A marked point process model for intraday financial returns: Modelling extreme risk

December 1, 2016

Abstract

Forecasting the risk of extreme losses is an important issue in the management of financial risk and has attracted a great deal of research attention. However, little attention has been paid to extreme losses in a higher frequency intraday setting. This paper proposes a novel marked point process model to capture extreme risk in intraday returns, taking into account a range of trading activity and liquidity measures. A novel approach is proposed for defining the threshold upon which extreme events are identified taking into account the diurnal patterns in intraday trading activity. It is found that models including covariates, mainly relating to trading intensity and spreads offer the best in-sample fit, and prediction of extreme risk, in particular at higher quantiles.

Keywords
Hawkes process, peaks over threshold, bid-ask spread, extreme risk, high frequency

JEL Classification Numbers C32, C53, C58.

1 Introduction

Modeling and forecasting the volatility of asset returns is an important issue in the management of financial risk and has attracted a great deal of research attention. In recent years, research attention has shifted its focus to the issue of accurate estimates of risk measures such as Value-at-Risk (VaR) that capture the risk of extreme losses. For a model to be successful in dealing with these extreme loss events it 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 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 process. This class of model has been applied in a range of different financial contexts. For instance, Gresnigt et al. (2015) developed an early warning system for market crashes, Herrera and Schipp (2014) a model for financial risk, Chavez-Demoulin and McGill (2012) a model for high-frequency returns, Herrera and Gonzalez (2014) and Clements et al. (2015) a model for risk in energy markets.

While Chavez-Demoulin and McGill (2012) and Liu and Tse (2015) consider extreme intraday risk, they ignore the possible important role of covariates reflecting intraday trading activity and liquidity. With the availability of high frequency financial data, recent studies have examined the link between order flow which reflects trading activity and volatility. Evans and Lyons (2002) and Berger et al. (2009) found that the persistence in order flow helps explain the observed persistence in volatility in the foreign exchange market, a finding also confirmed by Opschoor et al. (2014) in the bond market. Studies such as Chordia et al. (2001, 2002), Hasbrouck and Seppi (2001), Groß-Klußmann and Hautsch (2011), examine the dynamic behaviour of liquidity and its interaction with trading activity, returns and volatility. Ahn et al. (2001) investigate the relationship between transitory volatility and the depth of the order book, with their results showing that a rise in transitory volatility is followed by an increase in book depth, and an increase in book depth is usually followed by a decrease in transitory volatility. Næs and Skjeltorp (2006) find that volatility and volume are negatively related to the slope of the order book. While the links between liquidity and volatility are well understood, the role of these variables in the context of extreme losses is understood less.

Chavez-Demoulin et al. (2015) incorporate covariates into a Peaks Over Threshold (POT) model of extreme losses in the context of operational risk, however no study has considered how to incorporate such covariates into a model for intraday risk. To examine how trading activity and order book liquidly measures help account for the dynamics of extreme losses, a novel Hawkes-POT approach is developed extending the work of Chavez-Demoulin and McGill (2012). Here, careful attention is paid to the way in which extreme events are defined, with a method proposed for a time-varying threshold based on quantile regression. The time-varying threshold removes the commonly observed diurnal pattern in intraday trading activity. Otherwise extreme events are much more likely to be observed early in the trading day when volatility is higher, and hence returns of greater magnitude are expected to be observed. Of the covariates considered, order flow reflecting signed trading intensity, best bid-price depth and spread reflecting liquidity, and the rate of new bid limit order arrivals are important for describing the dynamics of extreme intraday losses, providing the best insample model fit. The importance of the covariates extend to the context of VaR prediction, where models containing these covariates produce the most accurate predictions, particularly at higher quantiles.

The paper proceeds as follows. Section 2 describes the intraday data used in the empirical analysis including the covariates considered. Section 3 develops the Hawkes-POT model proposed here, along with the time varying threshold used to define extreme events that accounts for the diurnal pattern in volatility. Section 4 describes the tests used to examine forecast accuracy. Sections 5 and 6 discuss the empirical results and provide concluding comments.

2 Data

This paper uses 1-minute data for the period January 4 2011 to December 29 2012 relating to BHP traded on the Australian Stock Exchange (ASX). Days with trading halts are removed, leaving 169,750 1-minute observations across 485 trading days. The trading day begins at 10:10 AM and ends at 4:00 PM, Monday to Friday. The ASX applies an algorithm to determine the opening price of the security by using the buy and sell orders available, with the goal of maximising the number of shares traded. This process begins at 10:00 AM, when the market technically opens, and finishes at approximately 10:10 AM. The data used begins at 10:10 AM until 4:00 PM to avoid this opening auction period, a common practice among studies using ASX data, Hall and Hautsch (2006, 2007). BHP was chosen for this analysis as it is one of the most actively traded stocks on the ASX. The data used in this study was obtained from the Securities Industry Research Centre of Asia-Pacific (SIRCA) database. The raw dataset consists of quote prices and volumes, and order book data, at a tick level, which is aggregated up to a 1-minute frequency where required. t represents each 1-minute interval within a day with negative returns are calculated as $X_t = -ln\left(\frac{C_t}{C_{t-1}}\right)$ where C_t is the closing price in an interval.

The first covariate of interest is order flow, a measure of trading activity denoted below as OF. Buyer (seller) initiated trades occur when a market buy (sell) order is executed against a sell (buy) limit order. Each buy is defined as positive and each sale is defined as negative volume. Therefore, order flow can be understood as signed volume within each 1-minute interval and reflects buying or selling pressure. Bid-ask spread (denoted below as S), the difference between the best ask and bid prices measured at the start of each 1-minute interval is considered given its common use as a measure of liquidity. The remaining trading and order-book measures relate to the bid (lower)

	Mean	SD	Min	Max	Skew	Kurt	L-B	JB	ADF
X_t	0.0000	0.0006	0.0027	-0.0028	-0.05	4.90	5.613	264*	-12.1*
OF	-800	14053	-90171	74556	-0.11	7.19	16.9*	1282*	-8.7*
AS	8921	10847	0.00	104217	2.54	12.54	204.3*	8515*	-7.9*
NBV	6041	7588	0.00	64568	2.43	11.18	468.2*	6598*	-5.7*
BD	7366	7319	1.00	83219	2.74	17.57	110.1*	17671*	-9.0*
S	0.015	0.006	0.01	0.04	0.805	2.93	100.2*	189*	-8.7*

Table 1: Descriptive statistics of daily negative log-returns and the different covariates. The Ljung-Box statistics significant at 1%, at a lag of 5 intervals are denoted by *.

side of the market as the central focus here is on extreme downward movements in price and hence negative returns. As an alternative to total OF, the role of the volume of aggressive sell trades in each 1-minute interval (denoted as AS) that eliminate all of the standing volume at the best bid price in the order book is also considered. AS reflects aggressive selling activity that will widen the spread, lower the best bid price and potentially lead to further falls in price. The depth of standing volume at the best bid price (denoted as BD) is considered as a measure of liquidity reflecting volume to buffer against selling pressure. The final covariate considered is the rate of new bid limit orders arriving within each interval (denoted as NBV) to capture the rate at which the standing volume at the best bid price in the order book is replenished.

Table 1 reports a range of summary statistics for the intraday returns and associated covariates. The first row shows that the returns, X_t are symmetric, exhibit excess kurtosis and are non-normal according to the Jarque-Bera test, exhibit no significant autocorrelation given the Ljung-Box test and are stationary according to the ADF test. The features of OF are broadly consistent with the returns. Aggressive sell volume, AS takes a minimum value of zero as there will be periods where no selling activity removes the entire best bid volume, is right skewed, autocorrelated but stationary. The rate of new bid volume arriving, NBV in fact shares the same characteristics as AS, also having a minimum value of zero as some intervals will contain no new bid limit orders. Bid depth, BD takes a minimum value of one, indicating that as a minimum, an order with a volume of one was sitting in the order book at the best bid price. However the mean level of depth is 7366 and BD is heavily skewed. The spread, S ranges between 0.01 and 0.04, with a mean of only 0.015 indicating that much of the time S sits at either 0.01 or 0.02. Along with all of the covariates, S is also persistent but stationary.

3 Methodology

This section introduces the self-exciting marked point process (SEMPP) methodology used here to describe the extreme value dynamics of 1-minute financial returns, in terms of the occurrence and size of the extreme events. In particular, a modified version of the Hawkes-POT point process of

Chavez-Demoulin et al. (2005) is used. Under this approach, the instantaneous behavior of random events and their magnitudes are based on the history of the process by means of its conditional intensity. This framework is modified in the paper to account for the salient features of intraday financial data.

