

Energy risk management through self-exciting marked point process

R. Herrera^{a,*}

^a*Departamento de Modelación y Gestión Industrial, Facultad de Ingeniería, Universidad de Talca, Camino Los Niches Km.1- Curicó. Chile*

Abstract

Crude oil is a dynamically traded commodity that affects many economies. We propose a collection of marked self-exciting point processes with dependent arrival rates for extreme events in oil markets and related risk measures. The models treat the time among extreme events in oil markets as a stochastic process. The main advantage of this approach is its capability to capture the short, medium and long-term behavior of extremes without involving an arbitrary stochastic volatility model or a prefiltration of the data, as is common in extreme value theory applications. We make use of the proposed model in order to obtain an improved estimate for the Value at Risk in oil markets. Empirical findings suggest that the reliability and stability of Value at Risk estimates improve as a result of finer modeling approach. This is supported by an empirical application in the representative West Texas Intermediate (WTI) and Brent crude oil markets.

JEL classification: C22; G15; G32; Q47

Keywords: Extreme value theory, energy market risk, energy forecasting, value at risk, marked self-exciting point process.

1. Introduction

In recent years oil industry has been continuously expanding and evolving with no sign of this changing in the near future. Price fluctuations in the crude oil markets worldwide have attracted significant attentions from both, managers and academics, due to the profound impact created on businesses and governments, and due to the high complexity and wide price swings in times of shortage or oversupply. In addition, oil futures are becoming an important investment instrument.

*Corresponding author

Email address: rodriherrera@utalca.cl (R. Herrera)

For example, oil futures became an important hedge during the Subprime mortgage crisis in 2007 to 2008. Oil is considered to be a hedge at such times as it goes against the trend of the stock market.

Good risk management begins with good risk framework; that is, understanding the likelihood and magnitude of extreme events. Unfortunately, due to the complicated unusual nature of oil price fluctuations, current energy risk management practice has remarkable difficulty assessing this class of events. By extreme events, we mean events that occur infrequently, even when looking broadly across the time. This limitation is unfortunate because infrequent, but potentially significant, events are precisely the kind of events that need the most attention and, with appropriate attention, they can also produce the greatest benefits. Proper measurement and management of risks due to unexpected oil price movements has been crucial from both, operational and strategic perspectives. An example of extreme events are the changes in crude oil prices which affect oil market traders, impact global economic activity, government policy and vice versa. For instance, the Persian Gulf War in 1991 is recognized as one of the most serious events that strongly affected crude oil markets in the 1990s.

Value at Risk (VaR) is the most popular and attractive method of measuring risk for simple concepts, and provides a single number that summarizes the total risk for financial assets. From the point of view of Extreme Value Theory (EVT), a few studies have considered VaR for crude oil markets (Krehbiel and Adkins, 2005; Marimoutou et al., 2009). Krehbiel and Adkins (2005) estimate tail parameters and construct risk statistics for unconditional distributions of daily logarithmic price changes of the NYMEX energy complex. They apply the conditional extreme value method to estimate VaR and related risk statistics from the tails of conditional distributions for these commodities. The results of this work indicate that, for the backtesting, the conditional extreme value approach is significantly more accurate for measuring the risk exposure of most of the series examined. Marimoutou et al. (2009) utilizes standard EVT to model VaR for long and short trading positions in the oil market by applying both unconditional and conditional EVT models to forecast VaR. They compare these models to the performances of other well-known modeling techniques, such as GARCH, Historical Simulation and Filtered Historical Simulation. The results of this work show that both conditional EVT and Filtered Historical Simulation procedures offer the greatest improvement over other conventional methods. Furthermore, their results confirm the importance of the filtering process for the success of standard approaches.

The peculiar properties of the extreme events related to the oil price fluctuations, such as the irregular spacing in time, the discreteness of price changes, the cluster of extremes as well as the presence of strong correlation among the time duration among extreme events, make new

econometric approaches, that take these issues into consideration necessary.

The contribution of this paper is two-fold. First, it explores dynamic EVT models that can capture the short-term dynamic of extreme events in crude oil markets without the use of an arbitrary stochastic volatility model, which certainly impacts the measures of risk, such as VaR. Taking the dynamical aspects within the cluster of extremes into account is absolutely necessary if one seeks to exploit the full amount of information in oil prices. The method to be introduced takes advantage of the structure of the model, thus allowing for more efficient use of the data. In this paper we will consider two alternative generalizations for the classical point process theory in EVT to the marked self-exciting point process (MSEPP) framework. A point process is defined as MSEPP, when the past evolution can impact the probability of future events.

The first class of models is formulated in relation to the time of occurrence of the extreme events (Chavez-Demoulin et al., 2005, Herrera and Schipp, 2009, and Chavez-Demoulin and McGill, 2012). We call this class of models time event peaks over threshold (TE-POT) and it would be able to generate power-law and exponential decay between extreme events and short-term cluster burst. The second class of models is the autoregressive conditional duration peaks over threshold (ACD-POT). It is focused on the intervals between extreme events, the inter-exceedance times (Herrera and Schipp 2012), and it is able to produce slow decay of autocorrelation and medium and long-term cluster burst.

The second contribution of this paper is the empirical application of the proposed models to two well known crude oil markets; the West Texas Intermediate (WTI), which trades on the New York Mercantile Exchange (NYMEX), and the Brent oil market, which is the leading global price benchmark for Atlantic basin crude oils and is used to price two thirds of the world's internationally traded crude oil supplies. Crude oil is not only the world's most actively traded commodity, but also the largest volume of futures trading of any physical commodity. Owing to its excellent liquidity and price transparency, crude oil contracts serve as a key international pricing benchmark. This is why estimating the risk in WTI and Brent crude oil is so important.

Our results show that the MSEPP models are stable and reliable, not only in-sample results but also in the backtesting, implying that these approaches of modeling extreme values can be used for further applications in energy markets. Moreover, major improvements in VaR forecasting are achieved in all aspects when accounting for the extreme event dynamics by means of the proposed models. In particular, VaR violation ratios are statistically equal to the theoretical values in all cases, and VaR violations are independent when using either the TE-POT model or the ACD-POT model, the latter being preferred overall.

The remainder of the paper is organized as follows. Section 2 offers a brief review of the

classical EVT approach. Section 3 introduces the new general MSEPP specification for extreme events. Furthermore, in this section we describe the two self-exciting approaches to be used in the empirical study. Section 4 contains the empirical applications. Section 5 concludes.

2. Peaks Over Threshold (POT) method

Suppose that we can observe the returns through the time of a crude oil market, as for example the Brent market. We denote these observations by Y_1, \dots, Y_n and we assume that the return series are independent and identically distributed (iid) random variables with common distribution function F .

Now, imagine that we observe this random distribution of observations $\{(t_i, y_i)\}$ whose values exceeds a high threshold u in a defined state space $\mathcal{T} \times \mathcal{Y} = [0, 1) \times [u, \infty)$, where the time has been rescaled for convenience to the interval $(0, 1)$. By looking at the dynamic of such observations, we concentrate only on the most extreme oil market returns.

On one hand, the time events t_i are the time of the i -th peak exceedance, i.e., the time in that a return exceeds a defined high threshold u . We refer to this process as the ground process. On the other hand, $y_i - u$ is the exceedance sizes or marks for a sufficiently high threshold u and we will call this process the process of the marks. A point process $N(A)$ can be viewed as the counter of these random observations in a set $A \subseteq \mathcal{T} \times \mathcal{Y}$. Pickands III, 1971 demonstrated that this two dimensional point process will look like as a non-homogeneous Poisson process with intensity defined for all subsets of the form $A = [t_1, t_2) \times [y, \infty)$ where t_1 and t_2 are times of occurrence of extreme events. This representation is as follows

$$\lambda(t, y) = \frac{1}{\sigma} \left(1 + \xi \frac{y - \mu}{\sigma} \right)_+^{-1/\xi - 1}, \quad (2.1)$$

where $y_+ = \max(y, 0)$. In this point process μ and σ determine the location and scale of the extremes, while ξ characterize the rate of decay of the tail of the distribution of extreme events. It follows from these characterizations that a complete summary of extremal behavior of this time series is contained in the three parameters μ, σ, ξ .

