 717–729.

Hua, J., Manzan, S., 2013. Forecasting the return distribution using high-frequency volatility measures. Journal of Banking & Finance 37 (11), 4381–4403.

Hull, J., White, A., 1987. The pricing of options on assets with stochastic volatilities. The Journal of Finance 42 (2), 281–300.

Ignatieva, K., Rodrigues, P., Seeger, N., 2015. Empirical analysis of affine versus nonaffine variance specifications in jump-diffusion models for equity indices. Journal of Business & Economic Statistics 33 (1), 68–75.

Johannes, M. S., Polson, N. G., Stroud, J. R., 2009. Optimal filtering of jump diffusions: Extracting latent states from asset prices. Review of Financial Studies 22 (7), 2759–2799.

Jones, C. S., 2003. The dynamics of stochastic volatility: Evidence from underlying and options markets. Journal of Econometrics 116 (1), 181–224.

Kaeck, A., Alexander, C., 2012. Volatility dynamics for the s&p 500: Further evidence from non-affine, multi-factor jump diffusions. Journal of Banking & Finance 36 (11), 3110–3121.

Kanas, A., 1998. Volatility spillovers across equity markets: European evidence. Applied Financial Economics 8 (3), 245–256.

Kindleberger, C. P., Aliber, R. Z., 2011. Manias, panics and crashes: A history of financial crises. Palgrave Macmillan.

King, M. A., Wadhwani, S., 1990. Transmission of volatility between stock markets. Review of Financial studies 3 (1), 5–33.

Koutmos, G., Booth, G. G., 1995. Asymmetric volatility transmission in international stock markets. Journal of International Money and Finance 14 (6), 747–762.

Lando, D., Nielsen, M. S., 2010. Correlation in corporate defaults: Contagion or conditional independence? Journal of Financial Intermediation 19 (3), 355–372.

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

Li, H., Wells, M. T., Cindy, L. Y., 2008. A bayesian analysis of return dynamics with lévy jumps. Review of Financial Studies 21 (5), 2345–2378.

Lin, W.-L., Engle, R. F., Ito, T., 1994. Do bulls and bears move across borders? international transmission of stock returns and volatility. Review of Financial Studies 7 (3), 507–538.

Lindsten, F., Jordan, M. I., Schön, T. B., 2014. Particle gibbs with ancestor sampling. Journal of Machine Learning Research 15 (1), 2145–2184.

Liniger, T. J., 2009. Multivariate hawkes processes. Ph.D. thesis, Swiss Federal Institute of Technology Zürich.

Liu, J., West, M., 2001. Combined parameter and state estimation in simulation-based filtering. In: Sequential Monte Carlo methods in practice. Springer, pp. 197–223.

Longin, F., Solnik, B., 1995. Is the correlation in international equity returns constant: 1960–1990? Journal of International Money and Finance 14 (1), 3–26.

Lynch, A. W., Mendenhall, R. R., 1996. New evidence on stock price effects associated with charges in the s&p 500 index, working paper, New York University.

Mashal, R., Zeevi, A., 2002. Beyond correlation: Extreme co-movements between financial assets, working paper, Colombia University.

Merton, R. C., 1976. Option pricing when underlying stock returns are discontinuous. Journal of Financial Economics 3 (1-2), 125–144.

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

Ogata, Y., 1981. On lewis' simulation method for point processes. IEEE Transactions on Information Theory 27 (1), 23–31.

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

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

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

Pan, J., 2002. The jump-risk premia implicit in options: Evidence from an integrated time-series study. Journal of Financial Economics 63 (1), 3–50.

Pesaran, M. H., Timmermann, A., 1995. Predictability of stock returns: Robustness and economic significance. The Journal of Finance 50 (4), 1201–1228.

Petersen, A. M., Wang, F., Havlin, S., Stanley, H. E., 2010. Market dynamics immediately before and after financial shocks: Quantifying the omori, productivity, and bath laws. Physical Review E 82 (3), 036114.

Pitt, M. K., dos Santos Silva, R., Giordani, P., Kohn, R., 2012. On some properties of markov chain monte carlo simulation methods based on the particle filter. Journal of Econometrics 171 (2), 134–151.

Pitt, M. K., Shephard, N., 1999. Filtering via simulation: Auxiliary particle filters. Journal of the American Statistical Association 94 (446), 590–599.

Poon, S.-H., Rockinger, M., Tawn, J., 2003. Modelling extreme-value dependence in international stock markets. Statistica Sinica 13 (4), 929–953.

Poon, S.-H., Rockinger, M., Tawn, J., 2004. Extreme value dependence in financial markets: Diagnostics, models, and financial implications. Review of Financial Studies 17 (2), 581–610.

Rasmussen, J. G., 2013. Bayesian inference for hawkes processes. Methodology and Computing in Applied Probability 15 (3), 623–642.

Rikitake, T., 1978. Biosystem behaviour as an earthquake precursor. Tectonophysics 51 (1), 1–20.

Roberts, G. O., Rosenthal, J. S., 2009. Examples of adaptive mcmc. Journal of Computational and Graphical Statistics 18 (2), 349–367.

Rogers, A. J., 1986. Modified lagrange multiplier tests for problems with one-sided alternatives. Journal of Econometrics 31 (3), 341–361.

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

Santa-Clara, P., Yan, S., 2010. Crashes, volatility, and the equity premium: Lessons from s&p 500 options. The Review of Economics and Statistics 92 (2), 435–451.

Selçuk, F., Gençay, R., 2006. Intraday dynamics of stock market returns and volatility. Physica A 367, 375–387.

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

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64 (4), 583–639.

Sun, W., Rachev, S., Fabozzi, F. J., Kalev, P. S., 2009. A new approach to modeling co-movement of international equity markets: Evidence of unconditional copula-based simulation of tail dependence. Empirical Economics 36 (1), 201–229.

Van Oordt, M. R., Zhou, C., 2012. The simple econometrics of tail dependence. Economics Letters 116 (3), 371–373.

Veen, A., Schoenberg, F. P., 2008. Estimation of space-time branching process models in seismology using an em-type algorithm. Journal of the American Statistical Association 103 (482), 614–624.

Weber, P., Wang, F., Vodenska-Chitkushev, I., Havlin, S., Stanley, H. E., 2007. Relation between volatility correlations in financial markets and omori processes occurring on all scales. Physical Review E 76 (1), 016109.

Yu, J., 2005. On leverage in a stochastic volatility model. Journal of Econometrics 127 (2), 165–178.

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

Zhuang, J., Ogata, Y., Vere-Jones, D., 2004. Analyzing earthquake clustering features by using stochastic reconstruction. Journal of Geophysical Research 109 (B5).

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

606. F. Gresnigt, *Identifying and predicting financial earthquakes using Hawkes processes*