3.1 The Hawkes-POT model

Let $N(t)$, $t \in \mathbb{R}$, be a SEMPP relating the occurrence of a stochastic process $\{(t_i, Y_i)\}$, which corresponds to all extreme events in the return on a stock $\{X_t\}_{t\geq 1}$ up to time t. Define extreme events (losses) as those observations occurring at time t_i over a time-varying high threshold $u_t > 0$, where $Y_i \in \mathbb{R}_+$ are the magnitudes attached to the occurrence of these events, denoted as marks, i.e., $Y_i = X_{t_i} - u_{t_i}$. These marks have a conditional probability density function $g(y \mid t, \mathcal{H}_t)$, where $\mathcal{H}_t := \{(t_i, Y_i) : t_i < t\}$ is the σ-field generated by $N(t)$, i.e., the history of the process up to but not including time t. A convenient and intuitive way of specifying the dynamics of a marked point process is by means of its conditional intensity function

$$\lambda(t, y \mid \mathcal{H}_t) = \lim_{h \to 0} \frac{1}{h} P(N(t+h) - N(t) > 0 \mid \mathcal{H}_t),$$

which can also be decomposed into two stochastic process

$$\lambda(t, y \mid \mathcal{H}_t) = \lambda_g(t \mid \mathcal{H}_t)\, g(y \mid t, \mathcal{H}_t).$$

The first process $\lambda_g(t \mid \mathcal{H}_t)$ is called the conditional ground intensity and characterizes the occurrence of extreme events, while the second process specifies the distribution of the marks given t and the history $\mathcal{H}_t$. Here, a modified version of the ground intensity of a Hawkes-POT process is employed:

$$\lambda_g(t \mid \mathcal{H}_t) = \mu + \eta \int_{(-\infty,t)} \phi(y)\, h(t-s)\, dN(s) = \mu + \eta \sum_{k: t_k < t} \phi(y)\, h(t - t_k) \tag{1}$$

where $\mu > 0$ is the baseline intensity reflecting exogenous Poisson arrivals, $h : \mathbb{R} \to \mathbb{R}^+$ is a decay kernel function capturing the self-exciting behavior through the time and $\phi : \mathbb{R} \to \mathbb{R}^+$ is an impact function representing the instantaneous influence of the marks on the ground intensity. These two last functions, $h(\cdot)$ and $\phi(\cdot)$, explain the clustering behavior through long-term and short-term components, respectively. Finally, the branching parameter $\eta > 0$ describes the degree of endogeneity in the extreme events. As usual in empirical applications, the decay kernel function is based on an exponential specification:

$$h(t - t_k) = \beta \exp(-\beta(t - t_k)), \ \beta \geq 0. \tag{2}$$

According to the Pickands-Balkema-de Haan theorem (Balkema and De Haan, 1974; Pickands III, 1975) the asymptotic tail distribution of the stochastic process for the marks converges to the generalized Pareto distribution (GPD) in the iid case. A conditional GPD with a time varying scale parameter is employed such that mean and variance of the distribution of the marks varies through the time. This captures volatility clustering behavior in the size of the extreme events. The probability density function of the proposed GDP is defined as:

$$g\left(y_i \mid t, \mathcal{H}_t\right) = \begin{cases} \frac{1}{\sigma_i}\left(1+\xi\frac{y_i}{\sigma_i}\right)_+^{-1-1/\xi}; \; y_i > 0, & \xi \neq 0 \\ \frac{1}{\sigma_i}\exp\left(\frac{y_i}{\sigma_i}\right) & \xi = 0, \end{cases} \tag{3}$$

where $\sigma_i > 0$ and ξ are the scale and shape parameters respectively. Note that $0 \leq y_i < \infty$ if $\xi \geq 0$, and $0 \leq y_i < -\sigma_i/\xi$ if $\xi < 0$. A simple linear functional form for the impact function is chosen:

$$\phi\left(y_i\right) = \frac{a + by_i}{a + bE\left[Y\right]} = \frac{\left(1 + by_i\right)\left(1 - \xi\right)}{\left(1 - \xi + b\sigma_i\right)}, \tag{4}$$

requiring $a = 1$ and $\xi < 1$ to ensure that the model is identified and the distribution of the marks have a finite mean. The scale parameter σ_i is driven by a set of covariates $\mathbf{Z} := \{OF,\ AS,\ NBV,\ BD,\ S\}$ related to the intraday trading process:

$$\log \sigma_i = \kappa_0 + \sum_{l=1}^{L} \kappa_l Z_{l,t_i-1}, \tag{5}$$

where κ_0 and κ_l are unrestricted coefficients, and Z_{l,t_i-1} are lagged covariates.

Under this specification the branching coefficient $0 < \eta < 1$ appears explicitly as a parameter in the model and corresponds on average to the fraction of extreme events endogenously generated. However, this advantage has a cost in terms of the specification of the impact function, which depends at the same time on the selection of the mark distribution function. Following Embrechts et al. (2011), in order to determine the existence and uniqueness of a stationary Hawkes-POT process the following condition is necessary:

Condition 1. *(Existence and uniqueness) A SEMPP $N(t)$, $t \in \mathbb{R}$, follows a well-defined Hawkes-POT process with ground conditional intensity given by Eq.(1) and distribution of the marks in Eq.(3) if and only if the following assumptions are satisfied*

- $\int_0^\infty h\left(\omega\right) d\omega = 1$
- $\int_0^\infty th\left(\omega\right) d\omega < \infty$ *and*
- $\mathbb{E}\left[\phi\left(y\right)\right] = 1.$

There are two major differences between the model proposed here and that of Chavez-Demoulin and McGill (2012). Firstly, restrictions are imposed on the impact function in order to determine the stationarity of the marked point process $N(t)$. Secondly, the scale parameter of the distribution of the marks is a discrete function depending on a set of covariates and it is only updated with the occurrence of the extreme events. Finally, by means of Condition 1, it is possible to obtain, in a very convenient way, the conditional expectation of the ground conditional intensity.

Preposition 1. *(Stationarity) A Hawkes-POT process with ground conditional intensity given by Eq.(1) and distribution of the marks in Eq.(3) is asymptotically stationary,* $\mathbb{E}\left[\lambda_g(t \mid \mathcal{H}_t)\right] < \infty$*, if satisfies Condition 1 and the branching coefficient* $0 < \eta < 1$.

Proof. By definition $\lambda_g(s \mid \mathcal{H}_s)ds = dN(s)$. Using the continuous representation of Eq.(1) and setting the expected intensity $\mathbb{E}\left[\lambda_g(t \mid \mathcal{H}_t)\right] = m_1$ to a finite value, results in:

$$\begin{aligned}
\mathbb{E}\left[\lambda_g(t \mid \mathcal{H}_t)\right] &= \mathbb{E}\left[\mu + \eta \int_{(-\infty,t)} \phi\left(y\right) h\left(t-s\right) dN(s)\right] \\
&= \mu + \eta \mathbb{E}\left[\phi\left(y\right)\right] \mathbb{E}\left[\int_{(-\infty,t)} h\left(t-s\right) \lambda_g(s \mid \mathcal{H}_s) ds\right] \\
&= \mu + \eta \int_{(-\infty,t)} h\left(t-s\right) \lambda_0 ds \\
&= \mu + \eta m_1 \int_{(0,\infty)} h\left(\omega\right) d\omega \\
&= \mu + \eta m_1,
\end{aligned}$$

where that $m_1 = \left(1-\eta\right)^{-1} \mu$ is obtained. Therefore, expected intensity $\mathbb{E}\left[\lambda_g(t \mid \mathcal{H}_t)\right]$ is finite if and only if $0 < \eta < 1$.

Assuming that we have observed the occurrence of events $\{(t_i, Y_i)\}$ of a stationary Hawkes-POT process in a interval of time $[0, T]$, the parameters of the ground process and the density for the marks defined in (1) and (3) respectively, are obtained by maximizing the following log-likelihood function:

$$\log L = \sum_{i=1}^{N(T)} \log \lambda_g\left(t \mid \mathcal{H}_t\right) - \int_0^T \lambda_g\left(s \mid \mathcal{H}_s\right) ds + \sum_{i=1}^{N(T)} \log g(y \mid \mathcal{H}_t, t).$$

Standard errors are obtained via the Hessian matrix of this log-likelihood, while under standard conditions, consistency, asymptotic normality, and efficiency of maximum likelihood estimators have been established (Ogata, 1978).

3.2 Intraday Threshold Selection

One of the most challenging tasks when analysing extreme events is to determine which observations out of the full sample are actually classified as extreme events, with these exceedances asymptotically approximated by means of a GPD. This choice of a threshold has an impact on model estimation given the trade off between bias and variance. A very high (low) threshold will reduce (increase) the bias in parameter estimates but increase (decrease) their variance.

There are different methods for threshold selection, although none provide a definitive solution. For instance, with non iid data Chavez-Demoulin and McGill (2012) recommends selecting thresholds between 92 % and 95 % through the mean excess plot. Chavez-Demoulin et al. (2014) chose 10% of the most extreme weekly events when dealing with non-stationary data. Herrera and Schipp (2013) propose a methodology based on a sensitivity analysis of VaR, choosing a threshold where the VaR risk estimates are most stable. They conclude that a threshold between the 0.90 and 0.92 quantiles may be optimal.

