Point process models for extreme returns: Harnessing implied volatility

Abstract

Forecasting the risk of extreme losses is an important issue in the management of financial risk. There has been a great deal of research examining how option implied volatilities (IV) can be used to forecast asset return volatility. However, the role of IV in the context of predicting extreme risk has received relatively little attention. The potential benefit of IV in forecasting extreme risk is considered within a range of models beginning with the traditional GARCH based approach, along with a number of novel point process models. Univariate models where IV is included as an exogenous variable are considered along with a novel bivariate approach where extreme movements in IV are treated as another point process. It is found that in the context of forecasting Value-at-Risk, the bivariate models produce the most accurate forecasts across a wide range of scenarios.

JEL classification: C32; C53; C58.

Keywords: Implied volatility, Hawkes process, Peaks over threshold, Point process, Extreme events

1. Introduction

Modeling and forecasting extreme losses is an important issue in the management of financial risk meaning that accurate estimates of risk measures such as Value-at-Risk (VaR) have attracted a great deal of research attention. A successful model for dealing with these extreme loss events must capture their tendency to cluster in time.

A number of approaches to deal with the clustering of events have been proposed. McNeil and Frey (2000) develop a two stage method where GARCH models are first applied to model the general time variation in volatility with extreme value theory (EVT) techniques then applied to the residuals. Chavez-Demoulin et al. (2005) propose a novel Peaks Over Threshold (POT) approach for modelling extreme events. To deal with event clustering they employ a self-exciting marked point process, specifically a Hawkes-POT process, where the intensity of the occurrence

of extreme events depends on the past events and their associated size or marks. Herrera and Schipp (2013) extend the Hawkes-POT framework of Chavez-Demoulin et al. (2005) in proposing a duration based model to capture the clustering in extreme loss events.

While they have not been considered in this specific context, option implied volatilities (IV) have been widely used for forecasting volatility. As the volatility of the returns on the underlying asset price is an input into option pricing models, an expectation (risk neutral) of volatility is required when pricing options. While IV is a risk neutral estimate, it is well known that IV indices are negatively correlated with the level of stock market indices and are an important measure of short-term expected risk (see, Bekaert and Wu, 2000; Wagner and Szimayer, 2004; Giot, 2005; Becker et al., 2009; Lin and Chang, 2010; Bekaert and Hoerova, 2014, among others), and have been found to be a useful forecast of physical spot volatility in many studies, see Poon and Granger (2003). Blair et al. (2001) find the inclusion of IV as an exogenous variable in GARCH models to be beneficial in terms of forecasting. While not focusing on forecasting, Becker et al. (2009) show that IV contains useful information about future jump activity in returns, which is likely to reflect extreme movements in prices.

Very few studies have focused on the complex extremal dependence between IV and equity returns. Aboura and Wagner (2016) investigate the asymmetric relationship between daily S&P 500 index returns and VIX index changes revealing a contemporaneous volatility-return tail dependence for negative extreme events though not for positive returns. Peng and Ng (2012) analyse the cross-market dependence between five of the most important equity markets and their corresponding volatility indices, finding evidence of asymmetric tail dependence. Hilal et al. (2011) propose a conditional approach for capturing extremal dependence between daily returns on VIX futures and the S&P500. Their empirical analysis shows that VIX futures returns are very sensitive to stock market downside risk.

In this paper, the analysis moves beyond the role of IV in forecasting total volatility to focus on the link to extreme losses and addresses two main questions.

1. How are extreme shocks in an IV index and extreme events in its respective stock market return related?
2. Can this relationship be harnessed to provide superior forecasts of extreme returns?

To address these issues, an approach utilising IV within intensity based point process models for extreme returns is proposed. The first model treats IV as an exogenous variable influencing the intensity and the size distribution of extreme events. A novel alternative view is also proposed based on a bivariate Hawkes-POT model. Extreme movements in IV are treated as events themselves, with their impact on extreme events in equity returns captured through a bivariate Hawkes-POT

model. Performance of the proposed methods will be analysed in the context of forecasting extreme losses within a VaR framework. The benchmark approach follows both the earlier forecasting literature in that IV is used as an exogenous variable within the GARCH-EVT framework and the bad environments, good environment (BEGE) model of Bekaert et al. (2015).

An empirical analysis is undertaken where forecasts of the risk of extreme returns are generated for five major equity market indices using their associated IV indices. These forecasts are based on GARCH-EVT, BEGE, univariate and bivariate Hawkes-POT models, and take the form of VaR estimates at a range of levels of significance. It is found that GARCH based forecasts which include IV are often inaccurate. Univariate Hawkes-POT and BEGE models where IV is treated as an exogenous variable outperform the GARCH forecasts, though their forecasts do fail a number of tests for VaR adequacy. The bivariate Hawkes-POT models, where the timing of past extreme increases in IV are treated as a point process, lead to the most accurate forecasts of extreme risk in the widest set of scenarios. The results of this paper show that while IV is beneficial for forecasting extreme risk in equity returns, the framework within which it is used is important. The superior approach is to treat extreme increases in IV as a point process within a bivariate model for extreme returns.

The paper proceeds as follows. Section 2 outlines the traditional GARCH-EVT framework, the BEGE model, and introduces the proposed univariate and bivariate Hawkes point process models. Section 3 describes how VaR forecasts are generated and evaluated. Section 3.1 outlines the equity market and associated IV indices. Section 4 presents in-sample estimation results for the full range of models considered along with the results from tests of forecast accuracy. Section 5 provides concluding comments.

2. Methodology

This section introduces the competing approaches for forecasting extreme losses in the context of VaR predictions. The first is based on the classic GARCH approach where IV is used as an exogenous variable. The specifications considered here are the standard GARCH model of Bollerslev (1986), the GJR-GARCH models of Glosten et al. (1993), and the exponential GARCH (EGARCH) of Nelson (1991). The next approach considered is the BEGE (Bad environment good environment) model of Bekaert et al. (2015) which offers a flexible conditional distribution to describe returns. The approach proposed here utilizes the Hawkes-POT framework introduced in the one-dimensional case by Chavez-Demoulin et al. (2005) which has been employed in a range of empirical applications from modeling equity risk to extreme spikes in electricity prices (Chavez-Demoulin and McGill, 2012; Herrera, 2013; Herrera and González, 2014). Here, the

one-dimensional approach is extended to include IV as an exogenous variable. A novel bivariate model is also developed to incorporate the intensity of the occurrence of extreme movements in IV. This approach will uncover potential bi-directional linkages between extreme movements in IV and extreme losses. Results from this analysis will reveal whether using IV itself, or the intensity of its extreme movements, lead to more precise prediction of the intensity and size of extreme equity market losses.

2.1. Conditional mean and volatility models

The conditional mean of the equity market returns is specified as an Auto Regressive Moving Average (ARMA) process

$$r_t = \mu + \sum_{i=1}^{m} a_i r_{t-i} + \sum_{j=1}^{n} b_j \varepsilon_{t-j} + \varepsilon_t. \tag{2.1}$$

Where r_t denotes the return on a stock market index at time t, μ a constant, a_i and b_j describe the autoregressive and moving average coefficients, respectively and ε_t denotes the residual term. The residuals are defined by

$$\varepsilon_t = \eta_t \sqrt{h_t}, \qquad \eta_t \sim iid(0,1), \tag{2.2}$$

where η_t is the standardized residual and h_t is the conditional variance. The GARCH specifications considered for the conditional variances which include IV as an exogenous variable are

$$\begin{aligned}
\text{GARCH(1,1)}: h_t &= \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1} + \gamma IV_{t-1} && (2.3)\\
\text{GJR-GARCH(1,1)}: h_t &= \omega + \alpha\varepsilon_{t-1}^2 + \delta \max(0, -\varepsilon_{t-1})^2 + \beta h_{t-1} + \gamma IV_{t-1} && (2.4)\\
\text{EGARCH(1,1)}: \ln h_t &= \omega + \alpha\varepsilon_{t-1} + \delta(|\varepsilon_{t-1}| - E|\varepsilon_{t-1}|) + \beta \ln h_{t-1} + \gamma \ln IV_{t-1}. && (2.5)
\end{aligned}$$

The GARCH model in Eq. (2.3) corresponds to the standard model of Bollerslev (1986), with $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$ and $\gamma \geq 0$ so that the conditional variance $h_t > 0$. The model is stationary if $|\alpha + \beta| < 1$ is ensured. The GJR-GARCH specification in Eq. (2.4) allows the conditional variance to respond asymmetrically to the sign of past returns by means of the parameter δ. Sufficient conditions for $h_t > 0$ are $\omega > 0$, $\alpha + \delta \geq 0$, $\beta \geq 0$ and $\gamma \geq 0$. Finally, the EGARCH specification in Eq. (2.5), allows for asymmetries in volatility if $\delta \neq 0$ while leverage exists if $\delta < 0$ and $\alpha < \delta < -\alpha$. To be consistent with the specification of the conditional variance in Eq. (2.5), IV indices are included in logarithmic form. These three conditional volatility specifications are

estimated assuming a Skew Student-t distribution. [1]

The BEGE model of Bekaert et al. (2015) describes the innovations in returns,

$$\begin{aligned}
&\varepsilon_t = \sigma_p \omega_{p,t} - \sigma_n \omega_{n,t}, \text{ where} \\
&\omega_{p,t} \sim \widetilde{\Gamma}(p_t, 1), \text{ and} \\
&\omega_{n,t} \sim \widetilde{\Gamma}(n_t, 1)
\end{aligned} \tag{2.6}$$

as a linear combination of two component shocks, where $\widetilde{\Gamma}(k,\theta)$ is a centred gamma distribution with shape and scale parameters, k and θ respectively. The two gamma distributions are assumed to have a constant scale, but time-varying shape parameters, p_t and n_t for the good and bad environments respectively. The shape parameters evolve according to a GJR-GARCH like structure

$$\begin{aligned}
p_t &= p_0 + \rho_p p_{t-1} + \frac{\sigma_p^+}{2\sigma_p^2}\varepsilon_t^2 \mathbf{I}_{\varepsilon_{t-1}\geq 0} + \frac{\sigma_p^-}{2\sigma_p^2}(1 - \mathbf{I}_{\varepsilon_{t-1}\geq 0}), \\
n_t &= n_0 + \rho_n n_{t-1} + \frac{\sigma_n^+}{2\sigma_n^2}\varepsilon_t^2 \mathbf{I}_{\varepsilon_{t-1}\geq 0} + \frac{\sigma_n^-}{2\sigma_n^2}(1 - \mathbf{I}_{\varepsilon_{t-1}\geq 0}).
\end{aligned} \tag{2.7}$$

A version of this model that includes lagged IV as an exogenous variable, with a common coefficient in both positive and negative components (denoted below as BEGE+IV) is also estimated.

2.2. Conditional intensity models

Marked point processes (MPP) are stochastic processes that couple the temporally clustered arrival times of extreme events, with a set of random variables, the so-called marks associated with each event. In EVT, the interest lies in the intensity of extreme event occurrences as well as the distribution of the exceedences over a pre-determined large or extreme threshold. This paper develops two approaches for investigating the role of IV in explaining the intensity and size of extreme loss events. In doing so, the nature of the extreme loss-IV relationship will be revealed.

2.2.1. Univariate Hawkes-POT model

The first point process approach is based on a univariate MPP, specifically the Hawkes-POT model introduced by Chavez-Demoulin et al. (2005) and applied by Chavez-Demoulin and McGill (2012). The Hawkes-POT model is generalised here by using the IV index as a covariate in the conditional intensity process for extreme loss events.

[1] In a preliminary version of the paper both a conditional Normal, and symmetric student t distribution were also considered. However assuming a skewed student t conditional distribution provides a superior fit to the data. Here the skewness is incorporated into the t-distribution using the method of Fernandez and Steel (1998).

In this context, let $\{(X_t, Y_t)\}_{t\geq 1}$ be a vector of random variables that represent the log-returns of a stock market index and the associated IV derived from options on that index. For ease of subsequent notation, assume returns are multiplied by -1. To determine the conditional intensity of extreme losses, return events whose size exceeds a pre-defined high threshold $u > 0$ are the focus. This will define a finite subset of observations $\{(t_i, w_i, z_i)\}_{i\geq 1}$, where $t_i \in \mathbb{R}$ corresponds to occurrence times, $w_i \in \mathbb{R}_+$ the magnitude of exceedences (the marks), and $z_i \in \mathbb{R}_+$ a covariate based on the IV index, with $w_i := X_{t_i} - u$, and $z_i := Y_{t_i}$. A general MPP $N(t)$ is proposed satisfying the usual conditions of right-continuity $N(t) := N(0,t] = \sum_{i\geq 1} \mathbb{1}\{t_i \leq t,\, w_i = w,\, z_i = z\}$ with past history or natural filtration $\mathscr{H}_t = \{(t_i, w_i, z_i)\ \forall i : t_i < t\}$ that includes times, marks and the covariates. According to the standard definition of an MPP, it may be characterized by means of its conditional intensity function

$$\lambda(t, w \mid \mathscr{H}_t) = \lambda_g(t \mid \mathscr{H}_t)\, g(w \mid \mathscr{H}_t, t), \tag{2.8}$$

which, broadly speaking, describes the probability of observing a new event in the next instant of time conditional on the history of the process.