If we accept that the point process of exceedances is an one-dimensional Poisson, then the process has independent increments, i.e., the number of events t_i that occur in disjoint time intervals are mutually independent, which implies lack of memory in the evolution of the process. In addition, the number of extreme events t_i in any interval of length $(t_2 - t_1)$ is Poisson distributed

with mean

$$\Lambda([t_1, t_2] \times [y, \infty)) = \int_{t_1}^{t_2} \int_y^{\infty} \lambda(l, s) ds dl = \Lambda_1([t_1, t_2]) \times \Lambda_2([y, \infty)).$$

Notice that we have divided the intensity measure Λ into two independent Poisson process with corresponding intensity measure Λ_1 and Λ_2 . The first, Λ_1 , models the random time at which the extreme events occur, while the second, Λ_2 , models the exceedance sizes.

Another important result of extreme value theory is the following limiting conditional probability, which characterizes the tail of the excess distribution function over the threshold u .

$$\mathbb{P}(Y - u \leq y \mid Y > u) = \frac{\Lambda_2([y + u, \infty))}{\Lambda_2([u, \infty))} = \left(1 + \frac{\xi y}{\sigma + \xi(u - \mu)}\right)^{-1/\xi} = \bar{G}_{\xi, \beta}(y), \quad (2.2)$$

which is just the survival function of the generalized Pareto distribution (GPD), i.e., $\bar{G} = 1 - G$, with scaling parameter $\beta = \sigma + \xi(u - \mu)$ for $0 \leq y < y_F$. Here y_F is the right endpoint with values $y_F = \infty$ if $\xi > 0$ and $y_F = -\beta/\xi$ if $\xi < 0$. We shall call this model the Peaks Over Thresholds or POT model.

Observe that this methodology does not take into account the time when these extreme events occur because this assumes that the observations are independent, and therefore, that past observations do not have influence on future observations. In the following, we will explain our interest in investigating extreme events in oil markets as a marked point process of exceedances. Figure 2.1 shows in the top panel the negative daily percentage log-returns of Brent shares between June 1, 1987 and December 31, 2010, and the times and sizes of the negative daily percentage log returns exceeding a threshold $u = 3.15$ (the 0.93 quantile). Notice that this contradicts the classical model assumption of no cluster at the extremes. Indeed, under a homogeneous Poisson process the inter-exceedance times should be independent exponential random variables. The lower left picture shows an exponential probability plot for the inter-event times, these are clearly far from exponential, giving evidence against a Poisson process of exceedances. Furthermore, the auto-correlogram plot suggests clustering of the inter-exceedance times. This hypothesis is moreover reaffirmed by the Ljung-Box statistic using 10 lags. The null hypothesis of white noise is easily rejected with the Ljung-Box statistic of 143.43 well above the critical value of 18.307 at the 5% level, rejecting the Null hypothesis.

Summarizing, the characteristics of oil markets, such as clustered extremes and serial dependence, typically violate the assumptions of independence in the model. As a consequence, the direct application of the classical POT framework seems to be nonviable.

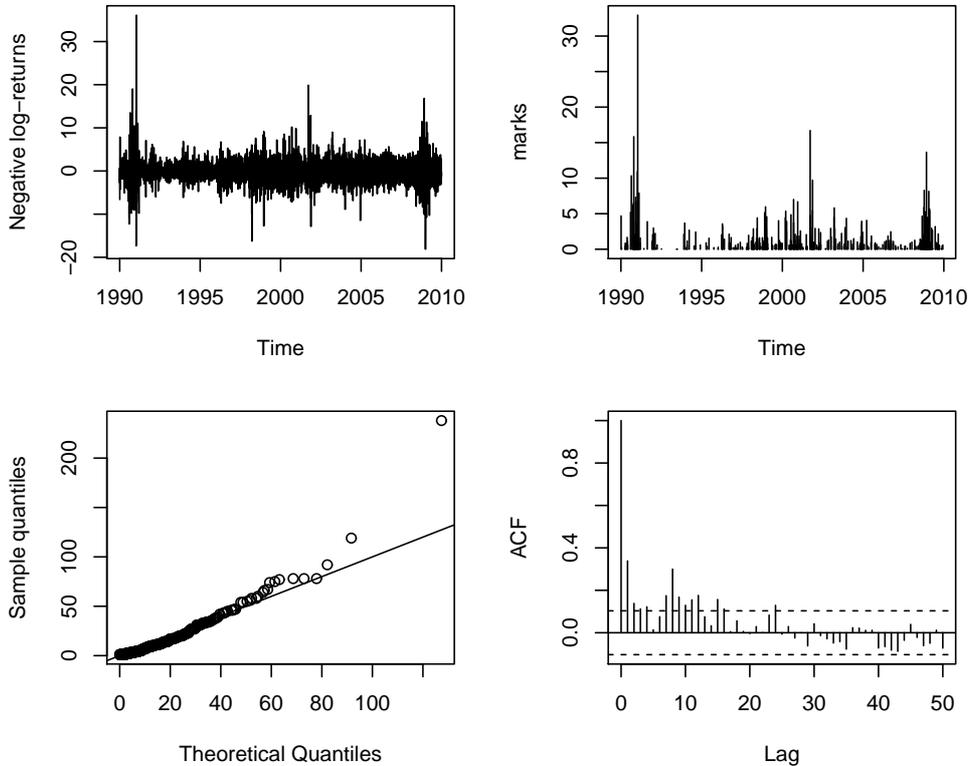


Figure 2.1: Upper left panel shows Brent daily percentage log-returns from 02.01.1990 to 18.01.2008. Upper right panel displays the 732 largest losses (the exceedances). Lower left panel shows a QQ-plot of inter-exceedance times against an exponential reference distribution. Finally, the lower right panel displays the autocorrelogram for the inter-exceedance times for the exceedances.

3. Extending the POT method to marked self-exciting point processes

In this paper we define a marked self-exciting point process (MSEPP) N , as a set of observations, occurrence times and marks $\{(t_i, y_i)\}$ on the space $\mathcal{T} \times \mathcal{Y}$, whose history $\mathcal{H}_t = (\{t_1, y_1\}, \dots, \{t_{t-1}, y_{t-1}\})$ consists only of the occurrence times and marks $\{t_1, y_1\}, \dots, \{t_{t-1}, y_{t-1}\}$ up to time t but not including t . In our case the time events t_i are the times in that the returns exceed a defined high threshold u , while $y_i - u$ is the size of the exceedances or marks for a sufficiently high threshold u . Observe that under this point of view, the marks arise as the component that carries the information about the events t in time and that may themselves have a stochastic structure and stochastic dependency relations, but they do not correspond to a second dimension (see Daley and Vere-Jones, 2003 for a more formal introduction).

As part of this point process, we define a ground point process N_g , which models the stochastic process of the inter-exceedance times with conditional intensity $\lambda_g(t | \mathcal{H}_t)$, and a generalized Pareto density function $f(y | \mathcal{H}_t, t)$ for the marks. This implies that the conditional expected

intensity for the marked point process N is given by

$$\lambda(t, y | \mathcal{H}_t) = \lambda_g(t | \mathcal{H}_t) f(y | \mathcal{H}_t, t). \quad (3.1)$$

Hereafter, we move away from Poisson models¹ for the occurrence times of exceedance of high thresholds and consider self-exciting models for the conditional intensity of the ground process. We focus on point processes which evolve with after-effects and which are conditionally orderly. In this sense, marked self-exciting point process constitutes a suitable framework because it allows the intensity of the ground process $\lambda_g(t | \mathcal{H}_t)$ to be modeled by a continuous function and a non-negative (stationary) random process and also allows past evolution to impact the probability structure of future events. Furthermore, it is particularly powerful for the modeling of multivariate processes. For instance, specifying in terms of an autoregressive process yields a dynamic intensity model which is particularly useful for capturing the clustering of extreme events in risk management.

The main idea in this class of models in extreme value theory is to replace the scale parameters β in (2.2) by dynamical alternatives based on time varying functions, e.g., $\beta(t, y | \mathcal{H}_t)$, so that the Poisson intensity parameters are now functions of the time t and the past history \mathcal{H}_t^2 .