In this paper, a two-step procedure for threshold selection is proposed. The issue of identifying the observations that exceed a high threshold $u_t > 0$ is complicated by the salient features of intraday financial data. In order to take the features of intraday returns into account, a quantile regression (Koenker and Bassett Jr, 1978) is used in the first step to characterize the threshold. Specifically, the approach of Koenker and Xiao (2006), who developed a quantile autoregressive model (QAR) of absolute value of returns accounting for intraday seasonality, is closely followed. Let u_t^τ denote the conditional τ-th quantile of $\{X_t\}_{t\geq 1}$, such that $P\left(X_t \leq u_t^\tau\right) = \tau$. Assume that the threshold takes the following autoregressive specification:

$$
\begin{aligned}
u_t^\tau := u_t\left(\tau \mid \mathcal{F}_t\right) = \omega\left(\tau\right) + \gamma_1\left(\tau\right)\left|X_{t-1}\right| + \gamma_2\left(\tau\right)\left|X_{t-2}\right| \\
+ \sum_{j=1}^{3} \varphi_j\left(\tau\right)\cos\left(2\pi j \frac{d}{D}\right) + \psi_j\left(\tau\right)\sin\left(2\pi j \frac{d}{D}\right),
\end{aligned} \tag{6}
$$

where $\mathcal{F}_t$ corresponds to the σ-algebra generated by $\{X_s :\ s < t\}$, $\tau \in (0, 1)$ is the specified quantile, $d = 1, \ldots, 350$ denotes the time events and D is the total number of intraday observations, 350 in this case.[1] The degree of the Fourier polynomials in the series expansion is set to three.[2] Estimation of the QAR model (6) involves solving the following problem:

$$
\min_{\Theta} \left\{ (1-\tau) \sum_{X_t < u_t^\tau} \left(u_t\left(\tau \mid \mathcal{F}_t\right) - X_t\right) + \tau \sum_{X_t > u_t^\tau} \left(X_t - u_t\left(\tau \mid \mathcal{F}_t\right)\right) \right\},
$$

with respect to $\Theta = (\omega, \gamma_1, \gamma_2, \varphi_1, \psi_1, \ldots, \varphi_3, \psi_3)$.

Recall, in the definition for the density of the marks (Eq. 3), the shape parameter ξ is the only constant parameter. This feature of the GDP is harnessed within the statistic proposed by Reiss and Thomas (2007) in order to determine the optimal τ quantile. The goal is to find the sample fraction k where the distribution of the shape parameter ξ in the GPD seems to be most stable in

[1]Other functional forms of returns are possible to capture the intraday behavior, for instance, squared returns. However, according to preliminary results, the choice of absolute values is preferred.

[2]We tried higher orders, but little is gained in the results observed.

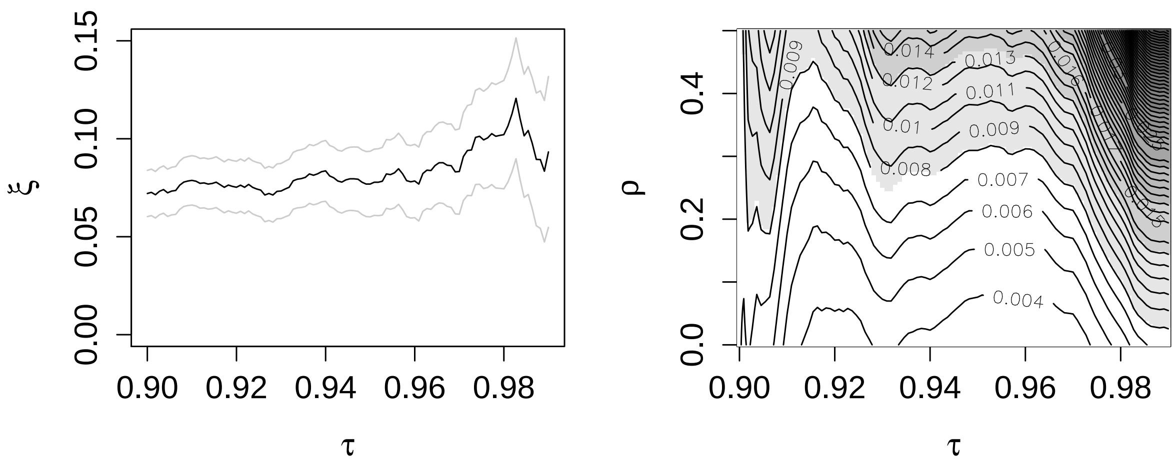

Figure 1: Threshold selection for the BHP time series returns of losses. The left plot shows estimates of the GPD shape parameter (ξ) at different thresholds (quantiles). The right plot shows the statistic proposed by Reiss and Thomas (2007). The gray rectangle displays the subset that seem to be the most stable for a given shape parameter (x-axis) and different tuning parameters ρ (y-axis).

relation to the statistic:

$$\arg\min_k f(k) = \frac{1}{k}\sum_{i=1}^{k} i^{\rho}\left|\hat{\xi}_i - \text{median}\left(\hat{\xi}_1, \ldots, \hat{\xi}_k\right)\right|, \tag{7}$$

where $\hat{\xi}_i$ is a shape parameter estimate for the sample fraction of the extremes above the upper order statistic i, and $\rho \in [0, 0.5]$ is a tuning parameter, and k is related to τ through $\tau = 1 - k/n$, with n being the number of observations of the returns $\{X_t\}_{t\geq 1}$. Figure 1 displays the results for the threshold selection for the BHP time series of losses. The left plot shows the estimates of the shape parameter at different thresholds (quantiles). The estimates seem to be stable across different thresholds from 0.90 to 0.95 and at around 0.96, $\hat{\xi}_i$ begins to increase. The contours in the right panel of Figure 1 represent estimates of the statistic proposed in (7) given selected values for the quantile τ and the tuning parameter ρ. The contours show that around $\tau = 0.95$, the estimate of $\hat{\xi}_i$ is the most stable in relation to this statistic, irrespective of the value of the tuning parameter. Therefore, a quantile of 0.95 is chosen so that 5% of the sample are selected as extreme losses, which corresponding to 8400 observations. This methodology provides, at a fixed quantile τ, a time varying threshold that accounts for the seasonality inherent in intraday financial data.

3.3 Forecasting Value at Risk

This section outlines how to generate high frequency forecasts of VaR for the 1-minute returns, taking into account the covariates. One can demonstrate that for Hawkes-POT models, the probability that the next return X_t will exceed the VaR at the α confidence level is a solution to the tail distribution $1 - F_{X_t|\mathcal{H}_t}\left(\mathrm{VaR}_\alpha^t\right) := \mathbb{P}\left(X_t > \mathrm{VaR}_\alpha^t \mid \mathcal{H}_t\right) = 1 - \alpha$, which can be decomposed into two components:

$$1 - F_{X_t|\mathcal{H}_t}\left(\mathrm{VaR}_\alpha^t\right) = \left\{1 - F_{X_t|\mathcal{H}_t}\left(u_t^\tau\right)\right\}\left\{1 - F_{X_t|\mathcal{H}_t}^{u_t^\tau}\left(\mathrm{VaR}_\alpha^t\right)\right\} \tag{8}$$

where $\left\{1 - F_{X_t|\mathcal{H}_t}\left(u_t^\tau\right)\right\} = 1 - \exp\left(-\int_{t_i}^t \lambda_g\left(s \mid \mathcal{H}_s\right) ds\right)$ which is related to the intensity of the events occurring over the threshold u_t^τ with t_i $(t_i < t)$ being the occurrence time of the last observed extreme event. On the other hand, $\left\{1 - F_{X_t|\mathcal{H}_t}^{u_t^\tau}\left(\mathrm{VaR}_\alpha^t\right)\right\} = \overline{G}\left(\mathrm{VaR}_\alpha^t - u_t^\tau \mid \mathcal{H}_t, t\right)$, corresponds to the conditional generalized Pareto survival function of the marks in Eq.(3). From this decomposition, the conditional asymptotic distribution of the log-returns can be obtained:

$$F_{X_t|\mathcal{H}_t}\left(\mathrm{VaR}_\alpha^t\right) = 1 - \frac{\left\{1 - \exp\left(-\int_{t_i}^t \lambda_g\left(s \mid \mathcal{H}_s\right) ds\right)\right\}}{\sigma_i}\left(1 + \xi\frac{\mathrm{VaR}_\alpha^t - u_t^\tau}{\sigma_i}\right)_+^{-1/\xi},$$

which leads to an estimate for the VaR at time t with confidence level α as follows:

$$\mathrm{VaR}_\alpha^t = u_t^\tau + \frac{\sigma_i}{\xi}\left\{\left(\frac{1 - \exp\left(-\int_{t_i}^t \lambda_g\left(s \mid \mathcal{H}_s\right) ds\right)}{1 - \alpha}\right)^{\xi} - 1\right\}. \tag{9}$$

4 Testing the Accuracy of Value of Risk Forecasts

To assess the accuracy of the conditional VaR forecasts, a range of tests of independence, unconditional and conditional coverage are applied, drawing on both long-standing techniques and recent developments in this area.