There are two components to the intensity of the MPP, a ground process $N^g(t) := \sum_{i\geq 1} \mathbb{1}\{t_i \leq t\}$ with conditional intensity $\lambda_g(t \mid \mathscr{H}_t)$ characterizing the rate of the extreme events over time, and the process for the marks, whose density function $g(w \mid \mathscr{H}_t, t)$ is conditional on the history of the process and time t. Observe that the covariate z_i does not directly enter into the definition of the conditional intensity in Eq. (2.8) even though it appears to be another mark in addition to w_i contained in the available information set, $\mathscr{H}_t$. Instead, the covariate z_i provides extra information to explain the behaviour of the process without being directly involved in the determination of likelihood in this stochastic process.

The conditional intensity $\lambda_g(t \mid \mathscr{H}_t)$ is characterized by the branching structure of a Hawkes process with an exponential decay function

$$\lambda_g(t \mid \mathscr{H}_t) = \nu + \vartheta \sum_{i:t_i<t} e^{\psi w_i + \rho z_i} \phi e^{-\phi(t-t_i)}, \tag{2.9}$$

where $\nu \geq 0$ is the intensity of exogenous events independent of the internal history $\mathscr{H}_t$. The branching coefficient $\vartheta \geq 0$ describes the frequency with which new extreme events arrive. The parameters $\psi \in \mathbb{R}$ and $\rho \in \mathbb{R}$ determine the contribution of the mark w_i and covariate z_i to the conditional intensity of the ground process, and $\phi > 0$ is a decay parameter. The exponential functions inside the sum define the impact function $f(w,z) = e^{\psi w + \rho z}$, and the kernel decay function $h(t - t_i) = \phi e^{-\phi(t-t_i)}$ which controls how offspring are generated by first order extreme events representing the main source of clustering in the model. This process is described as self-exciting as

the occurrence times and marks of past extreme events may make the occurrence of future extreme events more probable through the dependance on the history, $\mathscr{H}_t$.

To estimate risk measures such as VaR, an assumption regarding the probability distribution function of the most extreme return events, w_i conditional on the event that X_{t_i} exceeds the threshold $u > 0$ must be made. Motivated by the Pickands–Balkema–de Haan's theorem, [2] the extreme losses are assumed to follow a conditional Generalized Pareto Distribution (GPD) with a density function given by

$$g(w \mid \mathscr{H}_t, t) = \begin{cases} \frac{1}{\kappa(w|\mathscr{H}_t,t)} \left(1 + \xi \frac{w}{\kappa(w|\mathscr{H}_t,t)}\right)^{-1/\xi - 1} & , \quad \xi \neq 0 \\ \frac{1}{\kappa(w|\mathscr{H}_t,t)} \exp\left(-\frac{w}{\kappa(w|\mathscr{H}_t,t)}\right) & , \quad \xi = 0, \end{cases} , \tag{2.10}$$

where ξ is the shape parameter and $\kappa(w \mid \mathscr{H}_t, t)$ is a scale parameter specified as a self-exciting function of the arrival times of new extreme events and their sizes

$$\kappa(w \mid \mathscr{H}_t, t) = \kappa_0 + \kappa_1 \sum_{i:t_i<t} e^{\psi w_i + \rho z_i} \phi e^{-\phi(t-t_i)}.$$

Under this specification, $\kappa_0 \geq 0$ represents the baseline level for the scale, while $\kappa_1 \geq 0$ is an impact parameter related to the influence of new extreme event arrivals. The shape parameter is assumed to be constant through time due to the sparsity of events in the tail of the distribution which makes estimation of time-varying scale challenging (as evident in Chavez-Demoulin et al., 2005; Santos and Alves, 2012; Herrera, 2013).

The log-likelihood for the univariate Hawkes-POT model given a set of events $\{(t_i, w_i, z_i)\}_{i=1}^{N(T)}$ observed in the space $(0, T] \times [u, \infty)$ is obtained combining the conditional intensity of Eq.(2.8) and the density of the marks from Eq. (2.10) as follows

$$\begin{aligned} \ell &= \sum_{i=1}^{N(T)} \ln \lambda_g(t_i \mid \mathscr{H}_{t_i}) - \int_0^T \lambda_g(s \mid \mathscr{H}_s)\, ds + \sum_{i=1}^{N(T)} \ln g(w_i \mid \mathscr{H}_{t_i}, t_i) \qquad (2.11) \\ &= \sum_{i=1}^{N(T)} \ln \left(\nu + \vartheta \sum_{i:t_i<T} e^{\psi w_i + \rho z_i} \phi e^{-\phi(T-t_i)} \right) - \left\{ \nu T + \vartheta \sum_{i:t_i<T} e^{\psi w_i + \rho z_i} \left(1 - e^{-\phi(T-t_i)}\right) \right\} \\ &\quad - \left[(1/\xi + 1) \sum_{i=1}^{N(T)} \left\{ \ln \kappa(w_i \mid \mathscr{H}_{t_i}, t_i) + \ln\left(1 + \xi w_i / \kappa(w_i \mid \mathscr{H}_{t_i}, t_i)\right) \right\} \right] \end{aligned}$$

[2] See Pickands (1975) and Balkema and De Haan (1974).

assuming for ease of the exposition that $\xi \neq 0$. The resulting estimates are consistent, asymptotically normal and efficient, with standard errors obtained via the Fisher information matrix (Ogata, 1978).

2.2.2. Bivariate Hawkes-POT model

The novel bivariate approach proposed here moves beyond simply including IV as an exogenous covariate. Extreme increases in IV are treated as a second MPP and represent the second dimension in a bivariate model in addition to the extreme stock market losses. In this bivariate model, the marks influence the evolution of its respective ground process and vice versa, offering a framework to examine the impact of IV events on extreme stock market losses in terms of both the intensity and size of events.

The bivariate MPP is defined as a vector of point processes $\mathbf{N}(t) : \{N_1(t), N_2(t)\}$, where the first point process $N_1(t)$ is defined through the pairs $\left\{\left(t_i^1, w_i\right)\right\}_{i\geq 1}$; the subset of extreme events in the negative log-returns of the stock market occurring at time t_i^1 over a high threshold $u_1 > 0$, with $w_i := X_{t_i^1} - u_1$. Similarly, the second point process $N_2(t)$ is defined by the pairs of events $\left\{\left(t_i^2, z_i\right)\right\}_{i\geq 1}$ with $z_i := Y_{t_i^2} - u_2$, which also characterizes the subset of extreme events occurring in IV at time t_i^2 over a high threshold $u_2 > 0$. $\mathscr{H}_t = \left\{\left(t_i^1, w_i\right), \left(t_j^2, z_j\right)\ \forall i, j : t_i^1 < t \wedge t_j^2 < t\right\}$ denotes the combined history over all times and marks. This bivariate MPP includes a bivariate ground process $N_k^g(t) := \sum_{i\geq 1} \mathbb{1}\left\{t_i^k \leq t\right\}$ with conditional intensities

$$\begin{aligned}
\lambda_g^1(t \mid \mathscr{H}_t) &= \nu_1 + \vartheta_{11} \sum_{i:t_i^1<t} e^{\psi_1 w_i} \phi_1 e^{-\phi_1\left(t-t_i^1\right)} + \vartheta_{12} \sum_{i:t_i^2<t} e^{\rho_1 z_i} \phi_2 e^{-\phi_2\left(t-t_i^2\right)} \qquad (2.12)\\
\lambda_g^2(t \mid \mathscr{H}_t) &= \nu_2 + \vartheta_{21} \sum_{i:t_i^1<t} e^{\psi_2 w_i} \phi_1 e^{-\phi_1\left(t-t_i^1\right)} + \vartheta_{22} \sum_{i:t_i^2<t} e^{\rho_2 z_i} \phi_2 e^{-\phi_2\left(t-t_i^2\right)}
\end{aligned}$$

where $\nu_k \geq 0$ are the exogenous intensities, the branching coefficients $\vartheta_{jk} \geq 0$ describe the influence that dimension k will have on dimension j, the parameters $\psi_k \geq 0$ and $\rho_k \geq 0$ determine the contribution of the size of the extremes occurring at the returns and IV to the conditional intensity of the ground process, and $\phi_k > 0$ are again the decay parameters[3]. Thus, the impact functions $f_k(w) = e^{\psi_k w}$ and $f_k(z) = e^{\rho_k z}$, and the exponential decay kernel function $h_k\left(t - t_i^k\right) = \phi_k e^{-\phi_k\left(t-t_i^k\right)}$ account for mutual and cross excitation.

[3] Assuming the same rate of decay ϕ_1 and ϕ_2 in both dimensions, return and IV events, is common practice in such models see for instance Embrechts et al. (2011), Aït-Sahalia et al. (2015) and Aït-Sahalia and Hurd (2015). While in theory it is possible to have four different parameters, the model will suffer from identification problems. The approach we have taken here allowing ϕ_1 to differ from ϕ_2 is in fact a middle ground as other studies restrict all of the decay parameters to take one common value, Lee and Seo (2017).

A key feature of the proposed bivariate MPP is that it only includes a true mark for the point process of the stock market returns, with the distribution of the marks for the IV events always set to unity, $g(z \mid \mathscr{H}_t, t) = 1$ implying the conditional intensity for these events is

$$\lambda^2(t, z \mid \mathscr{H}_t) = \lambda_g^2(t \mid \mathscr{H}_t). \tag{2.13}$$

This assumption is invoked as the focus is on estimating measures of risk for the stock market returns given the behavior of IV at extreme levels (i.e., conditional intensity, occurrence times and size of extreme events in IV). To achieve this, it is not necessary to model the distribution of the extreme IV events thus reducing possible estimation error.

Similar to the univariate MPP, a generalized Pareto density for the stock market returns as in Eq.(2.10), is used again but with conditional scale parameter

$$\kappa(w \mid \mathscr{H}_t, t) = \kappa_0 + \kappa_1 \sum_{i:t_i^1<t} e^{\psi_1 w_i} \phi_1 e^{-\phi_1\left(t-t_i^1\right)} + \kappa_{12} \sum_{i:t_i^2<t} e^{\rho_1 z_i} \phi_2 e^{-\phi_2\left(t-t_i^2\right)}. \tag{2.14}$$

Under this specification $\kappa_{12} \geq 0$ is an impact parameter related to the influence of the arrival times and size of extreme events occurring in the IV index.

Define the occurrence of pairs of observations $\left\{\left(t_i^1, w_i\right)\right\}_{i=1}^{N_1(T)}$ and $\left\{\left(t_i^2, z_i\right)\right\}_{i=1}^{N_2(T)}$ in a set $(0,T] \times [u_1, \infty)$ and $(0,T] \times [u_2, \infty)$ respectively. The log-likelihood for this bivariate point process is obtained by linking the bivariate conditional intensity for the ground process in Eq.(2.12) and the density for the marks of the stock market returns Eq. (2.10) with scale parameter defined by in Eq.(2.14).

$$\begin{aligned}
\ell &= \sum_{k=1}^{2} \left\{ \sum_{i=1}^{N^k(T)} \ln \lambda_g^k\left(t_i^k \mid \mathscr{H}_{t_i^k}\right) - \int_0^T \lambda_g^k(s \mid \mathscr{H}_s)\, ds \right\} + \sum_{i=1}^{N^1(T)} \ln g\left(\mathscr{H}_{t_i^1}, t_i^1\right) \qquad (2.15) \\
&= \sum_{k=1}^{2} \sum_{i=1}^{N^k(T)} \ln \left(\nu_k + \vartheta_{k1} \sum_{j:t_j^1<t_i^k} e^{\psi_k w_j} \phi_1 e^{-\phi_1\left(t_i^k - t_j^1\right)} + \vartheta_{k2} \sum_{j:t_j^2<t_i^k} e^{\rho_k z_j} \phi_2 e^{-\phi_2\left(t_i^k - t_j^2\right)} \right) \\
&\quad - \left\{ (\nu_1+\nu_2)T + \sum_{k=1}^{2} \left\{ \vartheta_{k1} \sum_{j:t_j^1<T} e^{\psi_k w_j} \left(1 - e^{-\phi_k\left(T-t_j^1\right)}\right) + \vartheta_{k2} \sum_{j:t_j^2<T} e^{\rho_k z_j} \left(1 - e^{-\phi_2\left(T-t_j^2\right)}\right) \right\} \right\} \\
&\quad - \left[(1/\xi + 1) \sum_{i=1}^{N^1(T)} \left\{ \ln \kappa\left(w_i \mid \mathscr{H}_{t_i^1}, t_i^1\right) + \ln\left(1 + \xi w_i / \kappa\left(\mathscr{H}_{t_i^1}, t_i^1\right)\right) \right\} \right],
\end{aligned}$$

assuming once again for ease of the exposition that $\xi \neq 0$.

Following Embrechts et al. (2011), the following preposition outlines a number of weak conditions which ensure the existence of a Hawkes-POT process with stationary increments and asymptotically stationary conditional ground intensity.