As result of this approach, we now have a time-dependent intensity measure of the form

$$\lambda(t, y | \mathcal{H}_t) = \lambda_g(t | \mathcal{H}_t) f(y | \mathcal{H}_t, t) = \frac{\lambda_g(t | \mathcal{H}_t)}{\beta(t, y | \mathcal{H}_t)} \left(1 + \xi \frac{y - u}{\beta(t, y | \mathcal{H}_t)} \right)_+^{-1/\xi - 1}.$$

Notice that the distribution of the marks are assumed independent of the behavior of inter-exceedance times. Indeed, the implied distribution of the marks when an extreme event takes place is given by

$$\frac{\Lambda_2([u + y, \infty) | \mathcal{H}_t)}{\Lambda_2([u, \infty) | \mathcal{H}_t)} = \left(1 + \xi \frac{y - u}{\beta(t, y | \mathcal{H}_t)} \right)_+^{-1/\xi} = \bar{G}_{\xi, \beta(t, y | \mathcal{H}_t)}(y).$$

Note that the marginal distribution of the marks will now be conditionally independent of the associated ground process. Therefore, the product of mark densities simply has to be multiplied with the likelihood of the ground process. Let N be a MPP on $[t_0, T) \subseteq \mathcal{X} \times \mathcal{Y}$ for some finite positive T with realizations $(t_1, y_1), \dots, (t_{N(T)}, y_{N(T)})$ of N , and p_i be a family of conditional probability density functions for arrival time t_i , the log-likelihood L of such a realization in terms of

¹Observe that an alternative description of the Poisson process (2.1) is rewritten this as a special case of the MPP (3.1), with $\lambda_g(t | \mathcal{H}_t) = - \left(1 + \xi \frac{u - \mu}{\sigma} \right)_+^{-1/\xi}$ and $f(x | \mathcal{H}_t) = \frac{1}{\beta} \left(1 + \xi \frac{y - \mu}{\beta} \right)_+^{-1/\xi - 1}$, with $\beta = \sigma + \xi(u - \mu)$.

²We could also parametrize the shape parameter ξ . However, the behavior of the estimation is severely affected. For this reason it is reasonable to consider the shape parameter constant.

the conditional densities or intensities is given by

$$\begin{aligned}
L &= \sum_{i=1}^{N(T)} \log p_i(t_i | \mathcal{H}_t) + \sum_{i=1}^{N(T)} \log f_i(y_i | \mathcal{H}_t, t) \\
&= \sum_{i=1}^{N(T)} \log \lambda_g(t_i | \mathcal{H}_t) - \int_{t_0}^T \lambda_g(s | \mathcal{H}_t) ds + \sum_{i=1}^{N(T)} \log f_i(y_i | \mathcal{H}_t, t)
\end{aligned} \tag{3.2}$$

3.1. Dynamic risk measures

Risk management embodies the process and the tools used for evaluating, measuring and managing the various risks within a Company's portfolio of financial, commodity and other types of assets. In the case of oil markets, VaR can be used for instance to quantify the maximum oil price changes associated with a likelihood level, to avoid big losses due to price fluctuations or changing energy consumption patterns and to meet regulatory requirements that limit exposure to risk. For the MSEPP models the $\text{VaR}_\alpha^{t^*}$ is defined as the α -th quantile of a distribution at the time $t^* > t$ which is solution to $\mathbb{P}(y_{t^*} > y | \mathcal{H}_t) = 1 - \alpha$. Observe that under our framework this measure can be derived from the time-dependent intensity measure as follows

$$\begin{aligned}
\mathbb{P}(y_{t^*} > y | \mathcal{H}_t) &\approx \int_t^{t^*} \lambda(l, y | \mathcal{H}_t) dl \\
&= \lambda_g(t^* | \mathcal{H}_t) \times \Lambda_2([y, \infty)).
\end{aligned}$$

Thus, solving this equation for some value of $y \geq u$ the VaR is defined by

$$\text{VaR}_\alpha^{t^*} = u + \frac{\beta(t, y | \mathcal{H}_t)}{\xi} \left(\left(\frac{1 - \alpha}{\lambda_g(t^* | \mathcal{H}_t)} \right)^{-\xi} - 1 \right). \tag{3.3}$$

The last equation implies that the VaR is only defined for our models if $\lambda_g(t^* | \mathcal{H}_t) > 1 - \alpha$.

Novels MSEPP and applications to extreme value theory have been presented in Chavez-Demoulin et al. (2005), Herrera and Schipp (2009), Chavez-Demoulin and McGill (2012) and Herrera and Schipp (2012). In this paper we will consider two alternative generalizations for the classical point process methodology in extreme value theory. The first class of models is formulated in relation to the time of occurrence of the extreme events “the time event POT models”, while the second class of models is focused on the intervals between extreme events, the inter-exceedance times, “the autoregressive conditional duration POT models”.

3.2. Time event POT models (TE-POT)

This class of models is obtained by specifying, the intensity of the ground process $\lambda_g(t | \mathcal{H}_t)$ and the scale parameter $\beta(t, y | \mathcal{H}_t)$ as a (linear) self-exciting process given by

$$\begin{aligned}\lambda_g(t | \mathcal{H}_t) &= k + \phi \sum_{i:t_i < t} g(t - t_i, y_i) \\ \beta(t, y | \mathcal{H}_t) &= \beta_0 + \eta \sum_{i:t_i < t} g(t - t_i, y_i)\end{aligned}$$

where k and β_0 represent the background rate of events, which in most applications is assumed to be constant in time, while $\phi, \eta \geq 0$ and the kernel $g(\cdot)$ denote the magnitude of self-excitation and density at which self-excitation is triggered, respectively. Note that each time a new point arrives in this process, the conditional intensity (the time varying scale parameter) grows by ϕ (η) and then decreases exponentially back towards k (β). In other words, a point increases the chance of attaining other points immediately afterwards, and thus, this is the best model for clustered point patterns. Many forms for $g(\cdot)$ have been proposed in the literature, though in general the kernel is chosen so that the risk increases with extreme event magnitude and decreases in time away from each event. In this paper we use two different kernel functions.

$$g_H(t - t_i, y) = (1 + \delta y) \exp(-\gamma(t - t_i)) \quad \text{and} \quad g_E(t - t_i, y) = \frac{(1 + \delta y)}{\left(1 + \frac{t - t_i}{\gamma}\right)^{1+\rho}},$$

where $t - t_i$ denotes the time elapsed since an extreme event y has occurred at the time t . We refer to $g_H(t - t_i, y)$ as the Hawkes kernel and $g_E(t - t_i, y)$ ETAS kernel (Epidemic Type After Shock). For more on this class of models applied on risk management see Chavez-Demoulin et al. (2005), Herrera and Schipp (2009) and Chavez-Demoulin and McGill (2012). Observe that in the general formulation of these models, the marks are conditionally independent since the ground intensity does not depend on the past marks.

3.3. The Autoregressive conditional duration POT model (ACD-POT)

In the second alternative specification, the intensity is driven by an autoregressive process which is updated at each point of the process. These types of specifications were initially proposed by Herrera and Schipp (2012). This leads to a special type of point process model which does not originate from the classical point process literature but rather from the autoregressive conditional duration (ACD) literature (see Engle and Russell (1998)). We define a model for the conditional intensity of the ground point process of exceedances depending only on a fixed number of the most recent inter-exceedance times $x_i = t_i - t_{i-1}$. The ACD model is defined as follows

$$\begin{aligned}
x_i &= \psi_i \varepsilon_i \\
\psi_i &= \psi_i(x_{i-1}, \dots, x_1; \theta)
\end{aligned}$$

where $\psi_i(x_{i-1}, \dots, x_1; \theta)$ is the duration conditional on all information up to and including time t_{i-1} , θ is a parameter vector and ε_i are iid random variables. Two specifications are considered in this paper for the conditional expected duration. The first is the most popular autoregressive conditional duration model (ACD model), introduced by Engle and Russell (1998) and is based on a linear parameterization of the conditional mean function

$$\psi_i = w + \sum_{j=1}^p a_j x_{i-j} + \sum_{j=1}^q b_j \psi_{i-j},$$

where $w > 0$, $a, b \geq 0$. In order to ensure the stationarity and existence of the unconditional expected duration we need $\sum_{j=1}^p a_j + \sum_{j=1}^q b_j < 1$.

The second model is the logarithmic ACD (Log-ACD) model, introduced by Bauwens and Giot (2000) in order to prevent ψ_i becoming negative, in which the autoregression bears on the logarithm of the conditional expected duration³

$$\psi_i = \exp \left\{ w + \sum_{j=1}^p a_j \log x_{i-j} + \sum_{j=1}^q b_j \log \psi_{i-j} \right\}.$$

To find a general expression for the conditional intensity of the ground process $\lambda_g(x_i | \mathcal{H}_i; \theta)$, let f_ε and S_ε be the density function and the associated survival function of ε_i , respectively. One can easily show that the conditional expected intensity of the inter-exceedance times between extreme events, the ground process, can be expressed as a multiplicative effect between the baseline hazard function and a self-exciting point process shift given by the expected duration

$$\lambda_g(x_i | \mathcal{H}_i; \theta) = \lambda_0 \left(\frac{x_i}{\psi_i} \right) \frac{1}{\psi_i}. \quad (3.4)$$

where $\lambda_0(t) = f_\varepsilon(t) / S_\varepsilon(t)$ is defined as the baseline hazard function.