Let $\{I_t(\alpha)\}_{t=1}^n$ be a hit sequence 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. The first test is the unconditional coverage test (LR_{uc}) introduced by Kupiec (1995). This test only measures the distance between the nominal coverage rate α and the proportion of exceptions in the sample $\hat{\alpha}$ by means of the likelihood ratio:

$$LR_{uc} = -2\ln\left[L\left(\alpha; I_1, \ldots, I_n\right) / L\left(\hat{\alpha}; I_1, \ldots, I_n\right)\right] \sim \chi_1^2,$$

where $L\left(\alpha; I_1, \ldots, I_n\right) = \alpha^{n_1}\left(1 - \alpha\right)^{n_0}$ and $L\left(\hat{\alpha}; I_1, \ldots, I_n\right) = \hat{\alpha}^{n_1}\left(1 - \hat{\alpha}\right)^{n_0}$ are the likelihoods of a Binomial distribution with $\hat{\alpha} = n_1/(n_1+n_0)$, and n_1 being the number of exceptions and $n_0 = n - n_1$. The maximum likelihood estimator $\hat{\alpha}$ is the ratio of the number of violations to the total number of observations. This test implicitly assumes that the exceptions are independent.

The second test is a Markovian first-order test of independence of the exceptions (LR_{ind}), which tries explicitly to determine the independence of the exceptions based on a switching probability matrix:

$$\Pi = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix},$$

where π_{ij} and n_{ij} are the probability and number of observations with value j (0 or 1) on day t, given the value i on the previous day $t-1$. In this context, the probabilities can be defined as $\pi_{01} = \frac{n_{01}}{n_{00}+n_{01}}$ and $\pi_{11} = \frac{n_{11}}{n_{10}+n_{11}}$. The null hypothesis of interest is $H_0 : \pi_{01} = \pi_{11}$ (i.e., past hits do not contain information about current and future exceptions). By denoting $\pi = \pi_{01} = \pi_{11}$, a likelihood ratio test is given by

$$LR_{ind} = 2\ln\left[L\left(\tilde{\Pi}_1; I_1, \ldots, I_n\right)/L\left(\tilde{\Pi}_2; I_1, \ldots, I_n\right)\right] \sim \chi_1^2,$$

where

$$L\left(\tilde{\Pi}_1; I_1, \ldots, I_n\right) = (1-\pi_{01})^{n_{00}}\,\pi_{01}^{n_{01}}\,(1-\pi_{11})^{n_{10}}\,\pi_{11}^{n_{11}}$$

and

$$L\left(\tilde{\Pi}_2; I_1, \ldots, I_n\right) = (1-\pi)^{n_{00}+n_{10}}\,\pi^{n_{01}+n_{11}}.$$

The third test is the conditional coverage test (LR_{cc}), which is the result of a combination of the last two tests

$$LR_{cc} = LR_{uc} + LR_{ind} \sim \chi_2^2.$$

Engle and Manganelli (2004) proposed the dynamic quantile hit (DQ_{hit}) and dynamic quantile VaR (DQ_{VaR}) tests. To begin, define $Hit_t = I_t - \alpha$ as the de-meaned Hit exceptions. Engle and Manganelli (2004) test the null hypothesis that $\mathbf{Hit} = \{Hit_1, \ldots, Hit_p\}$ is orthogonal to the information contained in $\mathbf{X}$, which may include lags of Hit_t and VaR estimates amongst others. Under the null hypothesis, the proposed test statistic:

$$DQ = \frac{\mathbf{Hit}'\mathbf{X}\left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}\mathbf{Hit}'}{p\alpha\left(1-\alpha\right)},$$

follows a distribution χ_p^2, in which $p = \text{rank}(\mathbf{X})$, where the regressors should have no explanatory power. The DQ_{hit} denotes the test based on lagged demeaned hits of exceptions, while the DQ_{VaR} also includes VaR estimates as explanatory variables.

In a recent contribution to this literature, Ziggel et al. (2014) proposes a more flexible and powerful range of tests relative to the likelihood ratio tests described above. The advantage of these new tests is that all critical values for these tests are obtained utilizing Monte Carlo simulations and allow one- and two-tailed tests. For the unconditional coverage test they redefine the null hypothesis by

$H_0 : \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} I_t(\alpha)\right] = \alpha$ and propose to use the statistic

$$MCS_{uc} = \sum_{t=1}^{n} I_t(\alpha) + \epsilon,$$

where $\epsilon \sim N(0, 1e^{-6})$[3]. Comparing a large number of simulated realizations of this test statistic to the specified quantile permits a test of the unconditional coverage hypothesis. In the context of a test for independence, the Markov first-order test has very limited power against general forms of clustering in the exceptions, and therefore, different ways in which independence may be violated. Ziggel et al. (2014) state it is desirable that the exceptions should be independent and identically distributed. In particular, the null hypothesis of independence requires that $H_0 : \mathbb{E}[t_i - t_{i-1}] = \frac{1}{\alpha}$, i.e., waiting times between exceptions are geometrically distributed, motivating a squared duration based test better suited to detect exceptions which occur in clusters:

$$MCS_{iid,m} = t_1^2 + (n - t_m)^2 + \sum_{i=2}^{m} (t_i - t_{i-1})^2 + \epsilon,$$

where again $\epsilon \sim N(0, 1e^{-6})$. Here, m is the sum of observed VaR exceptions and $t_1, \ldots, t_m$ denote the times when VaR exceptions occur. The value of this statistic is minimised if the occurrence times are independent and identically distributed. In a similar fashion to the MCS_{uc} test, the critical values of the test statistic are obtained by simulating the finite sample null distribution.

The final tests are conditional coverage tests of the form:

$$MCS_{cc,m} = a \cdot f(MCS_{uc}) + (1 - a) \cdot g(MCS_{iid,m}),\ 0 \leq a \leq 1,$$

where $f(MCS_{uc}) = \left|\frac{MCS_{uc}/n - p}{p}\right|$ and $g(MCS_{iid,m}) = \frac{MCS_{iid,m} - \hat{r}}{\hat{r}} 1_{\{MCS_{iid,m} \geq \hat{r}\}}$ which measures the difference between the expected and observed proportions of exceptions, and sum of squared durations, respectively. The parameter a is a weighting factor that can be chosen arbitrarily according to the needs of a risk manager[4]. In the subsequent empirical applications, only results for $a = 0.5$ are presented. In the last term, $\hat{r}$ denotes an estimator of the expected value of the statistic $MCS_{iid,m}$ under the null hypothesis.

In the empirical application, both the MCS_{uc} and MCS_{cc} tests are applied to test separately whether a model is underestimating the VaR (i.e., an upper-tail test with $H_0 : \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} I_t(\alpha)\right] < \alpha$) or is too conservative (i.e., an lower-tail test with $H_0 : \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} I_t(\alpha)\right] > \alpha$) given the possible asymmetric consequences of these situations. In order to determine the critical values for these statistics, 10,000 Monte Carlo simulations are performed. For further details on these tests see Ziggel et al. (2014); Kuester et al. (2006); Engle and Manganelli (2004).

[3]The objective of this random variable ϵ with small variance is to help to break ties between test values.

[4]For instance, a risk manager could be more interested in the iid property of the exceptions during crisis period by choosing a lower level of a, instead a correct unconditional coverage.

Beyond testing whether the VaR forecasts satisfy certain optimality features (e.g., correct unconditional or conditional coverage), the predictive ability of the forecasts is also considered. To this end, the unconditional predictive ability tests (UPA) of Diebold and Mariano (1995) and the conditional predictive ability (CPA) tests of Giacomini and White (2006) are used, together with the smoothed quantile loss function:

$$\ell\left(X_t, VaR_\alpha^t\right) = \left(m_\delta\left(X_t, VaR_\alpha^t\right) - \alpha\right)\left(X_t - VaR_\alpha^t\right)$$

introduced by González-Rivera et al. (2004), where

$$m_\delta\left(X_t, VaR_\alpha^t\right) = \left[1 + \exp\left\{10^\delta\left(VaR_\alpha^t - X_t\right)\right\}\right]^{-1}$$

and δ is a parameter controlling the smoothness. Different values for the smoothness parameter $\delta = \{0.5, 2.5, 5\}$ are assumed. Selecting different δ parameters to control the smoothness permits different penalties to be applied to forecasts of VaR that are either too conservative or underestimate VaR [5]. Values of $\delta > 4$ make the smoothed quantile loss function indistinguishable from an the asymmetric loss function, where the VaR exceptions are more heavily penalized [6].