Preposition 1. *(Stationarity) The conditional ground intensities defined in Eqs. (2.9) and (2.12) are asymptotically stationary under the following stability conditions*

- ***Univariate model:*** *Define* $h(s) := \phi e^{-\phi(s)}$ *and* $f(w,z) := e^{\psi w + \rho z}$ *as the decay kernel and impact function, respectively. Then, given that the decay kernel function satisfies* $\int_0^\infty h(s)\,ds = 1$, *and the expectation of impact function exists* $\mathbb{E}[f(w,z)] = \mu_{wz}$, *the univariate model defined in Eq. (2.9) is asymptotically stationary, if and only if,*

$$0 < \vartheta \mu_{wz} < 1.$$

- ***Bivariate model:*** *Define* $h_k(s) := \phi_k e^{-\phi_k(s)}$ *as the decay function satisfying* $\int_0^\infty h_k(s)\,ds = 1$ *for* $k = 1,2$, *and* $f_k(w) := e^{\psi_k w}$ *and* $f_k(z) := e^{\rho_k z}$ *as the impact functions of the marks and covariates with expectations given by* $\mathbb{E}[f_k(w)] = \mu_w^k$ *and* $\mathbb{E}[f_k(z)] = \mu_z^k$, *respectively. In addition, denoting* $M := \left\{\left(\mu_w^k, \mu_z^k\right) : k \in \{1,2\}\right\}$ *and* $Q := \left\{\vartheta_{jk} : j,k \in \{1,2\}\right\}$ *as the* (2×2) *matrix representations of the expectations and branching coefficients. The bivariate model defined in Eq. (2.12) is asymptotically stationary, if and only if, the spectral radius of the matrix* $M \circ Q$ *is less than one, i.e.,*

$$Spr(M \circ Q) := \max\{|\varphi| : det(M \circ Q - \varphi \mathbf{1_2}) = 0\} < 1,$$

where $\mathbf{1_2}$ *is the* (2×2) *identity matrix,* φ *are the eigenvalues of* $M \circ Q$, *and* $\circ$ *denotes the Hadamard product.*

Proof. Given in Appendix A

3. Generating and evaluating forecasts of conditional risk measures

The accuracy of the forecasts of extreme events will be analysed in the context of conditional risk measures, namely VaR. VaR_α^t is the VaR computed at day $t-1$ for the negative log-return X_t as follows

$$1 - \alpha = P\left(X_t > VaR_\alpha^t \mid \mathscr{H}_t\right),$$

where the equality above assumes a continuous distribution for X_t. Most financial return series exhibit stochastic volatility, autocorrelation, and fat-tailed distributions limiting the direct estimation of VaR. For this reason, under the traditional benchmark approach, the first stage consists of filtering the returns series with an ARMA-GARCH process such that the residuals are closer to iid. Given the assumed dynamics for the conditional mean of returns in Eq. (2.1), and the conditional volatility proposed in Eq. (2.2) the following model for the returns is obtained

$$X_t = \mu + \sum_{i=1}^{m} a_i X_{t-i} + \sum_{j=1}^{n} b_j \varepsilon_{t-j} + \varepsilon_t, \tag{3.1}$$

where $\varepsilon_t = \eta_t \sqrt{h_t}$ and h_t is the stochastic conditional variance, $h_t \in \mathscr{H}_t$. The autoregressive specifications for the conditional variances including the GARCH, GJR-GARCH and EGARCH are shown in Eqs. (2.3), (2.4) and (2.5) respectively.

In the second stage, the corresponding VaR at the α confidence level of the assumed distribution of the residuals η_t , i.e., $VaR_\alpha(\eta_t) : \inf\{x \in \mathbb{R} : P(\eta_t > x) \leq 1-\alpha\}$ is used to obtain estimates for the conditional VaR for the returns. Observe that η_t are iid, and therefore $VaR_\alpha(\eta_t) = VaR_\alpha(\eta_{t-1}) = \cdots = VaR_\alpha(\eta_{t-j}) =: VaR_\alpha(\eta)$, implying that Eq. (3.1) can be rewritten as follows

$$VaR_\alpha^t = \mu_t + VaR_\alpha(\eta)\sigma_t,$$

where $\mu_t = \mu + \sum_{i=1}^{m} a_i X_{t-i} + \sum_{j=1}^{n} b_j \varepsilon_{t-j}$ and $\sigma_t = \sqrt{h_t}$ are the natural 1-step forecasts of the conditional mean and variance, respectively. Note that the history $\mathscr{H}_t$ in this type of model is generated in a discrete time framework, contrary to the filtration generated by the point process approach where time is continuous. Therefore all information relating to the stochastic process prior to (but not at) time t can be included.

VaR forecasts from the BEGE models are obtained by numerically inverting the BEGE cumulative distribution (used to numerically evaluate the probability distribution function and hence the likelihood) function at the required α confidence level, given forecasts of p_t and n_t. By doing so, this takes into account not only the conditional variance, but also the higher moments of the distribution when generating the VaR forecast.

The two Hawkes-POT models (univariate and multivariate) described in Section 2.2 can also be directly used to estimate VaR. The advantage of this approach is that it avoids the filtering of returns and the use of EVT. Observe that the conditional probability that the next daily return X_t will exceed the threshold $u > 0$ given that X_{t-1} has already exceeded this threshold is given by

$$\begin{aligned}\mathbb{P}(X_t > u \mid \mathscr{H}_t) &= 1 - \mathbb{P}\{N([t-1,t) = 0 \mid \mathscr{H}_t)\} \\ &= 1 - \exp\left(-\int_{t-1}^{t} \lambda_g(s \mid \mathscr{H}_s)\, ds\right), \\ &\approx \lambda_g(t \mid \mathscr{H}_t). \end{aligned} \tag{3.2}$$

On the other hand, the conditional probability of this event, of exceeding an even higher threshold $(u+x) > 0$ given that the high threshold $u > 0$ has been exceeded, is modeled using a generalized Pareto distribution.

$$\mathbb{P}(X_t - u > x \mid X_t > u, \mathscr{H}_t) = \overline{G}(x-u \mid \mathscr{H}_t, t), \tag{3.3}$$

where $\overline{G}(x-u \mid \mathscr{H}_t, t)$ corresponds to the survival function of the cumulative distribution function of Eq. (2.10). One can demonstrate that for Hawkes-POT models, the probability that the next daily return X_t will exceed the VaR at the α confidence level is a solution to the equation $\mathbb{P}(X_t > \text{VaR}_\alpha^t \mid \mathscr{H}_t) = 1-\alpha$, or alternatively,

$$\mathbb{P}\left(X_t > \text{VaR}_\alpha^t \mid \mathscr{H}_t\right) = \mathbb{P}(X_t > u \mid \mathscr{H}_t)\, \mathbb{P}\left(X_t - u > \text{VaR}_\alpha^t - u \mid X_t > u, \mathscr{H}_t\right). \tag{3.4}$$

Thus, given the conditional intensity for the ground process in Eq.(3.2) and the distribution for the marks in Eq. (3.3), a solution to Eq. (3.4) leads to a prediction of the VaR in the next instant at the α confidence level

$$\text{VaR}_\alpha^t = u + \frac{\kappa(w \mid \mathscr{H}_t)}{\xi}\left\{\left(\frac{\lambda_g(t \mid \mathscr{H}_t)}{1-\alpha}\right)^{\xi} - 1\right\}. \tag{3.5}$$

Depending on the approach, univariate or bivariate, the ground conditional intensity in Eq. (3.5) is replaced with either Eq. (2.9) or (2.12). The same occurs for the scale parameter.

To assess the accuracy of the competing approaches for predicting VaR, a range of statistical tests are employed. These are based on both long-standing methods along with very recent developments. For further details see Christoffersen (1998); Engle and Manganelli (2004); Ziggel et al. (2014). Let $\{I_t(\alpha)\}_{t=1}^n$ be a vector of ex-post indicator variables of VaR exceptions taking the value 1 if $X_t > VaR_\alpha^t$ and 0 if $X_t \leq VaR_\alpha^t$ at time t at the VaR coverage probability α. In addition, define the variable $Hit_t(\alpha) = I_t(\alpha) - \alpha$ as the de-meaned hits of exceptions.

The first test is the unconditional coverage test (LR_{uc}) introduced by Kupiec (1995) which is con-

cerned with whether the reported VaR exceptions occur more (or less) frequently than $\alpha \times 100\%$ of the time. The second test examines the independence of these exceptions (LR_{ind}) using a Markov test. The third test is the conditional coverage test (LR_{cc}), which is a combination of the previous two tests. The key point of this test is that an accurate VaR measure must exhibit both the independence and unconditional coverage properties. The next two tests are the regression based Dynamic Quantile tests introduced by Engle and Manganelli (2004), where the regressors are the lagged Hit_t in the Dynamic Quantile Hit (DQ_{hit}) test, whereas the Dynamic Quantile VaR (DQ_{VaR}) also includes past VaR estimates as an explanatory variable.

More recently Ziggel et al. (2014) proposed a new set of tests that, beside testing unconditional coverage and independence of exceptions, also test that exceptions are identically distributed. Another advantage of these new tests is that all critical values for these tests are distribution free and can be obtained utilizing Monte Carlo simulations, allowing for one- and two-tailed tests to be carried out. Under this framework, the null hypothesis of unconditional coverage test is satisfied if the expectation of VaR exceptions is equal on average to α, i.e., $H_0 : \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} I_t(\alpha)\right] = \alpha$. They propose the statistic:

$$MC_{uc} = \sum_{t=1}^{n} I_t(\alpha) + \varepsilon, \tag{3.6}$$

where ε is a continuous random variable with a small variance designed to help to break the link between the test statistics.

To test for iid VaR exceptions, Ziggel et al. (2014) utilizes the fact that waiting times between VaR exceptions should be geometrically distributed. In particular, they propose to test the null hypothesis $H_0 : \mathbb{E}[t_i - t_{i-1}] = \frac{1}{\alpha}$, by examining at the squared waiting times between VaR exceptions, which are better suited to detect exceptions which occur in clusters:

$$MC_{iid,m} = t_1^2 + (n - t_m)^2 + \sum_{i=2}^{m} (t_i - t_{i-1})^2 + \varepsilon, \tag{3.7}$$

where m is the number of observed VaR exceptions and $t_1, \ldots, t_m$ describe the occurrence times of VaR exceptions. Note, that the value of this statistic increases as the waiting times exhibit a greater degree of correlation and hence it is very useful for detecting clustering among the VaR exceptions.

The final test corresponds to a conditional coverage test whose specification is given by:

$$MC_{cc,m} = a \cdot f(MC_{uc}) + (1 - a) \cdot g\left(MC_{iid,m}\right), \; 0 \leq a \leq 1, \tag{3.8}$$

where $f(MC_{uc}) = \left|\frac{MC_{uc}/n - p}{p}\right|$ and $g\left(MC_{iid,m}\right) = \frac{MC_{iid,m} - \hat{r}}{\hat{r}} 1_{\left\{MC_{iid,m} \geq \hat{r}\right\}}$ measure the difference be-

tween the expected and observed proportions of VaR exceptions, and sum of squared waiting times, respectively. The parameter a is a weighting factor that can be chosen according to an individuals preference toward the importance of either the iid property or the correct unconditional coverage of the exceptions. In the subsequent empirical analysis, the importance of both properties are treated equal, and only results for $a = 0.5$ are presented. In the last term, $\hat{r}$ denotes an estimator of the expected value of the statistic $MC_{iid,m}$ under the null hypothesis. All critical values for these test statistics are obtained utilizing 10,000 Monte Carlo simulations of the finite sample null distribution. To ensure that the test statistics follow a continuous distribution, a continuous random variable with an arbitrarily small variance, $\varepsilon \sim N(0, 1e^{-6})$ is used in all applications[4]. For further details on the last three tests see Ziggel et al. (2014).

Backtesting considers whether each individual model produces VaR forecasts that are adequate in their own right and satisfy the coverage and independence properties. While this is important, these tests do not allow conclusions to be drawn on which model produces the most accurate VaR forecast. To directly measure forecast accuracy, the asymmetric quantile loss function

$$\ell\left(X_t, VaR_\alpha^t\right) = \left(I_t\left(\alpha\right) - \alpha\right)\left(X_t - VaR_\alpha^t\right) \tag{3.9}$$

proposed by González-Rivera et al. (2004) is used. $I_t\left(\alpha\right)$ is again the indicator function taking the value 1 when an exception occurs at α significance and 0 otherwise. The motivation behind this loss function is very intuitive in the context of risk management since VaR exceptions are penalised more heavily. Such a loss function would underly a fairly broad class of economic applications involving capital allocation in response to risk forecasts.

Given the quantile loss function in Eq. 3.9, significant differences in VaR forecast performance will be assessed using the Model Confidence Set (MCS) introduced by Hansen et al. (2011). The MCS approach avoids the specification of a benchmark model, and starts with a full set of candidate models $\mathscr{M}_0 = \{1, ..., m_0\}$. All loss differentials, $d_{ij,t}$, using Eq. 3.9, between models i and j are computed and the null hypothesis, $H_0 : \mathrm{E}(d_{ij,t}) = 0$ is tested for each pair. If H_0 is rejected at the significance level α_M, the worst performing model is removed and the process continues until non-rejection occurs with the set of surviving models being the MCS, $\widehat{\mathscr{M}}^*_{\alpha_M}$. If a fixed significance level α_M is used at each step, $\widehat{\mathscr{M}}^*_{\alpha_M}$ contains the best model from $\mathscr{M}_0$ with $(1 - \alpha_M)$ confidence. The null hypothesis is tested by means of the range statistic for combining individual t-statistics from the pairwise comparison of forecasts. An estimate of the asymptotic variance of the pairwise loss differentials is obtained from a bootstrap procedure described in Hansen et al. (2003). Re-

[4] According to Ziggel et al. (2014) the finite sample accuracy of the test statistics are not greatly affected by the choice of a continuous probability distribution function for ε, provided that its variance is small.

ported p-values are corrected to ensure consistency through the iterative testing framework. See Hansen et al. (2003) and Hansen et al. (2011) for more detail. In all subsequent empirical results, a level of 95% confidence will be used in the MCS analysis.