The second important ingredient in the parameterization of the ACD models is the distributional assumption for the innovation process. In this paper we propose the generalized gamma distribution⁴. The major advantage of this distribution is that this has a non-monotonic hazard

³For a meaningful comparison of alternatives in relation to TE-POT models and for simplicity, we limit the dynamic structure of the ACD-POT models to the first lag order only in the empirical study.

⁴During the first draft of this paper we take different ACD-models into account to find the best approach with

function taking bathtub shaped or inverted U-shaped form. In a bathtub shaped form the hazard rate initially decreases, during the middle phase the hazard rate is essentially constant, and in the final phase the hazard increases. Inverted U-shaped forms are the counterparts; the hazard rate initially increases, then becomes close to constant and ultimately decreases. This feature is of particular importance if we are interested in modeling risk measures such as the VaR or the expected shortfall.

Lunde (1999) as well Zhang et al. (2001) propose the use of a generalized gamma distribution to characterize the standardized durations because one can then obtain a non-monotonic hazard function and a time-varying conditional mean duration. A three parameter generalized gamma density is given by

$$f(x | \gamma, k) = \frac{\gamma x^{k\gamma-1}}{\lambda^{k\gamma} \Gamma(k)} \exp \left\{ - \left(\frac{x}{\lambda} \right)^\gamma \right\}, x > 0.$$

It includes the exponential distribution ($\gamma = k = 1$), the Weibull distribution ($k = 1$), the half-normal ($\gamma = 1/2, k = 1$) and the ordinary gamma distribution ($k = 1$). Under the restriction that $\lambda = 1$ we chose $\phi_i = \psi_i \frac{\Gamma(k)}{\Gamma(k + \frac{1}{\gamma})}$ which implies a conditional intensity for the ground process given by.

$$\lambda_g(x_i | \mathcal{H}_i; \theta) = \frac{\frac{\gamma x_i^{k\gamma-1}}{\phi_i^{k\gamma} \Gamma(k)} \exp \left\{ - \left(\frac{x_i}{\phi_i} \right)^\gamma \right\}}{I \left(k, \left(\frac{x_i}{\phi_i} \right)^\gamma \right)},$$

where $I \left(k, \left(\frac{x_i}{\phi_i} \right)^\gamma \right) = \int_{\left(\frac{x_i}{\phi_i} \right)^\gamma}^{\infty} u^{k-1} \exp(-u) du$ is the upper incomplete gamma integral. Note that if $k = 1$, then we get the Weibull-ACD model, while for $k = \gamma = 1$ the model reduces to an Exponential-ACD model.

In addition, for the scale parameter $\beta(t, y | \mathcal{H}_t)$ we consider a lineal parameterization such that it depends on the history.

$$\beta(t, y | \mathcal{H}_t) = \beta_0 + \beta_1 y_{i-1} + \beta_2 \psi_i$$

This feature implies that the marks are conditionally generalized Pareto distributed, given the history \mathcal{H}_t up to the time of the mark like the TE-POT models. These models assume that in a period of turmoil the temporal intensity of the inter-exceedance times and the magnitude of the marks linearly influence the scale parameter.

different distributional assumptions, as the Burr, Weibull or exponential for the standardized duration times. Our analysis with financial time series have suggested that by model comparisons based on the likelihood ratio statistic and AIC, the proposed models give the best fit keeping the formulation easy to understand.

4. Empirical analysis

One of the purposes of this work is to contribute to the literature on energy prices by studying the impact of extreme events on the determination of measures of risk for oil price returns. However, an important aspect for the implementation of the models proposed in the above section is the determination of the threshold beyond which the observations are assumed to follow a generalized Pareto distribution.

The issue of how to choose the threshold is similar to that of selecting the size of a block in classical EVT in the sense that both imply a balance between bias and variance. A low threshold u leads to failure in the asymptotic approximation of the model and a high threshold u provides few observations and then high variance. Actually, the choice of the optimal threshold is still considered an open problem and different approaches have been proposed to overcome this difficulty. For instance, Chavez-Demoulin et al. (2005) recommend choosing a threshold so that about between 5 and 10% of the data are excesses, while Herrera and Schipp (2012) propose a sensitivity analysis based on a mean squared error, to assess the stability of the VaR among different thresholds. In this paper we choose to work with the 7% of the maxima of the sample, the choice of the threshold is explained in detail in AppendixA.

In relation to the measures of goodness of fit in-sample we utilize the W-statistics to assess our success in modeling the temporal behavior of the exceedances of the threshold u . This statistic states that if the GPD parameter model is correct, then the residuals should approximately be independent unit exponential variables. In addition, to check that there is no further time series structure the autocorrelation function (ACF) for the residuals is also included. Similarly, to appraise the quality of the times component of our model, we employ the residual analysis for point process. All of these methods are resumed briefly in AppendixB.

4.1. Data set and summary statistics for WTI and Brent crude oil price returns

We use the daily closing price in both US West Intermediate Taxes (WTI) market and Europe Brent (Brent) market, which are two of the major marker markets in the world. The data source is the Energy Information Administration, Department of Energy, US. The sample period spans from 2 January 1990 to December 31, 2009. A second sample is used for backtesting the estimation of the VaR for the two markets from 4 January 2010 to August 22, 2011. In this study we only concentrate on the left tail, so that the daily returns are calculated as $r_t = -100\ln(p_t/p_{t-1})$, where p_t denotes the stock price at day t . In the backtest we daily update the new information that becomes available for the parameter estimates previously obtained. Thus, we dynamically adjust quantiles, which allows us to as accurately as possible improve the estimation of the risk measures.

Table 1 presents some relevant summary statistics regarding the unconditional distribution of the returns. The statistics show that all returns exhibit skewness to the losses as well as excess of kurtosis, hence we may deduce that each return has a leptokurtic distribution with a fat left tail. In other words, the returns analyzed here do not have the standard normal distribution. Verification is given by the results of Jarque Bera test. Serial correlation by means of a Ljung-Box test was not rejected for WTI and Brent returns with a statistical significance at the 1% and 5% level respectively. In addition, both returns are stationary series by means of ADF unit root test.

Indexes	mean	sd	min	max	skewness	kurtosis	Q(x)	Jarque-Bera	ADW
WTI	0.024	2.584	-40.639	18.867	-0.921	19.915	19.753*	61897.960*	-16.185*
Brent	0.025	2.436	-36.121	18.129	-0.749	18.120	11.647**	49357.420*	-15.827*

Table 1: Summary statistics for the stock market returns analysed. Asymptotic p-value are shown in the brackets. *,**,*** denote statistical significance at the 1, 5 and 10% level respectively. The Ljung-Box test statistic (Q) for serial correlation is calculated up to the 5-th order.

4.2. Model estimation and results

In order to adequately summarize the large quantity of empirical results obtained, we use the following classification scheme for the MSEPP models:

- TE-POTh: time event peaks over threshold model with Hawkes kernel.
- TE-POTe: time event peaks over threshold model with ETAS kernel.
- gACD-POT: ACD model for the expected conditional duration with generalized gamma distribution for the standardized residuals.
- gLogACD-POT: Log-ACD model for the expected conditional duration with generalized gamma distribution for the standardized residuals.

We have four models in total. Observe that we could obtain other submodels by restricting, for example, the scale parameter to be constant through the time.