5 Empirical Results

A range of models are estimated containing different combinations of the covariates described in Section 2. Model 1 contains all the covariates while Model 8 contains none, only utilising the history of extreme events. Models 2 and 3 only contain one measure of trading activity, OF and AS respectively. The remaining models do not include the arrivals of new limit orders, NBV. Models 4 and 5 contain AS and OF respectively along with the liquidity measures BD and S. Models 6 and 7 utilise only the measures of trading, AS and OF respectively.

5.1 Model Selection

Table 3 reports the full sample estimation results for all of the model specifications considered. According to the AIC, Model 1 offers the best fit, while under BIC, the best approach is Model 3 that excludes the covariate AS. A difference in the asymptotic properties of the selection criteria will account for the different conclusions here. As the length of the sample increases, selection

[5] From a risk management perspective, using different values of δ reflects different levels of risk-aversion. For instance, $\delta = 0.5$ gives a weight around $(0.6 - \alpha)$ to the observations for which $X_t < VaR_\alpha^t$, while that $\delta = 2.5$ and $\delta = 5$ give a weight around $(0.8 - \alpha)$ and $(1 - \alpha)$, respectively.

[6] This is the check function introduced by Koenker and Bassett Jr (1978) and is defined as $\ell\left(X_t, VaR_\alpha^t\right) = \left(\mathbf{1}\left\{X_t > VaR_\alpha^t\right\} - \alpha\right)\left(X_t - VaR_\alpha^t\right)$, where $\mathbf{1}\left\{X_t > VaR_\alpha^t\right\}$ is an indicator function that takes the value 1 when the return observed is larger than the VaR estimated for the day t at the given confidence level α and the value 0 otherwise.

based on AIC becomes equivalent to selection based on the maximized log-likelihood, therefore, AIC will favor more complex models. In contrast, BIC tends to penalize complex models more heavily, being appropriate for model selection to avoid overfitting. Overall, it appears that both measures of trading activity, either OF and, or AS, along with order book conditions provide the best model fit. While Models 1 and 3 have been identified as offering superior fit, the VaR forecast performance of all models will be considered.

Overall, the parameters relating to the ground intensity processes are stable across the different models and are highly significant. In particular, the estimate of the rate μ is around 0.015 meaning that of the 5% of the most extreme events, 1.5% of these correspond to exogenous events. Recall, the parameter η provides a direct measure of the fraction of endogenous extreme events and thus a measure of market self-excitation. The results show that η is around 0.7 for the models estimated indicating a high degree of endogeneity. This implies that each new extreme event arriving at rate μ, generates $\eta/(1-\eta) \approx 2.3$ descendant extreme events. The decay parameter β is close to 0.006 suggesting that the exponential kernel in Eq. (2) exhibits slow decay. In relation to the impact function, the linear specification through the parameter b, is positive indicating that the larger the size of past extreme losses, the larger the intensity of new events.

In relation to the GPD for the size of the extreme events, the shape parameter ξ, which characterizes the tail of returns, is significant and of similar magnitude (in the range of 0.052 to 0.077) for all the models. These results are similar to those obtained at different threshold levels as shown in Figure 1. The covariates all play an important role in explaining the size of the extreme events through the specification of the scale parameter in Eq. (5). OF is found to have a significant negative impact on the size of the losses, the lower the OF in terms of a higher rate of selling leads to larger subsequent losses. A similar pattern is observed in relation to AS, the greater the degree of selling activity eliminating the best bid level in the order book, the larger the extreme losses. A greater rate of new limit bid orders, NBV, are associated with larger subsequent losses indicating that the participant offer an increase rate of liquidity supply during periods of extreme losses. The effect of both measures of liquidity are consistent with prior expectations. The greater depth on the bid side of the order book, BD, reduces the size of extreme events while increases in spread, S lead to larger losses.

5.2 Forecast accuracy

Table 4 reports the p-values of all tests discussed in Section 3.3. Overall, the proposed Hawkes-POT framework does an adequate job of forecasting VaR as there are no widespread rejections of the coverage and independence tests observed. However a number of interesting patterns emerge. Model 8 which excludes all covariates produces a number of rejections under both likelihood ratio coverage tests (LR_{uc} and LR_{cc}) at the higher quantiles, 0.99 and above. The one-tailed MCS_{uc} tests help distinguish whether these rejections are due to overestimation or underestimation of the VaR.

According to the MCS_{uc}^{lt}, these rejections are indeed due to the VaR forecasts being too conservative.

Models 6 and 7 (containing either OF or AS), and Models 4 and 5 (also containing BD and S) produce a similar pattern in terms of rejections of the coverage tests though the strength of the rejections in a number of cases fall. Model 1 containing all of the covariates produces only a small number of relatively marginal rejections of the coverage tests. On the whole, the performance of Models 2 and 3 are not dissimilar to Model 1. These results are broadly consistent with the in-sample estimation results showing where Models 1 and 3 provided superior fit, and indicate that the chosen set of liquidity and trading activity covariates are useful in terms of forecasting extreme risk in intraday returns. Fewer rejections of the single tailed tests are also observed moving from Model 8 to Models 1 and 3. Once again, rejections only occur in the MCS_{uc}^{lt}, and hence MCS_{uc}^{tt} tests. On the other hand, the DQ statistics are not rejected in any instances, indicating that all specifications produce forecasts that ensure there is no residual predictability in the exceptions. None of the simulation based MCS_{cc} tests are rejected.

Table 5 reports the results of both the unconditional and conditional predictive ability at different VaR confidence levels [7]. The null hypothesis for both tests is that the simplest Hawkes-POT specification (Model 8) is not inferior to the alternative Hawkes-POT models that include explanatory covariates (Models 1 through 7). The entries in the rows correspond to the UPA or CPA test statistics, while the numbers in parentheses contain the p-values of these tests.

The overwhelming result here, is that the null hypothesis, that the accuracy of Model 8 containing no covariates is not inferior to the other models containing covariates is rejected in the many of the cases, somewhat more frequently by the unconditional test than the conditional test. At the relatively low quantile of 0.96, the inferiority of Model 8 is not rejected indicating that covariates add little in terms of benefit for forecast accuracy at this low quantile. There is little to distinguish between Models 7 (only includes OF) given there are very few rejections other than at the extreme quantile. This pattern changes moving to the models that include more covariates as there are many rejections of the null hypothesis. Overall, Models 1 through 4 offer the most frequent improvements over Model 8. However, interestingly, the Hawkes-POT models including AS (Models 2 and 4) instead of OF (Model 1 and Model 3) as the covariates provide the best forecasts at VaR confidence levels of 0.97 and higher, independent of the smoothness degree of the loss function. It appears as though the rate of aggressive selling (AS) provides useful information for measuring the extreme risk in returns across a range of levels of confidence and smoothness of the loss function. In the case of $\delta = 5$, reflecting an asymmetric loss function where exceptions are penalized more heavily, there are few rejections except at the very extreme quantiles. This implies that if a risk manger is very concerned with exceptions over a very extreme quantile, models with covariates provides superior forecasts. However, at the 0.96 VaR confidence level, no gains are observed in terms of predictive ability by including explanatory covariates.

[7] A Newey-West heteroskedasticity and autocorrelation-consistent (HAC) estimator is used to obtain the covariance matrix for these statistics.

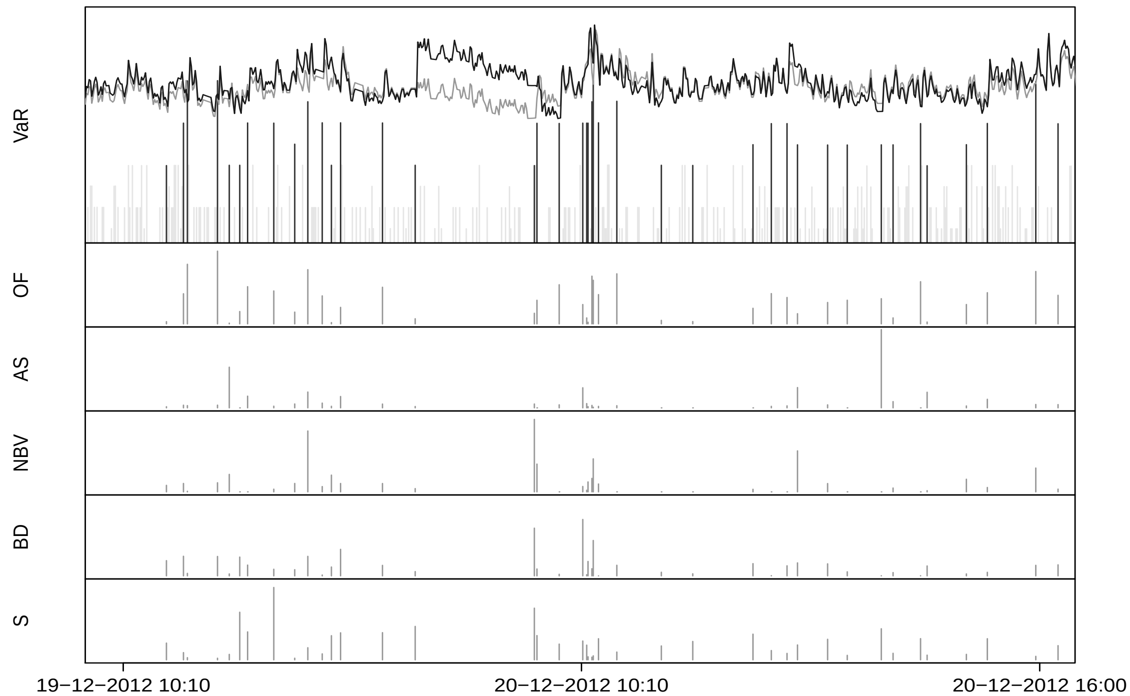

Figure 2: Top panel: 99% VaR estimates from Model 1, including all covariates (dark line) and Model 8 including no covariates (light line). Second Panel: Negative returns in light grey and extreme events observed during the backtesting period in black. The remaining panels show the observations of the covariates corresponding to the timing of the extreme events.