The estimation of VaR for horizons longer than one day is an important issue in determining financial risk. However, the extension of the Hawkes-POT model from a single prediction period to a longer horizon is not a trivial exercise, due to the dynamic specification for the intensity based on a stochastic counting process. As a final measure of the performance of the bivariate point process models, a simple attempt is made to obtain multi-period VaR estimates and examine whether they satisfy the standard tests discussed earlier. This is achieved by scaling the one period VaR by a factor $h^{\hat{\xi}_u}$

$$VaR_{\alpha}^{t+h} \approx h^{\hat{\xi}_u} VaR_{\alpha}^{t}$$

where h is the horizon time and $\hat{\xi}_u$ is the unconditional shape parameter obtained from the raw log-returns. The approach used here is based on EVT suggesting that the estimation of long-term VaR is actually possible for fat tailed distributions (see Danielsson and De Vries, 2000; Cotter, 2007, for empirical applications of this approach). The major advantage of this simple approach is that besides the estimation of the unconditional shape parameter $\hat{\xi}_u$, there is no need to re-estimate any additional parameters. Although VaR accuracy for multi-periods is complicated by the fact that the VaR exceptions are intrinsically autocorrelated, the unconditional coverage test (LR_{uc}) is applied over 5 and 10 day horizons, accounting for autocorrelation. Given the overlapping nature of the multi-period forecasts none of the more complex tests are undertaken.

3.1. Data

The data consists of daily returns for the S&P 500, Nasdaq, DAX 30, Dow Jones and Nikkei stock market indices, and their respective IV indices, VIX, VXN, VDAX, VXD, and VXJ. As the focus is on extreme increases in IV, events will be defined on daily log-changes in IV, $\triangle IV_t = \ln(IV_t/IV_{t-1})$ for each market. All data series used here are obtained from Bloomberg. For each pair of stock market index and IV, the longest sample of data available is collected (S&P 500: 02/01/90, DAX: 02/01/92, Dow Jones: 02/1/98, Nikkei: 05/01/98, Nasdaq: 01/01/00), with all series ending 31 December, 2013. The period ending 30 December 2011 is used for estimation while 2012-2013 is used for backtesting.

The VIX index was the first widely published IV index upon which trading was developed. IV for other U.S. indices (VXN and VXD) and both the European (VDAX) and Japanese markets (VXJ) all follow the same principle as the VIX. The VIX index was developed by the Chicago Board of Options Exchange from S&P 500 index options to be a general measure of the market's

Table 1: Descriptive statistics for the daily stock market returns and IV log-changes.

	S&P 500	VIX	DAX 30	VDAX	Dow Jones	VXD	Nikkei	VXJ	Nasdaq	VXN
mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
sd	0.01	0.03	0.01	0.05	0.01	0.06	0.02	0.06	0.02	0.05
min	-0.09	-0.15	-0.09	-0.27	-0.08	-0.33	-0.12	-0.42	-0.10	-0.31
max	0.11	0.22	0.11	0.31	0.11	0.53	0.13	0.58	0.11	0.36
skewness	-0.24	0.65	-0.10	0.68	-0.08	0.64	-0.32	1.45	-0.07	0.57
kurtosis	11.65	7.31	7.40	6.81	9.97	7.05	9.11	16.08	7.41	6.74
Ljung-Box	42.26*	109.82*	24.51*	37.77*	39.74*	59.09*	9.43***	31.01*	18.71*	31.41*
Jarque Bera	18916.57*	5118.41*	4292.14*	3637.99*	7639.51*	2837.02*	5806.37*	27554.81*	2632.06	2070.05*
ADF test	-18.20*	-19.81*	-16.94*	-17.83*	-15.31*	-16.49*	-15.38*	-15.97*	-14.54	-16.80*
Start of sample period	02/01/99		02/01/92		02/01/98		05/01/98		01/01/00	
Extreme events	606		533		379		370		326	
Comovements	340		308		201		197		160	

Notes: The Ljung-Box statistics are significant for a lag of 5 trading days. *, **, *** represent significance at 1%, 5% and 10% levels, respectively. All the samples end in December 31, 2013.

estimate of average S&P 500 volatility over the subsequent 22 trading days. It is derived from out-of-the-money put and call options that have maturities close to the fixed target of 22 trading days. For technical details relating to the construction of the VIX index, see CBOE (2003).

Descriptive statistics for each series are given in Table 1. It is clear that for all markets, the sample standard deviation of changes in IV are much larger than the corresponding equity index returns. All series exhibit high levels of kurtosis, stock market returns are negatively skewed and changes in IV are positively skewed. None of the series analysed are normally distributed based on the Jarque-Bera statistic. The Ljung-Box statistics reject the null of no autocorrelation at a lag of 5 trading days for all series. Augmented Dickey–Fuller (ADF) tests for the presence of unit roots show that all time series are stationary at 1% significance level.

Extreme movements in stock market returns or changes in IV are events for which the probability is small. An extreme event is defined as one that belongs to the 10% of the most negative returns in stock markets, or to the 10% of the most positive log-changes in IV. Table 1 reports the number of extreme events occurring in stock market returns and IV indices, independently and simultaneously. Of these, between 49% and 57% of the extreme events are joint events where extreme movements in equity returns and IV occur simultaneously, reflected in the comovements reported in the bottom row.

Fig. 3.1 gives an overview of the comovement at extreme levels in stock market returns and IV. The plots in the left column show the market returns and IV indices, with bars under the IV indicating the occurrence of the most negative extreme returns given the occurrence of a positive extreme change in IV. In a similar fashion to volatility itself, these extreme events tend to cluster

through time. The centre column of plots shows the relationship between changes in IV and stock market returns. There is a clear negative relationship between the two series reflecting the commonly observed asymmetry in the equity return-volatility relationship. The occurrence of extreme events, negative market returns and positive changes in IV, which are of central interest here, are represented by the black dots.

An alternative approach for measuring extremal comovement is the extremogram introduced by Davis et al. (2009). This is a flexible conditional measure of extremal serial dependence making it particularly well suited for financial applications. The plots in the right column in Fig. 3.1 show the sample extremograms for the 10% of the most negative stock returns conditional on the 10% of the most positive log-changes on the IV indices at different lags. The interpretation of the extremogram is similar to the correlogram. Given that the IV index has experienced an extreme positive change at time t, the probability of obtaining a negative extreme shock in stock market returns at time $t+k$ is reflected by the solid vertical lines in the sample extremogram for each lag k. The grey lines represent the .975 (upper) and .025 (lower) confidence intervals estimated using the stationary bootstrap procedure proposed by Davis et al. (2012), while the dashed line corresponds to this conditional probability under the assumption that extreme events in both markets are occurring independently (for more details on the estimation refer to Davis et al., 2012).[5] Observe that the speed of decay of the sample extremograms for all five markets is extremely slow given that the dependence of extreme movements in returns on IV shocks is significant out to about 10 lags in most cases.

4. Empirical results

In-sample estimation results are discussed in Section 4.1. As estimation results are not the main focus of the paper, a full detailed set of parameter estimates are relegated to the Internet Appendix with a discussion of the central points outlined here. Comparisons of forecast performance in terms of risk prediction are presented in Section 4.2.

4.1. Estimation results

Estimation results discussed in this section are based on data up to 30 December 2011. Estimation results for the various GARCH and BEGE specifications are reported in Tables IA.1 and IA.2. Results for models using a skew t-distribution are reported, because assuming either a conditional

[5] 2000 pseudo-series are generated for the estimation of the extremograms utilizing a stationary bootstrap with resampling based on block sizes from a geometric distribution with a mean of 200.

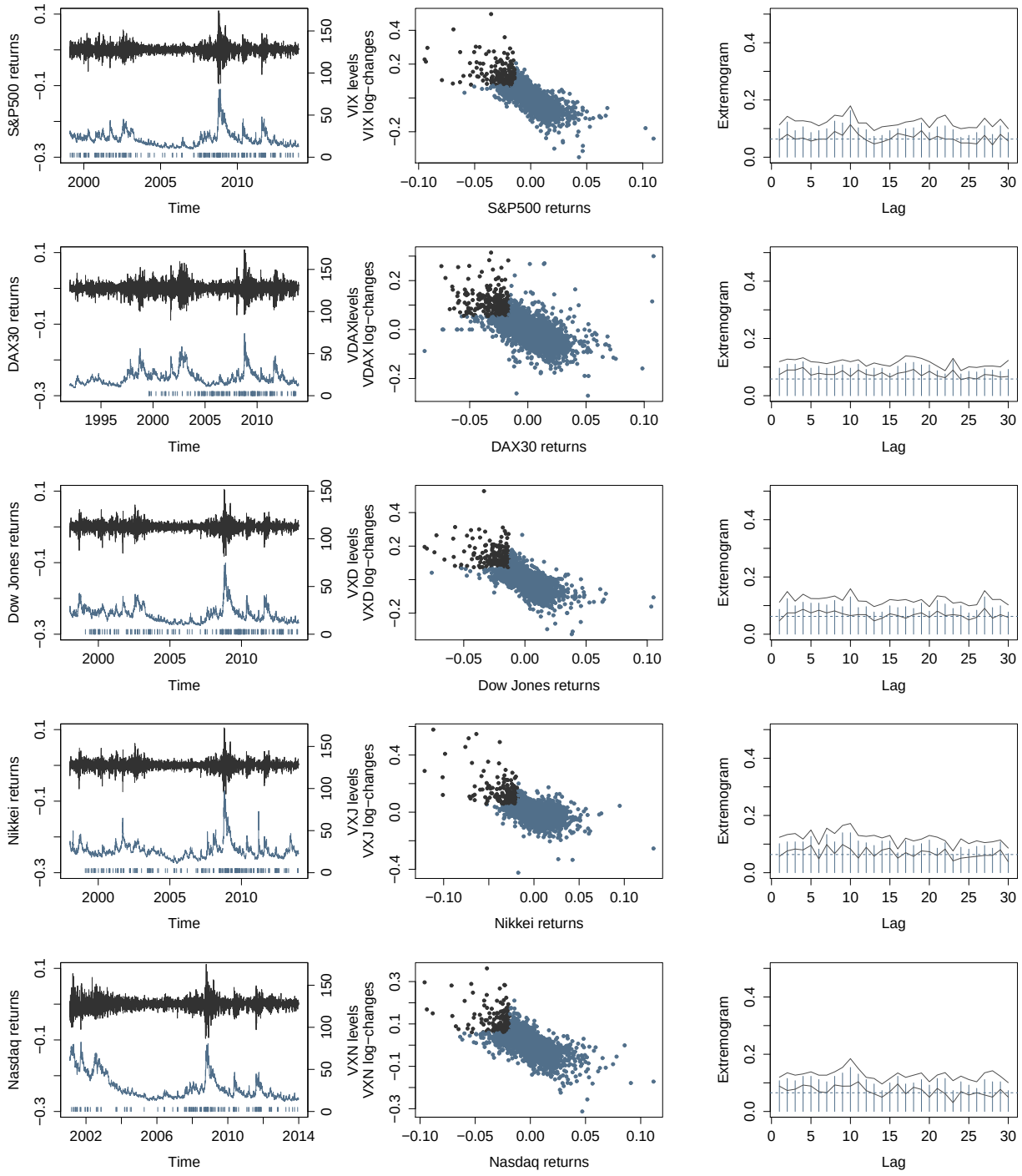

Figure 3.1: Left column: Extreme negative returns (grey color) and IV log-changes (blue color) to display the asymmetric association between them. Middle column: Scatter plot of IV log-changes and stock market returns. The 10% of the most extreme negative (positive) stock market returns (IV log-changes) are displayed in grey color. Right column: Sample extremograms for the 10% of the most negative stock returns conditional to the 10% of the most positive log-changes on the IV indices at different lags. Grey lines represent the .975 (upper) and .025 (lower) confidence interval estimated using a stationary bootstrap procedure proposed by Davis et al. (2012), while the dashed line (blue color) corresponds to this conditional probability under the assumption that extreme events between both markets occur independently.

normal or symmetric student t- distribution leads to inferior results and hence they are not reported here.

Estimates of the GARCH coefficients reveal a number of common patterns. For models that do not include IV as an exogenous regressor, estimates of the β coefficient (in Eqs. 2.3, 2.4 and 2.5) are in excess of 0.9 indicating a strong degree of volatility persistence. When IV is included, γ in the three models is found to be significant and the presence of IV helps explain a degree of the persistence in many of the cases with the estimates of β falling. As is to be expected, estimates of the asymmetry coefficient δ in both the GJR-GARCH and EGARCH (Eqs. 2.4 and 2.5) models are significant. Conditionally, returns are found to exhibit relatively heavy tails with estimates of the ν falling between 7 and 15. In all cases, estimates of the skew parameter ψ are significantly less than one indicating that returns are conditionally negatively skewed, supporting the choice of the skewed-t distribution. Of the competing models, EGARCH including IV offers the best model fit for all markets with the fit of the BEGE models close to both the GJR-GARCH and EGARCH models. While both positive and negative components of volatility int he BEGE model are found to be persistent, the negative component exhibits less persistence than the corresponding positive component ($\rho_n < \rho_p$), a result consistent with the findings of Bekaert et al. (2015). In three of the markets, the impact of IV on the shape parameters, p_t and n_t are significant and positive with the final two still positive though not significant.