Empirical results

The maximum log-likelihood estimates of the MSEPP models proposed in section 3 for the two markets are displayed in Table 2. The results lead to markedly favor the ACD-POT model approach. For the WTI market a gACD-POT model was estimated with AIC of 3662.91, while for the Brent market a gLog-ACD model was estimated with AIC of 3589.47. In addition, in both cases the time varying scale parameter leads to a better fit. Indeed, the results suggest that these models

WTI	Parameters of the conditional intensity $\lambda(t,y \mathcal{H}_t)$ of the MSEPP													Loglik	AIC
	Conditional intensity of the ground process $\lambda_g(t \mathcal{H}_t)$									Pareto generalized density $f(y \mathcal{H}_t, t)$					
	w	a_1	b_1	δ	γ	k	ϕ	η	ρ	ξ	β_0	β_1	β_2		
TE-POTe				0.215 (0.160)	2.449 (0.942)	0.022 (0.007)	0.050 (0.019)	0.612 (0.259)	0.161 (0.105)	0.286 (0.067)	0.616 (0.136)			-1842.643	3701.286
TE-POTh				0.345 (0.167)	0.014 (0.003)	0.035 (0.006)	0.014 (0.003)	0.196 (0.066)		0.267 (0.066)	0.729 (0.124)			-1829.61	3673.214
gACD	1.154 (0.458)	0.254 (0.058)	0.683 (0.061)		0.202 (0.160)	17.859 (28.011)				0.275 (0.144)	0.321 (0.067)	0.062 (0.052)	10.443 (2.121)	-1822.45	3662.91
gLog-ACD	0.301 (0.0952)	0.181 (0.033)	0.747 (0.052)		0.247 (0.116)	11.961 (11.108)				0.277 (0.0670)	0.260 (0.159)	0.059 (0.051)	11.406 (2.328)	-1823.35	3664.69
Brent	Parameters of the conditional intensity $\lambda(t,y \mathcal{H}_t)$ of the MSEPP													Loglik	AIC
	Conditional intensity of the ground process $\lambda_g(t \mathcal{H}_t)$									Pareto generalized density $f(y \mathcal{H}_t, t)$					
	w	a_1	b_1	δ	γ	k	ϕ	η	ρ	ξ	β_0	β_1	β_2		
TE-POTe				1.892 (1.132)	4.187 (1.597)	0.040 (0.007)	0.008 (0.004)	0.132 (0.081)	0.267 (0.168)	0.227 (0.069)	0.662 (0.122)	0.196 (0.662)		-1800.705	3617.41
TE-POTh				0.381 (0.151)	0.041 (0.008)	0.037 (0.005)	0.012 (0.009)	0.172 (0.098)		0.220 (0.074)	0.771 (0.113)			-1795.01	3604.02
gACD	0.9694 (0.4860)	0.1689 (0.045)	0.7743 (0.061)		0.1513 (0.0727)	31.9091 (30.506)				0.192 (0.066)	0.456 (0.143)	0.2007 (0.0694)	6.257 (1.816)	-1785.96	3589.92
gLog-ACD	0.310 (0.114)	0.150 (0.034)	0.767 (0.061)		0.1843 (0.0972)	21.623 (22.624)				0.186 (0.066)	0.416 (0.152)	0.1911 (0.0689)	6.976 (2.020)	-1785.74	3589.47

Table 2: Results of the estimation of all MSEPP models for the WTI and Brent market log-returns. Standard deviations are given in parentheses. Loglike are the results of the maximization of the log-likelihood estimation and AIC is the Akaike Information Criterion.

react more quickly to increasing and decreasing cluster frequency of extremes, which means that expected duration conditional of the inter-exceedances times has an effect on the probability of further exceedances in the near future. The β_2 values for all ACD-POT models are statistically significant. Interestingly, the size of the last exceedance represented by the coefficient β_1 is not as important as the expectation of the i -th inter-exceedance time.

We observe further that for the gamma distribution parameters we get $k\gamma > 1$ and $\gamma < 1$ for all ACD-POT fitted models. This means that for these models the conditional intensity is inverted U-shaped. This sort of flexibility in the shape of the conditional intensity was already noted to be relevant by Lunde (1999) and Zhang et al. (2001). The major difference between the TE-POT and the ACD-POT model results is that the latter allows a nonmonotonic conditional intensities, which is able to put a lot of probability mass on small durations but not too much probability mass on very small durations.

The results on the goodness of fit in-sample for the models fitted to the log-returns are displayed in Figures 4.1 and 4.2 for the WTI and Brent markets respectively.

Firstly, we assess the conditional GPD assumption of the marks in the models fitted. To this end, we provide the W-statistic explained in details in AppendixB. This statistic forms an iid sequence of exponential random variables with mean one if the marks are GPD. According to the QQ-plots displayed in Figure 4.1 for the WTI log-returns, we do not observe a substantial deviation from an exponential distribution in all cases. For instance the Kolmogorov-Smirnov test gives as result; $D = 0.0338$ with p-value 0.815 for the TE-POT_h model, $D = 0.0331$ with p-value 0.826 for the TE-POT_e model, $D = 0.0388$ with p-value 0.649 for the gACD-POT model, and $D = 0.038$ with p-value 0.677. In the case of the Brent log-returns, the results are very similar though some differences can be observed. The Kolmogorov-Smirnov test gives as result; $D = 0.0329$ with p-value 0.829 for the TE-POT_h model, $D = 0.078$ with p-value 0.023 for the TE-POT_e model, $D = 0.071$ with p-value 0.0506 for the gACD-POT model, and $D = 0.065$ with p-value 0.096. The above results indicates that for the TE-POT_e model the null hypothesis can be only accepted at the 1% significance level.

Furthermore, we would like to check that there is no further time series structure, for this reason the autocorrelation function (ACF) for the residuals (middle panel) is also included. The autocorrelations is negligible at nearly all lags for the WTI market. In the case of the Brent market we observe some differences, especially for the ACD-POT models. Examining the autocorrelation of the residuals formally by means of the Ljung-Box statistic under the null hypothesis that the first teen autocorrelations are zero, we found that the null hypothesis is not only rejected for the TE-POT models at the 5% significance level. In particular, the TE-POT_h model has a chi-squared