To give a more general view of the behaviour of the models and the resulting VaR estimates, Figure 2 plots the 99% VaR estimates from Models 8 and 1, along with the observed extreme events (in black color in the first panel), and all of the covariates observed at the times corresponding to the extreme events (for a two day period 19-20 December 2012). The top panel shows that the VaR predictions can either vary due to the arrival of extreme events or the behaviour of the covariates.

Overall, the VaR predictions from both Models 1 and 8 increase due to arrivals of extreme events as evident in the centre section of the plot corresponding to trading late on 19 December and early on 20 December. There are also periods where the VaR predictions from Models 1 and 8 diverge, highlighting the importance of the covariates. There is a period in the second half of 19 December where the VaR predictions from Model 1 rise substantially above those from Model 8. The difference in this period can be attributed to the larger spread S, observed at the time of an extreme event, with relatively low BD in the order book which significantly increases the VaR. At the beginning of the cluster of extreme events in the centre of the plot, BD is high to help offset the higher spread and hence the VaR estimates remain quite similar. In the subsequent period there is quite a number of extreme events driven by mainly higher trading intensity in terms of OF, with BD remaining high and S remaining low and hence the VaR predictions are similar.

6 Conclusion

Modeling and forecasting the risk of extreme financial losses is an important issue. This paper proposed a novel Hawkes-Peaks Over Threshold framework for dealing with the risk of extreme losses in an intraday setting. A new approach for a time-varying threshold, upon which extreme events are defined was proposed that accounts for the diurnal patterns observed in intraday volatility. An approach for including a range of covariates describing intraday trading activity and liquidity is also proposed. Here, covariates influence the scale of the distribution of the extreme events, and in turn the intensity which the extreme events occur.

The empirical exercise was based on a two year sample of high frequency BHP returns, and of the covariates considered, order flow reflecting signed trading intensity, best bid-price depth and spread reflecting liquidity, and the rate of new bid limit order arrivals are important for describing the dynamics extreme intraday losses, providing the best in-sample model fit. The importance of the covariates extend to the context of VaR prediction, where models containing these covariates produce the most accurate predictions, particularly at higher quantiles. It is shown that the VaR predictions can diverge significantly as a result of the variation in the trading covariates, clearly a desirable effect given the accuracy of the models that include the covariates.

References

Ahn, H., Bae, K., Chan, K., 2001. Limit orders, depth, and volatility: Evidence from the stock exchange of Hong Kong. The Journal of Finance 56 (2), 767–788.

Balkema, A. A., De Haan, L., 1974. Residual life time at great age. The Annals of probability, 792–804.

Berger, D., Chaboud, A., Hjalmarsson, E., 2009. What drives volatility persistence in the foreign exchange market. Journal of Financial Economics 94, 192–213.

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

Chavez-Demoulin, V., Embrechts, P., Hofert, M., 2015. An extreme value approach for modeling operational risk losses depending on covariates. Journal of Risk and Insurance.

Chavez-Demoulin, V., Embrechts, P., Sardy, S., 2014. Extreme-quantile tracking for financial time series. Journal of Econometrics 181 (1), 44–52.

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

Chordia, T., Roll, R., Subrahmanyam, A., 2001. Market liquidity and trading activity. Journal of Finance 56, 501–530.

Chordia, T., Roll, R., Subrahmanyam, A., 2002. Order imbalance, liquidity, and market returns. Journal of Financial Economics 65, 111–130.

Clements, A., Herrera, R., Hurn, A., 2015. Modelling interregional links in electricity price spikes. Energy Economics 51, 383 – 393.

Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics 13 (3), 253–265.

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

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

Evans, M., Lyons, R., 2002. Order flow and exchnage rate dynamics. Journal of Political Economy 110, 170–180.

Giacomini, R., White, H., 2006. Tests of conditional predictive ability. Econometrica 74 (6), 1545–1578.

González-Rivera, G., Lee, T.-H., Mishra, S., 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.

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

Groß-Klußmann, A., Hautsch, N., 2011. When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. Journal of Empirical Finance 18, 321–340.

Hall, A., Hautsch, N., 2006. Order aggressiveness and order book dynamics. Empirical Economics 30, 973–1005.

Hall, A., Hautsch, N., 2007. Modelling the buy and sell intensity in a limit order book market. Journal of Financial Markets 10, 249–286.

Hasbrouck, J., Seppi, D., 2001. Common factors in prices, order flows and liquidity. Journal of Financial Economics 59, 383–411.

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

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

Herrera, R., Schipp, B., 2014. Statistics of extreme events in risk management: The impact of the subprime and global financial crisis on the german stock market. The North American Journal of Economics and Finance 29, 218 – 238.

Koenker, R., Bassett Jr, G., 1978. Regression quantiles. Econometrica, 33–50.

Koenker, R., Xiao, Z., 2006. Quantile autoregression. Journal of the American Statistical Association 101 (475), 980–990.

Kuester, K., Mittnik, S., Paolella, M. S., 2006. Value-at-risk prediction: A comparison of alternative strategies. Journal of Financial Econometrics 4 (1), 53–89.

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

Liu, S., Tse, Y., 2015. Intraday value-at-risk: An asymmetric autoregressive conditional duration approach. Journal of Econometrics 189 (2), 437 – 446.

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

Næs, R., Skjeltorp, J. A., 2006. Order book characteristics and the volume-volatility relation: Empirical evidence from a limit order market. Journal of Financial Markets 9, 408–432.

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.

Opschoor, A., Taylor, N., van der Wel, M., van Dijk, D., 2014. Order flow and volatility: An empirical investigation. Journal of Empirical Finance.

Pickands III, J., 1975. Statistical inference using extreme order statistics. the Annals of Statistics, 119–131.

Reiss, R.-D., Thomas, M., 2007. Statistical analysis of extreme values: with applications to insurance, finance, hydrology and other fields. Birkhäuser Basel.

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

	OF	AS	NBV	BD	S
Model 1	•	•	•	•	•
Model 2		•	•	•	•
Model 3	•		•	•	•
Model 4		•		•	•
Model 5	•			•	•
Model 6		•			
Model 7	•				
Model 8					

Table 2: Covariates included in each model specification.

	Ground process				Generalized Pareto Distribution									
	μ	η	α	$b(\times 10e3)$	ξ	κ_0	$\kappa_1(OF)$	$\kappa_2(AS)$	$\kappa_3(NBV)$	$\kappa_4(BD)$	$\kappa_5(S)$	Log-like	AIC	BIC
Model 1	0.015	0.704	0.006	2.445	0.052	-8.369	-0.081	0.041	0.068	-0.038	15.705	26624	-53227	-53150
	(0.001)	(0.026)	(0.001)	(0.638)	(0.011)	(0.033)	(0.015)	(0.016)	(0.017)	(0.011)	(2.022)			
Model 2	0.015	0.704	0.006	2.414	0.057	-8.368		0.088	0.012	-0.023	15.141	26610	-53201	-53130
	(0.001)	(0.026)	(0.001)	(0.627)	(0.011)	(0.033)		(0.014)	(0.013)	(0.011)	(2.029)			
Model 3	0.015	0.703	0.006	2.455	0.053	-8.362	-0.103		0.097	-0.042	16.070	26621	-53223	-53152
	(0.001)	(0.026)	(0.001)	(0.638)	(0.011)	(0.033)	(0.013)		(0.013)	(0.011)	(2.019)			
Model 4	0.015	0.704	0.006	2.419	0.057	-8.364		0.093		-0.022	15.062	26610	-53202	-53139
	(0.001)	(0.026)	(0.001)	(0.628)	(0.011)	(0.033)		(0.012)		(0.010)	(2.027)			
Model 5	0.015	0.700	0.006	2.440	0.061	-8.306	-0.058			-0.028	15.731	26593	-53168	-53104
	(0.001)	(0.026)	(0.001)	(0.620)	(0.011)	(0.032)	(0.011)			(0.011)	(2.033)			
Model 6	0.015	0.695	0.006	2.395	0.067	-8.193		0.096				26580	-53145	-53096
	(0.001)	(0.026)	(0.001)	(0.607)	(0.011)	(0.018)		(0.012)						
Model 7	0.015	0.691	0.006	2.389	0.073	-8.131	-0.053					26559	-53104	-53055
	(0.001)	(0.026)	(0.001)	(0.589)	(0.012)	(0.016)	(0.011)							
Model 8	0.015	0.691	0.006	2.389	0.077	-8.131						26547	-53082	-53040
	(0.001)	(0.026)	(0.001)	(0.591)	(0.012)	(0.016)								