Three versions of the univariate model in Eq. (2.9) are estimated. Model 1 is the full model with marks ($\psi > 0$) and IV ($\rho > 0$). Model 2 only includes marks ($\psi > 0$) restricting $\rho = 0$. Model 3 includes neither marks nor covariates and restricts $\psi = 0$ and $\rho = 0$. Again, the full set of estimation results for the three univariate models are reported in the Internet Appendix in Table IA.3. In all cases, the unrestricted Model 1 offers the best overall fit. Estimates for ψ are significant in all instances, reflecting the importance of the size of past marks for future intensity. On the other hand, estimates of ρ are strongly significant only in the S&P500 and Nikkei markets meaning that the level of IV is only important for explaining the intensity of extreme events in these two markets. While ρ is marginally significant for the DAX, it is insignificant for the remaining markets.

Similar to the univariate case, four versions of the bivariate model are estimated. The ground intensities under Model 1 are generated by the full unrestricted model in Eq. (2.12) and contain the past times and marks of both extreme return and IV events, with $\psi_1, \psi_2, \rho_1, \rho_2 > 0$, with the scale of the return marks specified in Eq. (2.14). Model 2 also includes the past times and marks of both extreme return and IV events, with the restriction that $\psi_1 = \psi_2$ and $\rho_1 = \rho_2$, with the scale only driven by the arrival times and size of the past return events ($\kappa_{12} = 0$). Model 3 contains the times and marks of return events ($\psi_1, \psi_2 > 0$) but only the times of past IV events (i.e., $\rho_1 = \rho_2 = 0$)

with the scale only driven by the size of the past return events. The ground intensities under Model 4 are restricted to contain the times of past return and IV events, $\psi_1 = \psi_2 = 0$ and $\rho_1 = \rho_2 = 0$ with the scale of the marks being driven by the timing of past IV events, ($\psi_1, \psi_2 > 0$ and $\rho_1 = \rho_2 = 0$) and the dynamic introduced by the arrival times of the extreme events in IV ($\kappa_{12} > 0$).

Table IA.4 in the Internet Appendix reports the full estimation results for all four bivariate models. In all markets, Model 3 is found to provide the best fit to the data, where the ground intensities of extreme return (λ_g^1) and IV (λ_g^2) events are driven by the size of past return marks and the timing of past return and IV events, and the scale is driven by the size of past return marks. The impact of the timing of past IV events on the intensities is evident in the positive estimates of ϕ_2 which are significant in four of the five markets. The degree of self, or cross-excitation, is reflected in the combination of ϑ, ψ or ρ, and ϕ coefficients. Significant estimates of ϑ_{11}, ϕ_1 and ψ_1 for Model 3 reveal strong self-excitation in the return events with a similar pattern evident for IV events in terms of ϑ_{22} and ϕ_2. In terms of cross excitation the results are varied, estimates of ϕ_1 and ϕ_2 are nearly always significant with estimates of ϑ_{12} and ϑ_{21} being somewhat mixed. There appears to be bi-directional cross-excitation in the DAX and Nikkei markets, with excitation from returns to IV in both the S&P500 and Nasdaq markets.

4.2. Forecasting risk

In this section, results of the tests for VaR accuracy discussed in Section 3 are presented. These results are based on an out-of-sample backtesting period 2012-2013. Model estimation for forecasting purposes is initially based on the in-sample period ending 30 December 2011, and then on a recursive estimation window where the models are re-estimated every week moving through the 2012-2013 period.

Before moving to a formal analysis of VaR accuracy, Figs. 4.1 and 4.2 show VaR estimates and predictions at a significance level of 0.99, along with returns for the in- and out-of-sample (also with exceptions) periods respectively. Results are shown for the S&P 500 index for a selection of models across the different classes of models that provided the best in-sample fit, EGARCH+IV, BEGE+IV, univariate (M1) and bivariate Hawkes-POT (M3). Beginning with Fig. 4.1, it is clear that all the VaR estimates broadly follow the volatility of the overall market. Two observations emerge, the EGARCH+IV estimates appear to be somewhat more variable for much of the period and both Hawkes based VaR estimates adapt to a higher level during the peak of market volatility in 2009. The lower panels in Fig. 4.1 show the VaR estimates and associated returns during a number of important periods of crisis and heightened market volatility. It is evident that focusing in on these periods of interest highlights that the VaR estimates generated by both MPP models are less variable, certainly in comparison to those from the EGARCH+IV model. They do however

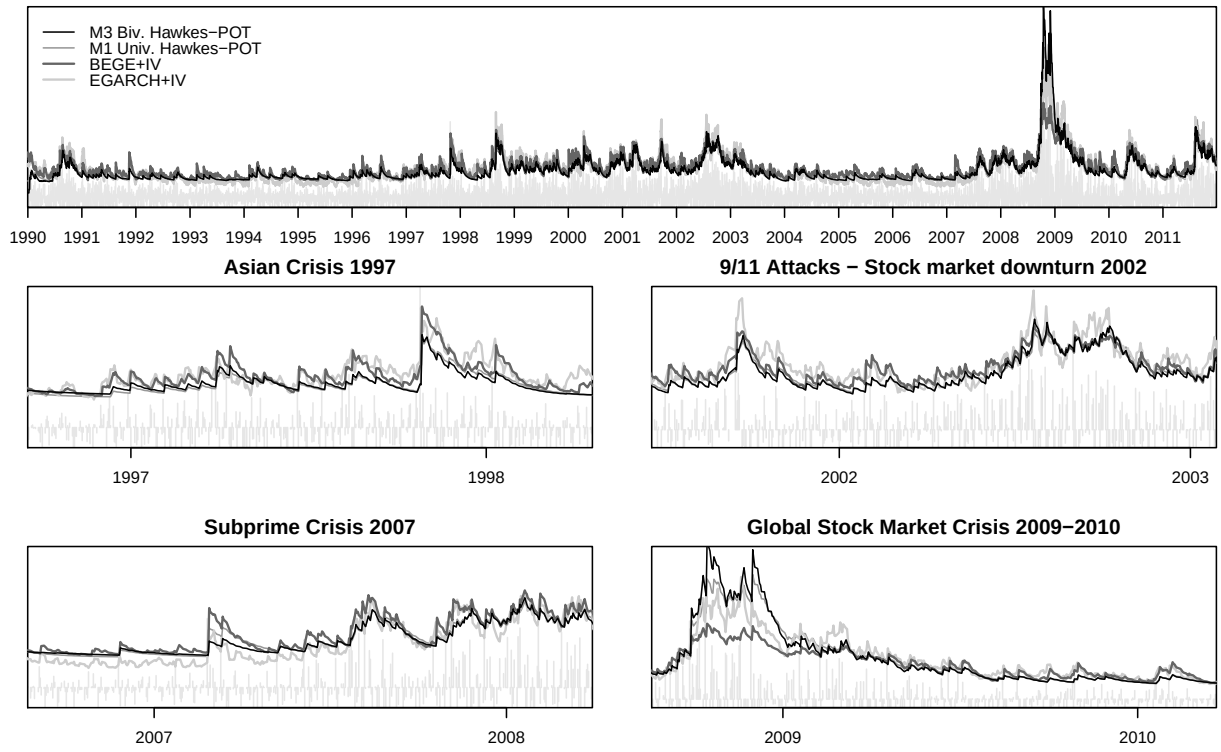

Figure 4.1: Plots of in-sample VaR estimates and returns (the negative of log returns are shown) on the S&P 500 index. VaR estimates are shown for four models across the different classes of models considered here, EGARCH + IV, BEGE+IV, univariate and bivariate Hawkes-POT. The top panel shows the full in-sample period, while the lower panels highlight various subperiods of interest.

adapt to noticeably higher levels during the peak of historically high volatility in 2009. Fig. 4.2 shows the corresponding VaR predictions during the backtesting period, 2012-2013. While all the forecasts vary with the overall volatility in returns, once again, the VaR forecasts from both self-exciting models are less variable than the EGARCH and BEGE equivalents. The exceptions from each model, BEGE+IV ($\times$), univariate Hawkes (+) and bivariate Hawkes ($\triangledown$) are also shown, with EGARCH+IV model producing no exceptions in this case. Visually speaking, there is no obvious clustering in the exceptions. It is clear the EGARCH+IV (and to some extent BEGE+IV) are not producing enough exceptions at $\alpha = 0.99$ and hence generating slightly conservative VaR predictions.

To begin the formal analysis, Table 2 reports results for the in-sample tests of VaR accuracy. To make the most efficient use of space here, and enhance the readability of the results, only a subset of the results are reported here. Results at $\alpha = 0.95, 0.99, 0.995$ for the GARCH, BEGE (those including IV) and Hawkes-POT (Univariate M1 and Bivariate M3) models with the best in-sample

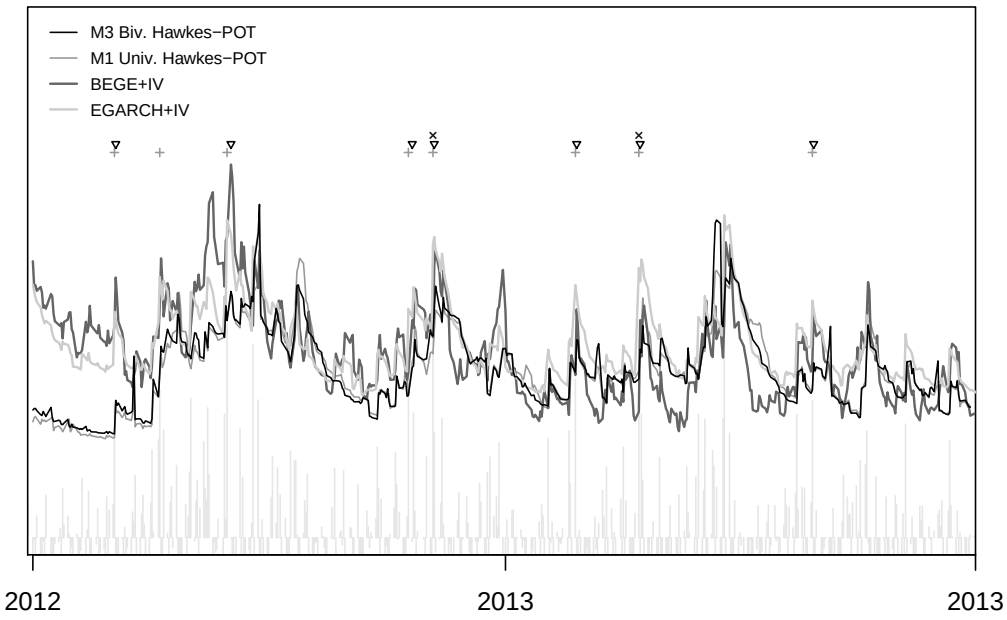

Figure 4.2: Plots of out-of-sample VaR prediction and returns (the negative of log returns are shown) on the S&P 500 index. VaR estimates are shown for four models across the different classes of models considered here, EGARCH + IV, BEGE+IV, univariate and bivariate Hawkes-POT. Exceptions from each model, BEGE+IV ($\times$), univariate Hawkes-POT (+) and bivariate Hawkes-POT ($\triangledown$) are also reported. The EGARCH+IV model produced no exceptions in this case.

performance are reported here. The full set of results, across all GARCH, BEGE and Hawkes-POT models are reported across Tables IA.5 (GARCH and BEGE) and IA.6 (Hawkes-POT) in the Internet Appendix. All in cases, the results are based on the full in-sample period ending in December 2011. Results are shown in the form of p-values which are shown in bold when a test is rejected at a significance level of $\alpha = 5\%$. Cells in the rows with the heading Exc. report the number of VaR exceptions in each case. Results in the Exc. rows show that in comparison to the GARCH models, the bivariate models tend to generate slightly fewer exceptions ($X_t > VaR_\alpha^t$) for most of the series. The bivariate models (M3, including IV) generate a similar number of rejections relative to the BEGE+IV model. In the vast majority of the cases, the tests are not rejected, indicating that the models accurately describe the in-sample behaviour of the extreme events in the context of VaR estimation. The majority of the rejections that do occur are found with the Nikkei, particularly at the highest (0.999) level of significance. Attention now turns to forecasting.

Table 3 reports results for tests of out-of-sample VaR forecast accuracy, with the results reported in the same format as Table 2. Again, only a subset of the results are presented for the GARCH,

Table 2: In-sample VaR accuracy test results.