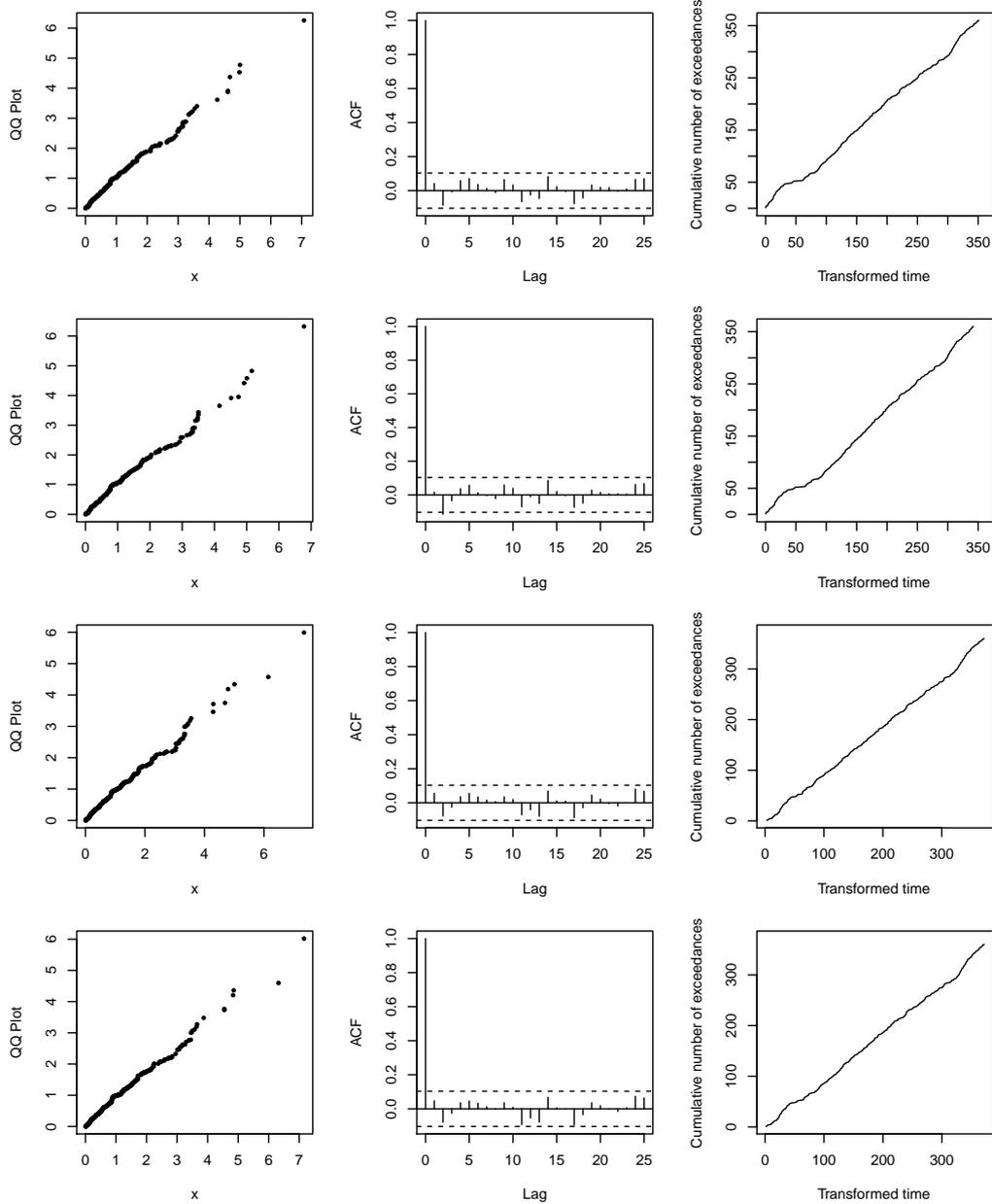


Figure 4.1: Goodness of fit in sample: QQ-plots of the residuals (left), autocorrelation function of the residuals (middle) and cumulative numbers of the residual process versus the transformed time $\{\tau_i\}$ (right), for the models applied to the log-returns of the WTI market. From top to bottom, the time event peaks over threshold model with Hawkes kernel (TE-POTh), the time event peaks over threshold model with ETAS kernel (TE-POTe), the ACD model for the expected conditional duration with generalized gamma distribution for the standardized residuals (gACD-POT), and the Log-ACD model for the expected conditional duration with generalized gamma distribution for the standardized residuals (gLog-ACD-POT).

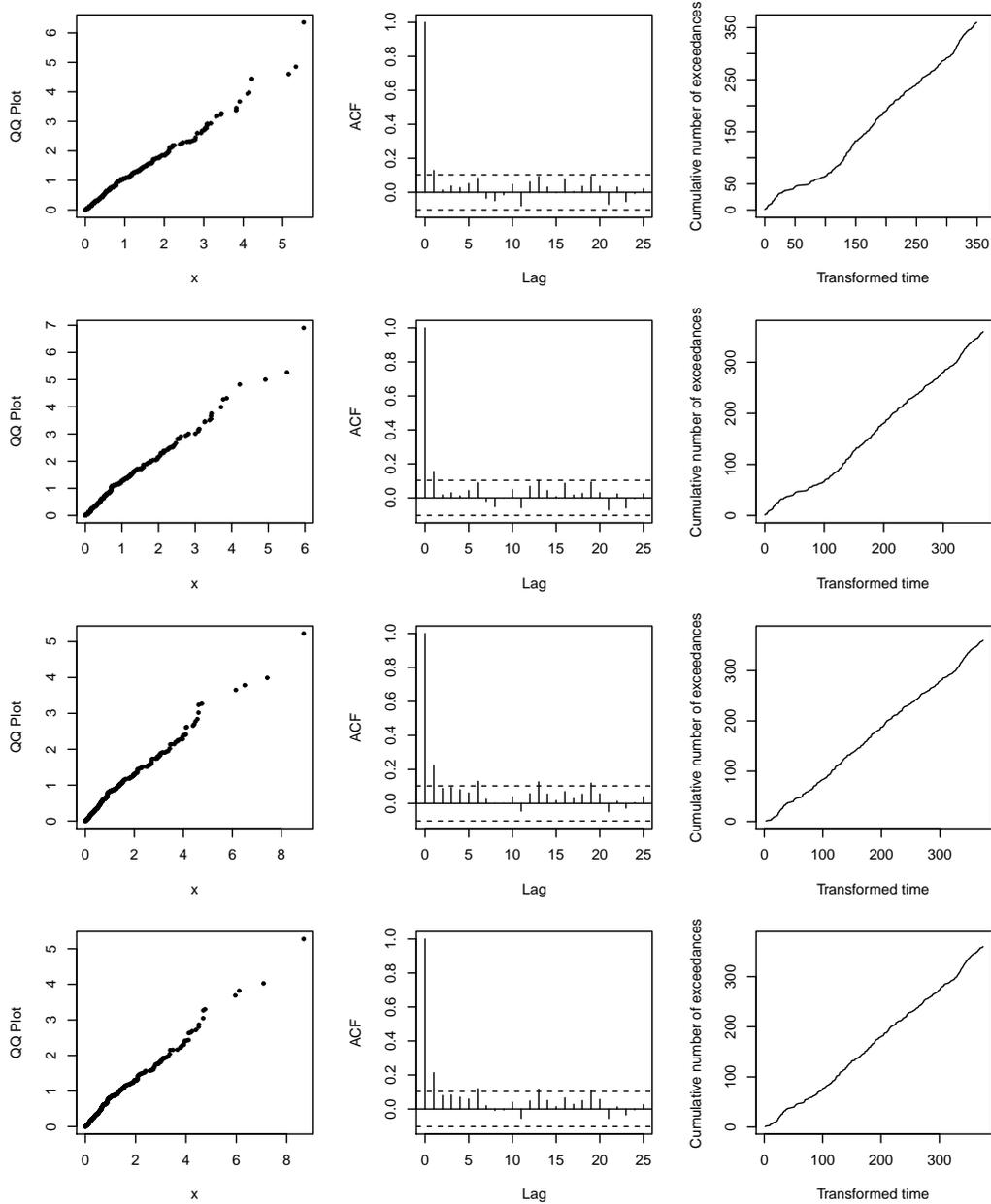


Figure 4.2: Goodness of fit in sample: QQ-plots of the residuals (left), autocorrelation function of the residuals (middle) and cumulative numbers of the residual process versus the transformed time $\{\tau_i\}$ (right), for the models applied to the log-returns of the Brent market. From top to bottom, the time event peaks over threshold model with Hawkes kernel (TE-POT_h), the time event peaks over threshold model with ETAS kernel (TE-POT_e), the ACD model for the expected conditional duration with generalized gamma distribution for the standardized residuals (gACD-POT), and the Log-ACD model for the expected conditional duration with generalized gamma distribution for the standardized residuals (gLog-ACD-POT).

statistic of 7.744 and a corresponding p-value of 0.171, while the TE-POTe model, has a chi-squared statistic of 9.861 and a corresponding p-value of 0.079. The results for the ACD-POT model could indicate that the specification of the scaling parameter is not flexible enough to model the marks or exceedances, since the scaling parameter specification only linearly depends on the temporal intensity of the inter-exceedance times and the magnitude of the marks.

Finally, in order to determine the quality of the times component of our models, i.e., the conditional intensity of the ground process λ_g , we employ the residual analysis method for point processes resumed briefly in the AppendixB. This is based on the change of time scale using the estimated conditional intensity. We investigated whether the transformed time-scale version of the data constitutes a homogeneous Poisson process according to the residual analysis introduced by Ogata (1988). The residual analysis for both markets indicates that all models seem to be acceptable in the changed time scale.

From the point of view of the market risk, we calculate the VaR in-sample from the different models for both markets, which are displayed in Figure 4.3. Taking a closer look at VaR estimates, there are clearly key aspects that mirror the complexity of capturing the extreme event dynamics by a model in response to an unpredictable, volatile and risky environment, as for example, the spikes reflected in the figures when the Gulf war and the Iraq War commenced in 1990 and 2003, respectively. A deeper analysis of the VaR will be done when we realize the backtesting of all models. The quotation from Aaron Brown (Risk Manager) in the June/July 2008 issue of the “Global Association of Risk Professionals” perfectly describes the importance of backtesting VaR models: “Value-at-Risk is only as good as its backtest. When someone shows me a Value-at-Risk number, I don’t ask how it is computed, I ask to see the backtest”.

For this reason, we include all models in the backtest in order to have a comparison of different alternatives, not only the best one in-sample.

Backtesting the models

Backtesting provides invaluable feedback about the accuracy of the models proposed to risk managers. The archetypal market risk model is a model that forecasts the VaR of a portfolio or stock market over one or more confidence levels, for a specified horizon. In this paper the backtest method consists of comparing the estimated conditional VaR for one day time horizon t , given knowledge of returns up to and including t for three different confidence levels (0.95, 0.99, and 0.999). For each day in the backtest we reestimate the models, something that immediately reveals possible stability problems of a model. Then, we reestimated the risk measures for each return series according to the equation (3.3).