Table 3: Estimation results of the Hawkes-POT models for the 1-min log-returns of the BHP stock market data. The sample period is from 10:10 hrs. AM, 2 January 2011 to 15:59 hrs. PM, 30 November 2012. The covariates Z_k correspond to: Order Flow (OF), Aggressive Bid (AB), Aggressive Sell (AS), New Bid Volume (NBV), New Ask Volume (NAV), Bid Depth (BD), Ask Depth (AD), and Spread (S). Standard deviations are shown in parentheses. Log-like are the results of the maximization of the log-likelihood estimation, while AIC and BIC are the Akaike and Bayesian Information Criterion, respectively.

	α-level	Exc.	LR_{uc}	LR_{ind}	LR_{cc}	DQ_{hit}	DQ_{VaR}	MCS_{uc}^{ut}	MCS_{uc}^{lt}	MCS_{uc}^{tt}	MCS_{iid}	MCS_{cc}^{ut}	MCS_{cc}^{lt}	MCS_{cc}^{tt}
Model 1	0.96	218	0.57	0.98	0.85	0.98	0.75	0.28	0.72	0.56	0.06	0.06	0.94	0.11
	0.97	155	0.84	0.51	0.79	0.52	0.39	0.57	0.43	0.85	0.29	0.30	0.70	0.59
	0.98	97	0.43	0.88	0.72	0.88	0.94	0.79	0.21	0.41	0.11	0.10	0.90	0.21
	0.99	38	**0.03**	0.28	0.06	0.28	0.36	0.98	**0.02**	**0.03**	0.69	0.70	0.30	0.60
	0.995	18	0.09	0.72	0.22	0.73	0.91	0.95	0.05	0.10	0.77	0.78	0.22	0.45
	0.999	1	**0.02**	0.98	0.07	0.98	0.40	0.98	**0.02**	0.05	0.95	0.95	0.05	0.09
Model 2	0.96	218	0.57	0.98	0.85	0.98	0.66	0.27	0.73	0.54	0.08	0.09	0.91	0.17
	0.97	157	0.97	0.55	0.84	0.56	0.48	0.51	0.49	0.97	0.34	0.34	0.66	0.68
	0.98	94	0.27	0.35	0.35	0.36	0.63	0.87	0.13	0.26	0.15	0.15	0.85	0.30
	0.99	34	**0.01**	0.22	**0.01**	0.22	0.45	1.00	**0.00**	**0.00**	0.26	0.27	0.73	0.53
	0.995	18	0.09	0.72	0.22	0.73	0.86	0.96	**0.04**	0.07	0.32	0.33	0.67	0.65
	0.999	1	**0.02**	0.98	0.07	0.98	0.23	0.98	**0.02**	0.05	0.95	0.95	0.05	0.09
Model 3	0.96	219	0.53	0.96	0.82	0.96	0.82	0.25	0.75	0.50	0.06	0.06	0.94	0.11
	0.97	157	0.97	0.55	0.84	0.56	0.61	0.51	0.49	0.99	0.23	0.24	0.76	0.47
	0.98	97	0.43	0.88	0.72	0.88	0.87	0.77	0.23	0.46	0.11	0.11	0.89	0.22
	0.99	36	**0.02**	0.25	**0.03**	0.25	0.42	0.99	**0.01**	**0.02**	0.66	0.67	0.33	0.66
	0.995	20	0.20	0.70	0.41	0.70	0.66	0.88	0.12	0.24	0.77	0.76	0.24	0.47
	0.999	1	**0.02**	0.98	0.07	0.98	0.46	0.98	**0.02**	0.05	0.95	0.95	0.05	0.10
Model 4	0.96	218	0.57	0.98	0.85	0.98	0.66	0.30	0.70	0.59	0.08	0.08	0.92	0.16
	0.97	157	0.97	0.55	0.84	0.56	0.48	0.51	0.49	0.98	0.34	0.34	0.66	0.69
	0.98	93	0.23	0.79	0.47	0.79	0.89	0.87	0.13	0.25	0.17	0.17	0.83	0.34
	0.99	34	**0.01**	0.22	**0.01**	0.22	0.45	1.00	**0.00**	**0.01**	0.25	0.26	0.74	0.52
	0.995	18	0.09	0.72	0.22	0.73	0.84	0.96	**0.04**	0.08	0.32	0.32	0.68	0.64
	0.999	1	**0.02**	0.98	0.07	0.98	0.25	0.98	**0.02**	**0.04**	0.96	0.95	0.05	0.09
Model 5	0.96	220	0.48	0.79	0.75	0.80	0.86	0.24	0.76	0.49	0.05	0.06	0.94	0.11
	0.97	155	0.84	0.51	0.79	0.52	0.34	0.56	0.44	0.88	0.29	0.29	0.71	0.58
	0.98	96	0.37	0.85	0.66	0.86	0.94	0.81	0.19	0.39	0.14	0.14	0.86	0.29
	0.99	33	**0.00**	0.52	**0.01**	0.52	0.81	1.00	**0.00**	**0.00**	0.33	0.33	0.67	0.66
	0.995	17	0.05	0.74	0.15	0.74	0.83	0.97	**0.03**	0.05	0.35	0.37	0.63	0.73
	0.999	1	**0.02**	0.98	0.07	0.98	0.54	0.98	**0.02**	**0.04**	0.95	0.95	0.05	0.09
Model 6	0.96	218	0.57	0.98	0.85	0.98	0.55	0.28	0.72	0.56	0.07	0.06	0.94	0.13
	0.97	155	0.84	0.51	0.79	0.52	0.39	0.57	0.43	0.86	0.29	0.29	0.71	0.58
	0.98	93	0.23	0.33	0.30	0.34	0.55	0.88	0.12	0.23	0.18	0.17	0.83	0.33
	0.99	31	**0.00**	0.54	**0.00**	0.55	0.83	1.00	**0.00**	**0.00**	0.69	0.70	0.30	0.60
	0.995	17	0.05	0.74	0.15	0.74	0.84	0.97	**0.03**	0.05	0.37	0.36	0.64	0.73
	0.999	1	**0.02**	0.98	0.07	0.98	0.17	0.98	**0.02**	**0.04**	0.95	0.95	0.05	0.09
Model 7	0.96	218	0.57	0.98	0.85	0.98	0.71	0.29	0.71	0.59	0.06	0.06	0.94	0.12
	0.97	154	0.78	0.49	0.76	0.50	0.31	0.62	0.38	0.77	0.22	0.23	0.77	0.46
	0.98	96	0.37	0.38	0.46	0.39	0.58	0.80	0.20	0.40	0.11	0.12	0.88	0.23
	0.99	31	**0.00**	0.54	**0.00**	0.55	0.69	1.00	**0.00**	**0.00**	0.66	0.65	0.35	0.69
	0.995	15	**0.02**	0.77	0.05	0.77	0.95	0.99	**0.01**	**0.01**	0.46	0.46	0.54	0.92
	0.999	1	**0.02**	0.98	0.07	0.98	0.42	0.97	**0.03**	0.06	0.95	0.95	0.05	0.09
Model 8	0.96	218	0.57	0.75	0.81	0.75	0.67	0.28	0.72	0.56	0.07	0.07	0.93	0.14
	0.97	154	0.78	0.49	0.76	0.50	0.34	0.60	0.40	0.79	0.22	0.22	0.78	0.43
	0.98	95	0.32	0.37	0.40	0.37	0.64	0.84	0.16	0.33	0.10	0.10	0.90	0.19
	0.99	30	**0.00**	0.56	**0.00**	0.56	0.60	1.00	**0.00**	**0.00**	0.65	0.66	0.34	0.68
	0.995	15	**0.02**	0.77	0.05	0.77	0.82	0.99	**0.01**	**0.02**	0.48	0.48	0.52	0.97
	0.999	1	**0.02**	0.98	0.07	0.98	0.24	0.98	**0.02**	**0.04**	0.95	0.95	0.05	0.09

Table 4: Backtesting accuracy test results for the Hawkes-POT models. The sample used for backtesting covers the period from 10:10 hrs. AM, 3 December 2012 to 15:59 hrs. PM, 21 December 2012 (5950 observations). Entries in the columns are the significance levels (p-values) of the respective tests, with the exception of the confidence level α for the VaR, and the number of VaR exceptions observed (Exc.). MCS_{uc}^{ut}, MCS_{uc}^{lt} and MCS_{uc}^{tt} refer to the upper-tail, lower-tail and two-tailed Monte Carlo simulation unconditional coverage based tests, respectively. Similarly, MCS_{cc}^{ut}, MCS_{cc}^{lt} and MCS_{cc}^{tt} correspond to the upper-tail, lower-tail and two-tailed Monte Carlo simulation conditional coverage based tests. Rejected null hypotheses are highlighted in bold type. The rejection rates for the Monte Carlo simulation tests were obtained through 10,000 samples of Bernoulli simulated VaR-exception sequences.