Returns	Statistics	Volatility Models *(In-sample)*									Hawkes-POT Models *(In-sample)*					
		GJRGARCH+IV			EGARCH+IV			BEGE+IV			Univariate (M1)			Bivariate (M3)		
S&P500 -VIX	α- level	0.95	0.99	0.999	0.95	0.99	0.999	0.95	0.99	0.999	0.95	0.99	0.999	0.95	0.99	0.999
	Exc.	285	51	7	293	60	7	247	43	10	245	52	8	259	49	8
	LRuc	0.63	0.54	0.55	0.34	0.54	0.55	0.06	0.08	0.09	**0.04**	0.64	0.33	0.26	0.38	0.33
	LRind	0.71	0.33	0.89	0.10	0.25	0.89	0.18	0.41	0.85	0.50	0.32	0.88	0.40	0.35	0.88
	LRcc	0.84	0.52	0.83	0.16	0.43	0.83	0.07	0.15	0.23	0.10	0.55	0.61	0.37	0.44	0.61
	DQhit	0.72	0.33	0.89	0.11	0.25	0.89	0.19	0.41	0.85	0.51	0.32	0.88	0.41	0.35	0.88
	DQVaR	0.73	0.52	0.99	0.27	0.42	0.99	0.05	**0.02**	**0.00**	**0.03**	0.51	0.99	0.65	0.53	0.99
	MCuc	0.64	0.54	0.43	0.34	0.50	0.41	0.06	0.10	0.08	0.05	0.63	0.25	0.26	0.43	0.35
	MCind	0.93	0.32	0.62	0.12	0.33	0.32	0.68	**0.01**	0.14	**0.00**	**0.01**	0.30	**0.00**	0.05	0.30
	MCcc	0.15	0.63	0.74	0.25	0.66	0.66	0.66	**0.03**	0.29	**0.00**	**0.02**	0.62	**0.01**	0.10	0.60
DAX - VDAX	Exc.	284	58	6	297	53	8	207	31	3	253	60	15	246	61	5
	LRuc	0.73	0.76	0.86	0.26	0.72	0.33	**0.00**	**0.00**	0.32	0.99	0.20	**0.00**	0.65	0.15	0.98
	LRind	0.35	**0.02**	0.91	0.19	0.11	0.88	0.08	0.54	0.95	0.13	0.74	0.77	0.37	0.22	0.92
	LRcc	0.61	0.08	0.98	0.22	0.25	0.62	**0.00**	**0.01**	0.61	0.33	0.41	**0.00**	0.61	0.17	0.99
	DQhit	0.36	**0.03**	0.91	0.20	0.11	0.88	0.08	0.54	0.95	0.15	0.75	0.77	0.39	0.22	0.92
	DQVaR	0.61	0.06	0.99	0.44	0.20	0.99	**0.00**	0.64	0.81	0.33	0.63	0.92	0.59	0.31	0.99
	MCuc	0.72	0.74	0.73	0.26	0.76	0.38	**0.00**	**0.00**	0.24	0.95	0.20	**0.00**	0.68	0.14	0.79
	MCind	0.65	0.77	0.22	0.22	0.11	0.17	0.35	0.19	0.99	**0.01**	0.88	0.49	**0.00**	0.94	0.11
	MCcc	0.70	0.45	0.45	0.43	0.20	0.33	0.67	0.38	**0.01**	**0.01**	0.24	0.98	**0.00**	0.12	0.22
DJI - VXD	Exc.	180	37	5	187	37	5	157	29	6	161	34	10	164	36	12
	LRuc	0.76	0.76	0.46	0.40	0.76	0.46	0.13	0.28	0.23	0.26	0.86	**0.00**	0.38	0.87	**0.00**
	LRind	0.21	0.41	0.91	0.11	0.41	0.91	0.08	0.49	0.89	0.82	0.41	0.81	0.62	0.39	0.77
	LRcc	0.44	0.68	0.75	0.20	0.68	0.75	0.07	0.43	0.48	0.52	0.71	**0.02**	0.60	0.68	**0.00**
	DQhit	0.22	0.41	0.91	0.12	0.41	0.91	0.09	0.49	0.89	0.82	0.42	0.81	0.63	0.39	0.77
	DQVaR	0.39	0.64	**0.03**	0.20	0.62	**0.03**	0.22	0.05	0.24	0.81	0.56	0.97	0.83	0.53	0.96
	MCuc	0.73	0.79	0.33	0.38	0.73	0.51	0.15	0.30	0.18	0.25	0.87	**0.00**	0.37	0.78	**0.00**
	MCind	**0.01**	**0.00**	0.33	**0.01**	**0.00**	0.62	0.73	0.11	0.87	0.58	0.38	**0.02**	0.14	0.30	**0.02**
	MCcc	**0.02**	**0.00**	0.66	**0.02**	**0.00**	0.77	0.53	0.25	0.26	0.85	0.76	**0.04**	0.27	0.59	0.05
Nikkei - VXJ	Exc.	173	33	5	171	39	6	149	27	6	171	39	10	178	38	7
	LRuc	0.93	0.81	0.43	0.95	0.44	0.21	0.07	0.19	0.21	0.97	0.43	**0.00**	0.61	0.53	0.09
	LRind	0.43	0.33	**0.00**	0.23	0.08	**0.01**	0.28	0.21	**0.01**	0.87	0.08	**0.02**	0.55	0.44	**0.01**
	LRcc	0.73	0.60	**0.01**	0.49	0.16	**0.01**	0.10	0.19	**0.01**	0.99	0.16	**0.00**	0.74	0.61	**0.01**
	DQhit	0.44	0.33	**0.00**	0.25	0.08	**0.01**	0.29	0.21	**0.01**	0.87	0.08	**0.02**	0.56	0.44	**0.01**
	DQVaR	0.45	0.53	**0.02**	0.51	0.18	**0.02**	0.10	**0.01**	**0.00**	0.29	0.16	0.07	0.39	0.57	**0.03**
	MCuc	0.96	0.79	0.49	0.96	0.45	0.25	0.07	0.21	0.19	0.97	0.38	**0.00**	0.60	0.46	0.07
	MCind	0.46	0.50	0.42	0.20	0.50	0.53	0.42	0.40	**0.02**	**0.01**	0.87	0.76	**0.02**	0.91	0.67
	MCcc	0.92	1.00	0.84	0.40	0.98	0.94	0.82	0.79	**0.04**	**0.03**	0.25	0.48	**0.04**	0.18	0.68
Nasdaq - VXN	Exc.	141	27	2	133	25	3	121	17	2	121	26	5	123	28	5
	LRuc	0.74	0.93	0.64	0.71	0.63	0.88	0.15	**0.03**	0.64	0.17	0.80	0.22	0.23	0.89	0.22
	LRind	0.17	0.46	0.96	**0.03**	0.50	0.94	0.26	0.65	0.96	0.52	0.25	0.89	0.47	0.29	0.89
	LRcc	0.37	0.76	0.89	0.10	0.71	0.99	0.18	0.09	0.89	0.31	0.50	0.46	0.38	0.57	0.46
	DQhit	0.18	0.47	0.96	**0.04**	0.50	0.94	0.26	0.65	0.96	0.53	0.25	0.89	0.48	0.30	0.89
	DQVaR	0.21	**0.03**	0.98	**0.03**	0.10	0.56	0.53	0.90	0.95	0.59	0.39	0.99	0.60	0.44	0.99
	MCuc	0.74	0.97	0.77	0.71	0.66	0.98	0.15	**0.04**	0.81	0.16	0.78	0.22	0.24	0.80	0.14
	MCind	0.94	0.33	0.70	**0.02**	**0.03**	0.42	0.12	0.06	0.69	**0.00**	**0.00**	0.16	**0.00**	**0.00**	0.16
	MCcc	0.12	0.65	0.62	0.05	0.06	0.82	0.25	0.11	0.63	**0.00**	**0.00**	0.34	**0.00**	**0.00**	0.34

Notes: Results are based on the full in-sample estimation period ending in December 31, 2011. Results for the volatility and Hawkes-POT models producing the best in-sample fit are reported here. Results are in the form of p-values for each of the respective tests (with p-values < 5% shown in bold) and the number of exceptions observed (Exc.). at each confidence level α

Table 3: Out-of-sample VaR accuracy test results.

Returns	Statistics	Volatility Models *(Out-sample)*									Hawkes-POT Models *(Out-sample)*					
		GJRGARCH+IV			EGARCH+IV			BEGE+IV			Univariate (M1)			Bivariate (M3)		
S&P500 - VIX	α- level	0.95	0.99	0.999	0.95	0.99	0.999	0.95	0.99	0.999	0.95	0.99	0.999	0.95	0.99	0.999
	Exc.	5	0	0	10	0	0	18	2	0	38	8	1	23	3	1
	LRuc	**0.00**	**0.00**	0.32	**0.00**	**0.00**	0.32	0.13	0.12	0.32	**0.02**	0.23	0.54	0.63	0.32	0.54
	LRind	0.75	1.00	1.00	0.52	1.00	1.00	0.25	0.90	1.00	0.56	0.61	0.95	0.14	0.85	0.95
	LRcc	**0.00**	**0.01**	0.61	**0.00**	**0.01**	0.61	0.16	0.30	0.61	0.05	0.42	0.83	0.30	0.60	0.83
	DQhit	0.76	1.00	1.00	0.53	1.00	1.00	0.26	0.90	1.00	0.58	0.61	0.95	0.15	0.85	0.95
	DQVaR	0.79	1.00	1.00	0.42	1.00	1.00	0.17	0.91	1.00	0.84	0.71	0.99	0.30	0.95	1.00
	MCuc	**0.00**	**0.01**	0.87	**0.00**	**0.00**	0.25	0.15	0.16	1.00	**0.01**	0.25	0.63	0.69	0.39	0.60
	MCind	0.78	1.00	1.00	0.75	1.00	1.00	0.94	0.91	1.00	0.23	0.76	0.91	0.84	0.42	0.91
	MCcc	0.45	**0.00**	0.88	0.53	1.00	0.41	0.12	0.18	1.00	0.45	0.48	0.20	0.34	0.83	0.19
DAX - VDAX	Exc.	19	6	0	24	6	0	21	5	0	27	3	0	29	2	0
	LRuc	0.19	0.67	0.32	0.82	0.67	0.32	0.36	0.98	0.31	0.69	0.33	0.32	0.43	0.12	0.32
	LRind	0.75	0.70	1.00	0.88	0.70	1.00	0.89	0.75	1.00	0.67	0.85	1.00	0.55	0.90	1.00
	LRcc	0.41	0.85	0.61	0.96	0.85	0.61	0.65	0.95	0.60	0.85	0.61	0.61	0.61	0.30	0.61
	DQhit	0.75	0.70	1.00	0.89	0.70	1.00	0.89	0.75	1.00	0.68	0.85	1.00	0.56	0.90	1.00
	DQVaR	0.58	**0.02**	1.00	0.91	0.17	1.00	0.74	0.13	1.00	0.17	0.95	1.00	0.27	0.97	1.00
	MCuc	0.22	0.66	1.00	0.91	0.74	1.00	0.34	0.87	1.00	0.63	0.43	1.00	0.45	0.17	1.00
	MCind	0.89	0.27	1.00	0.60	0.28	1.00	0.81	0.92	1.00	0.40	0.41	1.00	0.30	0.60	1.00
	MCcc	0.20	0.53	1.00	0.79	0.54	1.00	0.37	0.16	1.00	0.78	0.82	1.00	0.61	0.80	1.00
DJI - VXD	Exc.	15	1	0	18	1	1	14	1	0	24	3	0	25	3	0
	LRuc	**0.03**	**0.03**	0.32	0.13	**0.03**	0.54	**0.01**	**0.03**	0.32	0.84	0.33	0.32	0.99	0.33	0.32
	LRind	0.34	0.95	1.00	0.67	0.95	0.95	0.37	0.95	1.00	0.88	0.85	1.00	0.52	0.85	1.00
	LRcc	0.05	0.09	0.61	0.28	0.09	0.82	**0.03**	0.09	0.61	0.97	0.62	0.61	0.81	0.62	0.61
	DQhit	0.35	0.95	1.00	0.68	0.95	0.95	0.38	0.95	1.00	0.88	0.85	1.00	0.53	0.85	1.00
	DQVaR	0.51	0.06	1.00	0.78	0.07	0.08	0.68	0.24	1.00	0.33	0.96	1.00	0.32	0.97	1.00
	MCuc	**0.04**	0.06	1.00	0.16	**0.04**	0.57	**0.01**	**0.04**	1.00	0.87	0.45	1.00	0.96	0.54	1.00
	MCind	0.74	0.36	1.00	0.78	0.35	0.36	0.80	0.36	1.00	0.68	0.41	1.00	0.59	0.41	1.00
	MCcc	0.53	0.72	1.00	0.44	0.72	0.72	0.40	0.72	1.00	0.62	0.84	1.00	0.84	0.82	1.00
Nikkei - VXJ	Exc.	19	1	1	25	2	1	18	2	1	28	6	1	26	6	1
	LRuc	0.19	**0.03**	0.54	0.98	0.12	0.54	0.15	0.13	0.53	0.50	0.64	0.53	0.78	0.64	0.53
	LRind	0.71	0.95	0.95	0.85	0.90	0.95	0.68	0.90	0.95	0.63	**0.01**	0.95	0.76	0.05	0.95
	LRcc	0.40	0.09	0.82	0.98	0.30	0.82	0.33	0.32	0.82	0.71	**0.01**	0.82	0.92	0.14	0.82
	DQhit	0.71	0.95	0.95	0.85	0.90	0.95	0.69	0.90	0.95	0.64	**0.01**	0.95	0.77	0.06	0.95
	DQVaR	0.91	0.75	0.76	0.97	0.30	0.99	0.83	0.52	0.34	0.85	**0.02**	1.00	0.94	0.15	1.00
	MCuc	0.24	0.07	0.64	0.95	0.22	0.28	0.18	0.20	0.66	0.54	0.63	0.22	0.72	0.65	0.46
	MCind	0.33	0.30	0.30	0.32	0.25	0.30	0.46	0.26	0.31	0.22	0.25	0.30	0.33	0.10	0.29
	MCcc	0.67	0.59	0.60	0.64	0.51	0.61	0.91	0.52	0.60	0.43	0.50	0.62	0.66	0.20	0.61
Nasdaq - VXN	Exc.	28	9	0	31	16	2	18.00	2.00	0.00	25	3	0	26	5	0
	LRuc	0.56	0.11	0.32	0.24	**0.00**	0.11	0.13	0.12	0.32	0.89	0.36	0.32	0.73	0.95	0.32
	LRind	0.07	0.57	1.00	**0.04**	0.30	0.90	0.67	0.90	1.00	0.82	0.85	1.00	0.75	0.75	1.00
	LRcc	0.16	0.23	0.61	0.07	**0.00**	0.28	0.28	0.30	0.61	0.97	0.64	0.61	0.90	0.95	0.61
	DQhit	0.08	0.57	1.00	0.05	0.31	0.90	0.68	0.90	1.00	0.83	0.85	1.00	0.76	0.75	1.00
	DQVaR	0.06	0.26	1.00	0.14	0.52	0.92	0.37	0.93	1.00	0.18	0.95	1.00	0.15	0.90	1.00
	MCuc	0.60	0.08	0.13	0.22	**0.00**	**0.03**	0.11	0.25	1.00	0.90	0.53	1.00	0.76	0.86	1.00
	MCind	0.28	0.09	**0.00**	0.35	0.22	0.70	0.84	0.78	0.00	0.67	0.19	1.00	0.74	0.96	1.00
	MCcc	0.57	0.17	**0.00**	0.71	0.45	0.60	0.31	0.44	1.00	0.68	0.37	1.00	0.51	0.08	1.00

Notes: Results are based on the backtesting period of 2012-2013. Results for the volatility and Hawkes-POT models producing the best in-sample fit are reported here. Results are in the form of p-values for each of the respective tests (with p-values < 5% shown in bold) and the number of exceptions observed (Exc.). at each confidence level α

Table 4: MCS results for comparing VaR forecast performance.