In addition, we provide empirical evidence on the accuracy of actual VaR measures derived

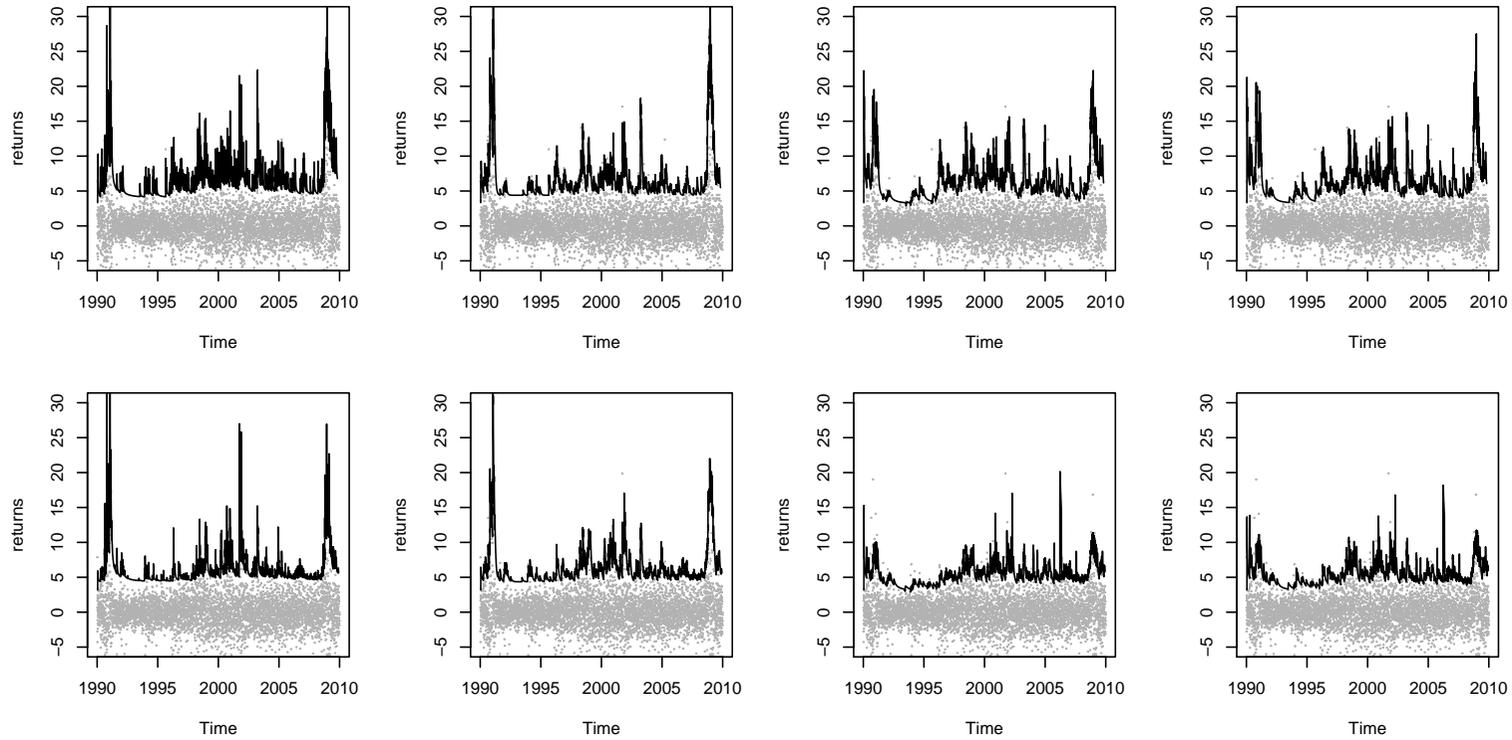


Figure 4.3: In-sample VaR estimates at the 0.99 confidence level for the WTI (top panel) and Brent (bottom panel) markets for their negative log-returns. The sample period spans from 2 January 1990 to December 31, 2009. The black line is the VaR estimation. From right to left, the time event peaks over threshold model with Hawkes kernel (TE-POT_h), the time event peaks over threshold model with ETAS kernel (TE-POT_e), the ACD model for the expected conditional duration with generalized gamma distribution for the standardized residuals (gACD-POT), and the Log-ACD model for the expected conditional duration with generalized gamma distribution for the standardized residuals (gLog-ACD-POT).

from the models. The first test is an unconditional coverage (LR_{uc}) test (Christoffersen, 1998). The idea is to test if the fraction of violations obtained for a particular risk measure significantly differs from the theoretical one. A violation of the VaR or Hit is defined as occurring when the ex-post return is lower than the VaR. A second test proposed by Christoffersen (1998) is a test of independence (LR_{ind}) among VaR violations, where under the null hypothesis, a violation today has no influence on the probability of a violation tomorrow. The third test is a combination of the last two tests which is known as the conditional coverage (LR_{cc}) test. The fourth approach, proposed by Berkowitz et al. (2009), tests for uncorrelatedness among the VaR violations. In particular, we suggest the well-known Ljung-Box (BT) test of the violation sequence's autocorrelation function. The last test, named the Dynamic Quantile (DQ) test, was introduced by Engle and Manganelli (2004). The idea is to regress the violations on the VaR for the present period on a judicious choice of explanatory variables. In our case, denoted by the DQ_{hit} , the regressor vector contains one constant and lagged VaR violations. All of these measures are reviewed briefly in the AppendixC.

Table 3 reports the results on the VaR backtesting exercise for all confidence levels. Entries in the columns are the significance levels (p-values) of the respective tests. A p-value less than or equal to 0.05 will be interpreted as evidence for rejecting the null hypothesis.

We observe that for all the models the results are more than satisfactory. These indicate that no severe clustering of exceedances is present and that the VaR violations can be considered as independent at all the confidence levels. The major difference between the TE-POT and ACD-POT models is that the latter have the lowest average VaR ($\overline{VaR_\alpha}$) across all VaR levels. In other words, ACD-POT models on average bring about the lowest capital requirement.

Overall, the assessment of our results shows that the MSEPP models are stable and reliable, implying that this approach of modeling extreme values can be used for further application of extreme events. Moreover, these models allow us to take the heavy-tailness or the stochastic nature of the cluster of extreme events into consideration.

The results for the MSEPP models can be summarized as follows:

- The results in-sample lead to markedly favor the ACD-POT model approach according to the AIC. However, the results in goodness of fit indicate that the TE-POT models performance best. This is probably due to the specification of the scaling parameter in the ACD-POT model. This means that the parametric specification for the scaling parameter may be not flexible enough, because it depends linearly on the temporal intensity of the inter-exceedance times and the magnitude of the marks.
- The quality of the conditional intensity of the ground process fit by means of residual anal-

Models for WTI	α	% Viol.	LR_{uc}	LR_{ind}	LR_{cc}	BT	DQ_{hit}	\overline{VaR}_α
TE-POTe	0.95	19	0.71	0.27	0.51	0.21	0.29	3.33
	0.99	3	0.56	0.83	0.83	0.88	0.83	5.23
	0.999	1	0.44	0.94	0.74	0.96	0.83	9.91
TE-POTh	0.95	17	0.40	0.23	0.34	0.38	0.24	3.48
	0.99	3	0.56	0.83	0.83	0.88	0.83	5.70
	0.999	1	0.44	0.94	0.74	0.96	0.83	11.35
gACD	0.95	19	0.71	0.27	0.51	0.21	0.29	3.35
	0.99	5	0.67	0.73	0.86	0.80	0.73	5.02
	0.999	1	0.44	0.94	0.74	0.96	0.73	9.09
gLog-ACD	0.95	18	0.55	0.22	0.40	0.15	0.24	3.39
	0.99	6	0.38	0.67	0.63	0.76	0.68	5.11
	0.999	1	0.44	0.94	0.74	0.96	0.67	9.34
Models for Brent	α	% Viol.	LR_{uc}	LR_{ind}	LR_{cc}	BT	DQ_{hit}	\overline{VaR}_α
TE-POTe	0.95	13	0.07	0.36	0.12	0.51	0.37	3.48
	0.99	2	0.24	0.89	0.50	0.92	0.89	5.68
	0.999	1	0.44	0.94	0.74	0.96	0.89	10.46
TE-POTh	0.95	13	0.07	0.36	0.12	0.51	0.37	3.39
	0.99	2	0.24	0.89	0.50	0.92	0.89	5.57
	0.999	1	0.44	0.94	0.74	0.96	0.89	10.33
gACD	0.95	14	0.11	0.32	0.18	0.47	0.33	3.42
	0.99	6	0.38	0.67	0.63	0.76	0.68	5.14
	0.999	1	0.44	0.94	0.74	0.96	0.67	8.72
gLog-ACD	0.95	13	0.07	0.36	0.12	0.51	0.37	3.53
	0.99	5	0.67	0.73	0.86	0.80	0.73	5.35
	0.999	1	0.44	0.94	0.74	0.96	0.73	9.13