α-level	Model 8 vs Model 7 $\delta=0.5$	$\delta=2.5$	$\delta=5$	Model 8 vs Model 6 $\delta=0.5$	$\delta=2.5$	$\delta=5$	Model 8 vs Model 5 $\delta=0.5$	$\delta=2.5$	$\delta=5$	Model 8 vs Model 4 $\delta=0.5$	$\delta=2.5$	$\delta=5$	Model 8 vs Model 3 $\delta=0.5$	$\delta=2.5$	$\delta=5$	Model 8 vs Model 2 $\delta=0.5$	$\delta=2.5$	$\delta=5$	Model 8 vs Model 1 $\delta=0.5$	$\delta=2.5$	$\delta=5$
Unconditional Predictive Ability Tests (Diebold and Mariano)																					
0.96	-6.89	-6.70	0.19	0.16	-0.03	1.16	-6.93	-6.96	0.36	-1.42	-1.64	1.11	-12.14	-12.40	0.60	-2.16	-2.42	1.13	-8.99	-9.24	0.84
	(1.00)	(1.00)	(0.42)	(0.44)	(0.51)	(0.12)	(1.00)	(1.00)	(0.36)	(0.92)	(0.95)	(0.13)	(1.00)	(1.00)	(0.27)	(0.98)	(0.99)	(0.13)	(1.00)	(1.00)	(0.20)
0.97	-7.40	-7.34	-1.37	2.42	2.43	1.20	7.27	7.09	0.46	10.47	10.26	1.79	9.83	9.45	1.07	11.58	11.31	1.92	8.82	8.55	1.30
	(1.00)	(1.00)	(0.91)	**(0.01)**	**(0.01)**	(0.12)	**(0.00)**	**(0.00)**	(0.32)	**(0.00)**	**(0.00)**	**(0.04)**	**(0.00)**	**(0.00)**	(0.14)	**(0.00)**	**(0.00)**	**(0.03)**	**(0.00)**	**(0.00)**	(0.10)
0.98	-0.06	0.25	-1.25	3.37	3.67	1.30	26.16	25.95	0.51	20.37	20.47	2.08	20.26	19.72	-0.06	21.81	21.73	2.12	20.91	20.62	0.76
	(0.53)	(0.40)	(0.89)	**(0.00)**	**(0.00)**	(0.10)	**(0.00)**	**(0.00)**	(0.31)	**(0.00)**	**(0.00)**	**(0.02)**	**(0.00)**	**(0.00)**	(0.53)	**(0.00)**	**(0.00)**	**(0.02)**	**(0.00)**	**(0.00)**	(0.22)
0.99	5.38	6.02	-0.39	5.37	6.12	0.98	39.29	39.00	1.31	28.11	28.79	1.95	23.66	24.82	0.62	29.43	29.85	1.98	27.55	28.09	1.10
	(0.00)	**(0.00)**	(0.65)	**(0.00)**	**(0.00)**	(0.16)	**(0.00)**	**(0.00)**	(0.09)	**(0.00)**	**(0.00)**	**(0.03)**	**(0.00)**	**(0.00)**	(0.27)	**(0.00)**	**(0.00)**	**(0.02)**	**(0.00)**	**(0.00)**	(0.14)
0.995	8.46	9.41	-0.09	7.73	9.17	0.67	47.37	46.98	1.63	33.97	35.47	1.50	26.74	30.42	0.51	35.24	36.35	1.46	32.33	34.21	0.90
	(0.00)	**(0.00)**	(0.54)	**(0.00)**	**(0.00)**	(0.25)	**(0.00)**	**(0.00)**	(0.05)	**(0.00)**	**(0.00)**	(0.07)	**(0.00)**	**(0.00)**	(0.30)	**(0.00)**	**(0.00)**	(0.07)	**(0.00)**	**(0.00)**	(0.18)
0.999	13.88	15.73	13.82	13.59	17.84	0.60	62.59	61.18	6.84	46.52	50.72	2.48	34.33	48.39	3.66	47.82	51.27	2.44	42.84	49.82	3.28
	(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	(0.27)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.01)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.01)**	**(0.00)**	**(0.00)**	**(0.00)**
Conditional Predictive Ability Tests (Giacomini and White)																					
0.96	-2.22	-2.47	0.19	0.06	-0.01	1.17	-2.13	-2.36	0.36	-0.51	-0.67	1.05	-3.60	-4.15	0.62	-0.75	-0.97	1.06	-2.88	-3.34	0.86
	(0.99)	(0.99)	(0.43)	(0.48)	(0.50)	(0.12)	(0.98)	(0.99)	(0.36)	(0.69)	(0.75)	(0.15)	(1.00)	(1.00)	(0.27)	(0.77)	(0.83)	(0.14)	(1.00)	(1.00)	(0.20)
0.97	-1.63	-1.70	-1.37	0.62	0.68	1.21	1.87	1.95	0.46	2.79	3.01	1.80	2.35	2.44	1.06	3.07	3.31	1.93	2.11	2.22	1.30
	(0.95)	(0.96)	(0.91)	(0.27)	(0.25)	(0.11)	**(0.03)**	**(0.03)**	(0.32)	**(0.00)**	**(0.00)**	**(0.04)**	**(0.01)**	**(0.01)**	(0.14)	**(0.00)**	**(0.00)**	**(0.03)**	**(0.02)**	**(0.01)**	(0.10)
0.98	-0.01	0.05	-1.26	0.66	0.78	1.32	5.88	6.17	0.50	4.49	4.90	2.08	3.39	3.52	-0.06	4.80	5.21	2.12	3.88	4.11	0.76
	(0.50)	(0.48)	(0.90)	(0.25)	(0.22)	(0.09)	**(0.00)**	**(0.00)**	(0.31)	**(0.00)**	**(0.00)**	**(0.02)**	**(0.00)**	**(0.00)**	(0.52)	**(0.00)**	**(0.00)**	**(0.02)**	**(0.00)**	**(0.00)**	(0.22)
0.99	0.83	0.98	-0.39	0.89	1.07	0.98	7.49	7.91	1.32	5.26	5.72	1.95	3.27	3.67	0.61	5.43	5.85	1.98	4.20	4.56	1.10
	(0.20)	(0.16)	(0.65)	(0.19)	(0.14)	(0.16)	**(0.00)**	**(0.00)**	(0.09)	**(0.00)**	**(0.00)**	**(0.03)**	**(0.00)**	**(0.00)**	(0.27)	**(0.00)**	**(0.00)**	**(0.02)**	**(0.00)**	**(0.00)**	(0.14)
0.995	1.23	1.45	-0.09	1.25	1.58	0.67	8.45	8.98	1.63	6.13	6.83	1.50	3.57	4.41	0.51	6.26	6.87	1.46	4.71	5.37	0.90
	(0.11)	(0.07)	(0.54)	(0.11)	(0.06)	(0.25)	**(0.00)**	**(0.00)**	(0.05)	**(0.00)**	**(0.00)**	(0.07)	**(0.00)**	**(0.00)**	(0.30)	**(0.00)**	**(0.00)**	(0.07)	**(0.00)**	**(0.00)**	(0.18)
0.999	1.93	2.39	2.09	2.14	3.07	0.60	10.56	11.32	6.77	8.13	9.61	2.48	4.60	7.53	3.62	8.21	9.55	2.44	6.02	8.00	3.27
	(0.03)	**(0.01)**	**(0.02)**	**(0.02)**	**(0.00)**	(0.28)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.01)**	**(0.00)**	**(0.00)**	(0.00)	**(0.00)**	**(0.00)**	**(0.01)**	**(0.00)**	**(0.00)**	**(0.00)**

Table 5: Results of pairwise tests of equal unconditional and conditional predictive ability for the Hawkes-POT models for the period from 10:10 hrs. AM, 3 December 2012 to 15:59 hrs. PM, 21 December 2012. The null hypothesis for both tests is that the simplest Hawkes-POT specification (Model 8) is not inferior to the alternative Hawkes-POT models including explanatory covariates. By selecting different δ parameters we control the smoothness and discriminate among different weights to penalize too conservatives or underestimated VaR estimates. The entries correspond to the statistics, while the numbers in parentheses contain the p-values of the tests. P-values in bold type indicates rejection of the null hypothesis