		S&P500 - VIX			DAX - VDAX			DJI - VXD			Nikkei - VXJ			Nasdaq - VXN		
	α-level	0.95	0.99	0.999	0.95	0.99	0.999	0.95	0.99	0.999	0.95	0.99	0.999	0.95	0.99	0.999
Uni. Hawkes-POT	Model 1			*	*	*	*	*				*	*	*	*	*
	Model 2	*	*	*	*	*	*	*				*	*	*	*	*
	Model 3				*	*	*	*			*	*	*	*	*	
Biv. Hawkes-POT	Model 1	*	*	*	*	*	*	*	*	*	*	*	*			
	Model 2	*	*	*	*	*	*	*		*	*	*	*	*	*	*
	Model 3	*	*	*	*	*	*	*		*	*	*	*	*	*	*
	Model 4				*		*	*			*	*	*	*	*	
Volatility Models	GJRGARCH			*	*		*	*					*	*	*	
	GJRGARCH+IV			*	*		*	*					*	*	*	
	EGARCH						*	*					*	*	*	
	EGARCH+IV				*		*	*					*	*	*	
	GARCH						*	*					*	*	*	
	GARCH+IV						*	*					*	*	*	
	BEGE				*	*	*	*				*	*	*	*	
	BEGE+IV						*	*				*	*	*	*	*

Notes: The MCS results are based on the asymmetric quantile loss function in Eq. (3.9), at each VaR level. The MCS results are reported for a level of significance of $\alpha_M = 5\%$, with an * indicating that the model is a member of the final MCS.

BEGE (those including IV) and Hawkes-POT (Univariate M1 and Bivariate M3) models with the full set of out-of-sample results reported across Tables IA.7 (GARCH and BEGE) and IA.8 (Hawkes-POT) in the Internet Appendix. The results are based on 1-day ahead VaR forecasts for the final backtesting period, 2 January, 2012 to 31 Decembers, 2013. The first result that stands out is the frequent rejections of the LRuc, and often MCuc tests for the GARCH models for all markets except the DAX and NASDAQ. This indicates that the GARCH models are producing inaccurate VaR forecasts as the average rate of rejection is significantly different from the given level of significance in many cases. While the BEGE models also produce a number of rejections, they are less frequent than those based on the GARCH forecasts. This improvement reflects the ability of the more flexible BEGE distribution to capture tail behaviour. Apart from a number of rejections of the LRuc and MCuc tests in the case of the S&P 500, the univariate Hawkes-POT models produce few other rejections. In contrast, there are no rejections produced under the bivariate Hawkes-POT forecasts across the five markets considered, indicating that treating the IV events as an additional MPP offers gains in forecast accuracy.

Table 4 reports the MCS results based on the asymmetric quantile loss function in Eq. (3.9) and a level of significance of $\alpha_M = 5\%$. Table 4 shows a * when a model is included in the final MCS at a confidence level of 95%. All models are included here in Table 4 as the initial set of models considered in the MCS testing procedure covers the full range of models. The most significant result is that the bivariate models are included in the final MCS in nearly every case, across all markets and VaR levels. Of these models, Models 1-3, which include both the timing and size of

Table 5: VaR adequacy tests results for $h = 5$ and $h = 10$ day ahead forecasts.

		S&P500 - VIX			DAX - VDAX			DJI - VXD			Nikkei - VXJ			Nasdaq - VXN		
	α-level	0.95	0.99	0.999	0.95	0.99	0.999	0.95	0.99	0.999	0.95	0.99	0.999	0.95	0.99	0.999
VaR for $h = 5$																
Uni. Hawkes-POT	Model 1	31	1	0	30	1	0	34	1	0	35	2	0	34	5	0
		(0.22)	(**0.03**)	(0.32)	(0.33)	(**0.03**)	(0.32)	(0.07)	(**0.03**)	(0.32)	(0.05)	(0.13)	(0.32)	(0.05)	(0.94)	(0.33)
	Model 2	29	2	0	31	2	0	33	1	0	38	1	0	36	5	0
		(0.40)	(0.13)	(0.32)	(0.24)	(0.12)	(0.32)	(0.10)	(**0.03**)	(0.32)	(**0.01**)	(**0.03**)	(0.32)	(**0.02**)	(0.94)	(0.33)
	Model 3	9	0	0	23	1	0	27	2	0	32	2	0	26	3	0
		(**0.00**)	(**0.00**)	(0.32)	(0.66)	(**0.03**)	(0.32)	(0.65)	(0.13)	(0.32)	(0.15)	(0.13)	(0.32)	(0.70)	(0.37)	(0.33)
Biv. Hawkes-POT	Model 1	29	2	0	34	6	0	33	1	0	35	2	0	37	6	0
		(0.40)	(0.13)	(0.32)	(0.05)	(0.67)	(0.32)	(0.10)	(**0.03**)	(0.32)	(0.05)	(0.13)	(0.32)	(**0.01**)	(0.61)	(0.33)
	Model 2	29	2	0	34	6	0	33	2	0	35	2	0	36	5	0
		(0.40)	(0.13)	(0.32)	(0.05)	(0.67)	(0.32)	(0.10)	(0.13)	(0.32)	(0.05)	(0.13)	(0.32)	(**0.02**)	(0.94)	(0.33)
	Model 3	31	2	0	34	6	0	34	1	0	42	2	0	37	5	0
		(0.22)	(0.13)	(0.32)	(0.05)	(0.67)	(0.32)	(0.07)	(**0.03**)	(0.32)	(**0.00**)	(0.13)	(0.32)	(**0.01**)	(0.94)	(0.30)
	Model 4	16	2	0	23	2	0	27	2	0	38	3	0	27	3	0
		(0.05)	(0.13)	(0.32)	(0.66)	(0.12)	(0.32)	(0.65)	(0.13)	(0.32)	(**0.01**)	(0.34)	(0.32)	(0.56)	(0.37)	(0.33)
VaR for $h = 10$																
Uni. Hawkes-POT	Model 1	19	1	0	20	0	0	30	1	0	36	3	0	24	3	0
		(0.23)	(**0.03**)	(0.32)	(0.30)	(**0.00**)	(0.32)	(0.27)	(**0.03**)	(0.32)	(**0.03**)	(0.35)	(0.32)	(0.98)	(0.38)	(0.33)
	Model 2	19	1	0	21	0	0	30	1	0	37	3	0	24	3	0
		(0.23)	(**0.03**)	(0.32)	(0.42)	(**0.00**)	(0.32)	(0.27)	(**0.03**)	(0.32)	(**0.02**)	(0.35)	(0.32)	(0.98)	(0.38)	(0.33)
	Model 3	8	1	0	15	0	0	24	2	0	36	0	0	18	2	0
		(**0.00**)	(**0.03**)	(0.32)	(**0.03**)	(**0.00**)	(0.32)	(0.92)	(0.13)	(0.32)	(**0.03**)	(**0.00**)	(0.32)	(0.2)	(0.15)	(0.33)
Biv. Hawkes-POT	Model 1	18	1	0	31	1	0	30	1	0	36	3	0	23	5	0
		(0.15)	(**0.03**)	(0.32)	(0.22)	(**0.03**)	(0.32)	(0.27)	(**0.03**)	(0.32)	(**0.03**)	(0.35)	(0.32)	(0.85)	(0.92)	(0.33)
	Model 2	18	1	0	31	1	0	32	2	0	36	3	0	24	3	0
		(0.15)	(**0.03**)	(0.32)	(0.22)	(**0.03**)	(0.32)	(0.14)	(0.13)	(0.32)	(**0.03**)	(0.35)	(0.32)	(0.98)	(0.38)	(0.33)
	Model 3	8	1	0	31	1	0	33	1	0	41	6	0	24	3	0
		(**0.00**)	(**0.03**)	(0.32)	(0.22)	(**0.03**)	(0.32)	(0.09)	(**0.03**)	(0.32)	(**0.00**)	(0.63)	(0.32)	(0.98)	(0.38)	(0.33)
	Model 4	12	1	0	16	1	0	25	2	0	40	2	0	18	2	0
		(**0.00**)	(**0.03**)	(0.32)	(0.05)	(**0.03**)	(0.32)	(0.92)	(0.13)	(0.32)	(**0.00**)	(0.14)	(0.32)	(0.20)	(0.15)	(0.33)
Expected		25.10	5.02	0.50	25.35	5.07	0.51	25.00	5.00	0.50	25.00	5.00	0.50	24.40	4.88	0.49
Shape $\hat{\xi}_u$			0.476			0.436			0.468			0.375			0.388	

Notes: Results of the LRuc test are shown in the form of p-values. p-values less than 5% are shown in bold to highlight where the rejections of the accuracy tests are occurring.

IV events are virtually always included in the MCS. Model 4, which includes only the timing of the IV events is excluded in the majority of cases. These results once again support the notion that treating IV as an additional point process and considering the size and timing of these events leads to the greatest benefit in terms of forecast accuracy. The univariate Hawkes-POT models are included in the MCS in well over half the cases. Of these models, Model 2, which only includes the size of past return events, is most frequently included in the MCS. The BEGE models follow closely in terms of forecast performance and remain in the MCS in about half of the cases. Finally, the GARCH models are inferior, and are excluded from the MCS in the majority of cases.

Overall these results reveal that harnessing information from IV, when treated as its own point process is beneficial. Given the bivariate Hawkes-POT models produce the most accurate forecasts (that pass all tests) across the widest range of scenarios, indicating that information regarding the timing and size of past IV extreme events is of benefit in a multivariate setting for forecasting VaR in equity markets. The benefit of including IV in a univariate point process model or the BEGE framework, is somewhat more limited, and of little use in the context of GARCH models.

While the bivariate MPP models appear to dominate at the one day horizon, the final analysis determines whether adequate VaR forecasts can be generated from the MPP models at longer horizons. Based on 5- and 10-day VaR forecasts using the methodology discussed earlier in Section 3, Table 5 reports results for the LRuc test based on both the univariate and bivariate MPP models. Given the small number of MPP models, results for all the MPP models are presented in Table 5. Again, the results shown in bold indicate when a rejection at 5% is observed. At the $h = 5$ day horizon, the adequacy of the coverage is only rejected in 22% and 12% of the cases for the univariate and bivariate models, once again indicating the information in IV is best harnessed through a bivariate MPP. While unsurprisingly, the rejection rates do rise moving to the longer horizon of $h = 10$, the adequacy of the coverage is only rejected in a quarter of cases for the bivariate models (33% of cases for the univariate models).

In summary, the bivariate Hawkes-POT models that include IV as an additional MPP produce the best performing model across the widest range of scenarios. They pass all of the individual tests of VaR forecast adequacy, they are most frequently found to be amongst the most accurate under asymmetric quantile loss, and are able to generate adequate VaR forecasts in most cases at a longer one-week ahead (somewhat less at two-weeks ahead) forecast horizon.

5. Conclusion

Modelling and forecasting the occurrence of extreme events in financial markets is crucially important. While there have been many studies considering the role of implied volatility (IV) for forecasting volatility, this has not been the case when dealing with extreme events. This paper addressed how best to use IV to generate forecasts of the risk of extreme events in the form of Value-at-Risk (VaR).

The BEGE model, along with traditional GARCH models including IV as an exogenous variable, coupled with EVT formed the benchmark set of models. More recent advances in VaR prediction have employed marked point process (MPP) models that treat the points as the occurrence of extreme events and marks their associated size. This paper proposed a number of novel MPP models that include IV. A number of univariate models for extreme return events are developed, where the size and timing of past return events and IV are included. In addition, novel bivariate MPP models were also proposed that move beyond simply including IV as an exogenous covariate. The second dimension in the bivariate models, apart from extreme stock market losses, were extreme increases in IV which were treated as a second MPP.

The empirical analysis here focused on a number of major equity market indices and their associated IV indices, where the full range of models are used to generate estimates of VaR. In terms

of an in-sample explanation of extreme events in equity markets, the bivariate models satisfied all backtests of VaR adequacy, while the univariate models and the BEGE models passed most. The GARCH models produced relatively frequent rejections. A similar pattern was observed in 1-day ahead predictions of VaR. GARCH style models including IV generated inaccurate forecasts of VaR and failed a number of tests relating to the frequency of the VaR exceptions. Univariate MPP models and BEGE models provided more accurate forecasts though still produced a number of rejections in backtesting. It was also shown that longer horizon VaR forecasts from the bivariate MPP models satisfied most tests. Overall, the bivariate models that included the extreme IV events produced the most accurate forecasts of VaR across the full range of levels of significance and markets. A direct comparison of VaR forecast accuracy showed that the bivariate MPP models that consider the size and timing of past IV events were among the most accurate in the widest range of cases. These results show that while IV is certainly of benefit for predicting extreme movements in equity returns, the framework within which it is used is important. It is shown that the novel bivariate MPP model proposed here leads to superior forecasts of extreme risk in a VaR context.

Acknowledgements

Herrera acknowledges the Chilean CONICYT funding agency for financial support (FONDECYT 1150349) for this project. The authors also wish to thank Dominik Weid for proving computer code for implementing the Monte Carlo simulation based backtests.

References

Aboura, S. and N. Wagner (2016). Extreme asymmetric volatility: Stress and aggregate asset prices. *Journal of International Financial Markets, Institutions and Money 41*, 47–59.

Aït-Sahalia, Y., J. Cacho-Diaz, and R. J. Laeven (2015). Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics 117*(3), 585–606.

Aït-Sahalia, Y. and T. R. Hurd (2015). Portfolio choice in markets with contagion. *Journal of Financial Econometrics 14*(1), 1–28.

Balkema, A. A. and L. De Haan (1974). Residual life time at great age. *The Annals of Probability 5 (2)*, 792 – 804.

Becker, R., A. Clements, and A. McClelland (2009). The jump component of S&P 500 volatility and the VIX index. *Journal of Banking & Finance 33*(6), 1033–1038.

Bekaert, G., E. Engstrom, and A. Ermolov (2015). Bad environments, good environments: A non-gaussian asymmetric volatility model. *Journal of Econometrics 186*(1), 258 – 275.

Bekaert, G. and M. Hoerova (2014). The vix, the variance premium and stock market volatility. *Journal of Econometrics 183*(2), 181 – 192.

Bekaert, G. and G. Wu (2000). Asymmetric volatility and risk in equity markets. *Review of Financial Studies 13*(1), 1–42.

Blair, B. J., S. H. Poon, and S. J. Taylor (2001). Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns. *Journal of Econometrics 105*, 5–26.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*, 307–327.

CBOE (2003). VIX CBOE volatility index. *Chicago Board Option Exchange, White Paper, www. cboe. com*.

Chavez-Demoulin, V., A. Davison, and A. McNeil (2005). A point process approach to value-at-risk estimation. *Quantitative Finance 5*, 227–234.

Chavez-Demoulin, V. and J. McGill (2012). High-frequency financial data modeling using hawkes processes. *Journal of Banking & Finance 36*(12), 3415–3426.

Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review 39*, 841–862.

Cotter, J. (2007). Varying the var for unconditional and conditional environments. *Journal of International Money and Finance 26*(8), 1338 – 1354.

Danielsson, J. and C. De Vries (2000). Value-at-risk and extreme returns. *Annales d'Economie et de Statistique 60*, 239–270.

Davis, R. A., T. Mikosch, et al. (2009). The extremogram: a correlogram for extreme events. *Bernoulli 15*(4), 977–1009.

Davis, R. A., T. Mikosch, and I. Cribben (2012). Towards estimating extremal serial dependence via the bootstrapped extremogram. *Journal of Econometrics 170*(1), 142–152.

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

Engle, R. and S. Manganelli (2004). Caviar. *Journal of Business and Economic Statistics 22*(4), 367–381.

Fernandez, C. and M. Steel (1998). On Bayesian Modeling of Fat Tails and Skewness. *Journal of the American Statistical Association 93*(441), 359–371.

Giot, P. (2005). Relationships between implied volatility indexes and stock index returns. *The Journal of Portfolio Management 31*(3), 92–100.

Glosten, L., R. Jagannathan, and D. Runkle (1993). On the relation between the expected value and the volatility of

the nominal excess return on stocks. *Journal of Finance 48*, 1779–1801.

González-Rivera, G., T.-H. Lee, and S. Mishra (2004). Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of Forecasting 20*(4), 629–645.

Hansen, P., A. Lunde, and J. Nason (2003). Choosing the best volatility models: the model confidence set approach. *Oxford Bulletin of Economics and Statistics 65*, 839–861.

Hansen, P., A. Lunde, and J. Nason (2011). The model confidence set. *Econometrica 79*, 453–497.

Herrera, R. (2013). Energy risk management through self-exciting marked point process. *Energy Economics 38*, 64–76.

Herrera, R. and N. González (2014). The modeling and forecasting of extreme events in electricity spot markets. *International Journal of Forecasting 30*(3), 477–490.

Herrera, R. and B. Schipp (2013). Value at risk forecasts by by extreme value models in a conditional duration framework. *Journal of Empirical Finance 23*, 33–47.

Hilal, S., S.-H. Poon, and J. Tawn (2011). Hedging the black swan: Conditional heteroskedasticity and tail dependence in S&P500 and VIX. *Journal of Banking & Finance 35*(9), 2374–2387.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivates 3*(2), 73–84.

Lee, K. and B. K. Seo (2017). Marked hawkes process modeling of price dynamics and volatility estimation. *Journal of Empirical Finance 40*, 174 – 200.

Lin, Y.-N. and C.-H. Chang (2010). Consistent modeling of S&P 500 and VIX derivatives. *Journal of Economic Dynamics and Control 34*(11), 2302–2319.

McNeil, A. and R. Frey (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance 7*, 271–300.

Nelson, D. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica 59*, 347–370.

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

Peng, Y. and W. L. Ng (2012). Analysing financial contagion and asymmetric market dependence with volatility indices via copulas. *Annals of Finance 8*(1), 49–74.

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics 3*(1), 119 – 131.

Poon, S.-H. and C. W. J. Granger (2003). Forecasting volatility in financial markets: a review. *Journal of Economic Literature 41*, 478–539.

Santos, P. A. and M. F. Alves (2012). Forecasting value-at-risk with a duration-based POT method. *Mathematics and Computers in Simulation 94*, 295 – 309.

Wagner, N. and A. Szimayer (2004). Local and spillover shocks in implied market volatility: evidence for the U.S. and Germany. *Research in International Business and Finance 18*(3), 237 – 251.

Ziggel, D., T. Berens, G. N. Weiß, and D. Wied (2014). A new set of improved value-at-risk backtests. *Journal of Banking & Finance 48*, 29–41.

AppendixA. Proofs

Proof. (Proposition 1) Using the continuous representation of a Hawkes process and setting the expected intensity $\mathbb{E}[\lambda_g(t \mid \mathscr{H}_t)] = \lambda_0 < \infty$, gives for the univariate case

$$\begin{aligned}\mathbb{E}[\lambda_g(t \mid \mathscr{H}_t)] &= \mathbb{E}\left[\nu + \vartheta \int_{(-\infty,t)\times\mathbb{R}_+^2} f(w,z)h(t-s)\,N(ds\times dw\times dz)\right] \\ &= \nu + \vartheta\mathbb{E}[f(w,z)]\,\mathbb{E}\left[\int_{(-\infty,t)} h(t-s)\,\lambda_g(s \mid \mathscr{H}_s)ds\right]\end{aligned}$$

and by assuming $\mathbb{E}[f(w,z)] = \mu_{wz}$ and by defining $\lambda_g(s \mid \mathscr{H}_s)ds = N(ds)$ leads to

$$\begin{aligned}\mathbb{E}[\lambda_g(t \mid \mathscr{H}_t)] &= \nu + \vartheta\mu_{wz}\int_{(-\infty,t)} h(t-s)\,\lambda_0 ds \\ &= \nu + \vartheta\mu_{wz}\lambda_0\int_{(0,\infty)} h(s)\,ds \\ &= \nu + \vartheta\mu_{wz}\lambda_0,\end{aligned}$$

where finally $\lambda_0 = (1-\vartheta\mu_{wz})^{-1}\nu$ is obtained. Hence, the expectation of the ground conditional intensity is finite in the univariate case, if and only if, $0 < \vartheta\mu_{wz} < 1$.

In the bivariate model the demonstration follows the same steps. Assume that the expected intensity $\mathbb{E}\left[\lambda_g^k(t \mid \mathscr{H}_t)\right] = \lambda_0^k < \infty$, for $k = 1,2$. Then, by taking the unconditional expectation in (2.12) leads to

$$\begin{aligned}\mathbb{E}\left[\lambda_g^1(t \mid \mathscr{H}_t)\right] &= \mathbb{E}\left[\nu_1 + \vartheta_{11}\int_{(-\infty,t)\times\mathbb{R}_+} f_1(w)h_1(t-s)\,N_1(ds\times dw)\right] \\ &\quad + \mathbb{E}\left[+\vartheta_{12}\int_{(-\infty,t)\times\mathbb{R}_+} f_1(z)h_2(t-s)\,N_2(ds\times dz)\right] \\ \mathbb{E}\left[\lambda_g^2(t \mid \mathscr{H}_t)\right] &= \mathbb{E}\left[\nu_2 + \vartheta_{21}\int_{(-\infty,t)\times\mathbb{R}_+} f_2(w)h_1(t-s)\,N_1(ds\times dw)\right] \\ &\quad + \mathbb{E}\left[\vartheta_{22}\int_{(-\infty,t)\times\mathbb{R}_+} f_2(z)h_2(t-s)\,N_2(ds\times dz)\right]\end{aligned}$$

by assuming $\mathbb{E}[f_k(w)] = \mu_w^k$ and $\mathbb{E}[f_k(z)] = \mu_z^k$, the expectations are reduced to

$$\begin{aligned}
\mathbb{E}\left[\lambda_g^1(t \mid \mathscr{H}_t)\right] &= \nu_1 + \vartheta_{11}\mu_w^1 \mathbb{E}\left[\int_{(-\infty,t)} h_1(t-s)\, N_1(ds)+\right] + \vartheta_{12}\mu_z^1 \mathbb{E}\left[\int_{(-\infty,t)} h_2(t-s)\, N_1(ds)\right] \\
\mathbb{E}\left[\lambda_g^2(t \mid \mathscr{H}_t)\right] &= \nu_2 + \vartheta_{21}\mu_w^2 \mathbb{E}\left[\int_{(-\infty,t)} h_1(t-s)\, N_1(ds)+\right] + \vartheta_{22}\mu_z^2 \mathbb{E}\left[\int_{(-\infty,t)} h_2(t-s)\, N_2(ds)\right].
\end{aligned}$$

Since the kernel functions satisfy $\int_0^\infty h_k(s)\, ds = 1$ and $\lambda_g^k(s \mid \mathscr{H}_s)ds = N_k(ds)$ for $k = 1,2$, it is possible to express

$$\begin{aligned}
\mathbb{E}\left[\lambda_g^1(t \mid \mathscr{H}_t)\right] &= \nu_1 + \vartheta_{11}\mu_w^1 \mathbb{E}\left[\int_{(0,\infty)} h_1(s)\, \lambda_g^1(s \mid \mathscr{H}_s)ds\right] + \vartheta_{12}\mu_z^1 \mathbb{E}\left[\int_{(0,\infty)} h_2(s)\, \lambda_g^2(s \mid \mathscr{H}_s)ds)\right] \\
\mathbb{E}\left[\lambda_g^2(t \mid \mathscr{H}_t)\right] &= \nu_2 + \vartheta_{21}\mu_w^2 \mathbb{E}\left[\int_{(0,\infty)} h_1(s)\, \lambda_g^1(s \mid \mathscr{H}_s)ds\right] + \vartheta_{22}\mu_z^2 \mathbb{E}\left[\int_{(0,\infty)} h_2(s)\, \lambda_g^2(s \mid \mathscr{H}_s)ds)\right],
\end{aligned}$$

which in turn is equivalent to

$$\begin{aligned}
\mathbb{E}\left[\lambda_g^1(t \mid \mathscr{H}_t)\right] &= \nu_1 + \vartheta_{11}\mu_w^1\lambda_0^1 + \vartheta_{12}\mu_z^1\lambda_0^2 \\
\mathbb{E}\left[\lambda_g^2(t \mid \mathscr{H}_t)\right] &= \nu_2 + \vartheta_{21}\mu_w^2\lambda_0^1 + \vartheta_{22}\mu_z^2\lambda_0^2,
\end{aligned}$$

or in matrix representation

$$\lambda_{\mathbf{0}} = \nu + (M \circ Q)\, \lambda_{\mathbf{0}},$$

where $\nu = (\nu_1, \nu_2)^T$, $M = \begin{pmatrix} \mu_w^1 & \mu_z^1 \\ \mu_w^2 & \mu_z^2 \end{pmatrix}$ and $Q = \begin{pmatrix} \vartheta_{11} & \vartheta_{12} \\ \vartheta_{21} & \vartheta_{22} \end{pmatrix}$. Hence the unconditional expectation of the ground intensity given by $\lambda_0 = (\mathbf{1_2} - M \circ Q)^{-1}\nu$ exists, if and only if, the spectral radius of the matrix $M \circ Q$ is less than one.