Table 3: Goodness of fit to assess the predictive performance in the backtest of the models under consideration for the WTI and Brent market returns. Entries in the columns are the significance levels (p-values) of the respective tests, with exception of the level α and the number of violations at the VaR (% Viol.). The first test is an unconditional coverage (LR_{uc}) test (Christoffersen, 1998). A second test proposed by Christoffersen (1998) is a test of independence (LR_{ind}) between violations of the VaR, where under the null hypothesis a violation today has no influence on the probability of a violation tomorrow. The third test is a combination of the last two test which is known as the conditional coverage (LR_{cc}) test. The fourth approach proposed by Berkowitz et al. (2009) tests for uncorrelatedness of the violations. In particular, we suggest the well-known Ljung-Box (BT) test of the violation sequence's autocorrelation function. The last test, named the Dynamic Quantile (DQ) test, was introduced by Engle and Manganelli (2004). The idea is to regress the violations on the VaR for the present period on a judicious choice of explanatory variables. In our case, denoted by the DQ_{hit} , the regressor vector contains one constant and lagged violations of the VaR. \overline{VaR}_α denotes the average value of the VaR estimates.

ysis method for point processes indicates that all models seem to be acceptable in the transformed time-scale version of the data, which constitutes a homogeneous Poisson process.

- Major improvements in VaR predictions (backtesting) are achieved in all aspects when accounting for the extreme event dynamics by means of the proposed models. In addition, VaR violation ratios are statistically equal to the theoretical values in all cases, and VaR violations are independent when using either the TE-POT model or the ACD-POT model, the latter being preferred overall.

5. Conclusions

The impact of extreme events on crude oil markets is of great importance in crude oil price analysis due to the fact that those events generally not only show strong impact on crude oil markets but also in world stock markets. Although, sometimes those extreme movements in crude oil prices can be due to the behavior of events exogenous to the macroeconomy, as for example; the outbreak of the 1990 Gulf War or the 2003 Iraq War.

For better estimation of the impact of extreme events on crude oil price, this study attempts to use a marked self-exciting point process (MSEPP) approach for the task. In the proposed method, we make use of two new classes of MSEPP models that seem particularly well suited. The idea was to create a model being able to incorporate stylized facts such as clustering of extreme events and autocorrelation of the inter-exceedance times of extreme events, i.e., properties that are observed in crude oil markets. The first class of models is formulated in relation to the time of occurrence of the extreme events. We call this class of models Time event POT (TE-POT) models and it is able to generate power-law or exponential decay between extreme events and short-term cluster burts. The second class of models can be interpreted as a combination between the classical Peaks Over Threshold (POT) model from Extreme Value Theory and the class of Autoregressive Conditional Duration (ACD) models that are popular in finance for high-frequency data analysis. For this reason we call it ACD-POT models.

The main conclusions that can be drawn from our empirical investigation in the WTI and Brent oil markets can be summarized as follows. The VaR estimates under different high confidence levels exhibit strong stability through a range of the selected thresholds, implying the accuracy and reliability of the estimated risk measures.

Overall, the assessment of our results shows that the MSEPP models are stable and reliable, not only in-sample results but also in the backtesting, implying that these approaches of modeling extreme values can be used to further applications. Finally, credible regions and simulation of

future risks could be derived, which would provide interested organizations and managers with a valuable measure of short term uncertainty.

Acknowledgments

The author thanks the DAAD-CRUCH initiative for the financial support via a Fellowship and the Chilean Agency CONICYT for their financial support. FONDECYT No. 11110247.

References

- Bauwens, L., Giot, P., 2000. The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks. *Annales d'Economie et de Statistique* 60, 117–149.
- Berkowitz, J., Christoffersen, P., Pelletier, D., 2009. Evaluating value-at-risk models with desk-level data. *Forthcoming in Management Science*.
- 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., McGill, J., 2012. High-frequency financial data modeling using hawkes processes. Submitted to *Journal of Banking and Finance*.
- Christoffersen, P., 1998. Evaluating interval forecasts. *International economic review*, 841–862.
- Daley, D., Vere-Jones, D., 2003. *An Introduction to the Theory of Point Processes*. Springer.
- Engle, R., Manganello, S., 2004. Caviar. *Journal of Business and Economic Statistics* 22 (4), 367–381.
- Engle, R., Russell, J., 1998. Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* 66, 1127–1162.
- Herrera, R., Schipp, B., 2009. Self-exciting extreme value models on stock market crashes. In: Schipp, B., Krämer, W. (Eds.), *Statistical Inference, Econometric Analysis and Matrix Algebra*. Physica, pp. 209–231.
- Herrera, R., Schipp, B., 2012. Var forecasts by extreme value models in a conditional duration framework. Submitted to *Journal of Empirical Finance*.
- Krehbiel, T., Adkins, L. C., 2005. Price risk in the nymex energy complex: An extreme value approach. *Journal of Futures Markets* 25 (4), 309–337.
URL <http://dx.doi.org/10.1002/fut.20150>
- Lunde, A., 1999. A generalized gamma autoregressive conditional duration model. Tech. rep., Aalborg University.
- Marimoutou, V., Raggad, B., Trabelsi, A., 2009. Extreme value theory and value at risk: Application to oil market. *Energy Economics* 31 (4), 519 – 530.
URL <http://www.sciencedirect.com/science/article/pii/S0140988309000322>
- Ogata, Y., 1988. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association* 83, 9–27.
- Pickands III, J., 1971. The two-dimensional poisson process and extremal processes. *Journal of applied Probability*, 745–756.
- Zhang, M., Russell, J., Tsay, R., 2001. A nonlinear conditional autoregressive duration model with applications to financial transactions data. *Journal of Econometrics* 104, 179–207.

AppendixA. Threshold choice for EVT

Observe that under the MSEPP method only the tail index ξ remains constant, while the scale parameter varies through time. From the point of view of the risk measures, a robust fit to a sample of extreme events and a good estimate of risk measures, as for example VaR, would be relatively

insensitive to departures from the model. This is valuable in actual financial problems where one of the most important objectives is to obtain a robust measure of risk. However, EVT implementation faces many challenges, one of the most important being the fact that EVT is designed for independent observations. Crude oil market returns tend to be dependent, and therefore, a standard methodology for threshold selection does not exist.

The critical point in threshold selection is that by increasing the number of observations for the series of maxima, some observations from the center of the distribution are introduced in the series, and that the tail index ξ as well as the VaR estimate are more precise but biased (i.e., there is less variance). On the other hand, choosing a high threshold reduces the bias but also makes the estimates more unstable.

Thus, the main objective in this section is to determine how sensitive the MSEPP framework is to the choice of the threshold u , and in particular the VaR estimates obtained by means of these models. To this end, we choose the optimal threshold indirectly, by choosing an interval where the threshold quantile seems to be more stable in relation to the VaR estimate. To compare the different intervals, we computed the mean squared error (MSE) pointwise of the estimators as follows:

- We fix in advance a grid of size 10×10 of possible threshold-quantiles q ($q_k \in [0.85, \dots, 0.94]$ for $k = 1, \dots, 10$ with, $q_k < q_{k+1}$ for all k), and different confidence levels α for the VaR estimates ($\alpha_j \in [0.95, \dots, 0.999]$ for $j = 1, \dots, 10$) that will be estimated through a TE-POT or ACD-POT Model.
- We choose a quantile threshold q_k and estimate a suitable TE-POT or ACD-POT model. Since the estimate of the VaR are time varying we compute a mean value $\overline{VaR}(q_k, \alpha_j)$ for each VaR level α_j (the results of these estimations are displayed in Figure A.1).
- To compare the different estimates, we calculate the following MSE

$$MSE_{\Sigma\alpha}(q_k - q_{k+1}) = \frac{1}{9} \sum_{j=1}^9 (\overline{VaR}(q_k, \alpha_j) - \overline{VaR}(q_{k+1}, \alpha_j))^2$$

- Finally, we choose the threshold by selecting an interval where the threshold quantile seems to be more stable.

For example, in Figure A.1 we display the $\overline{VaR}(q_k, \alpha_j)$ estimates for the analysis of threshold selection for the proposed models. The results indicate that at least for all the returns that we have considered, the threshold selection seems to have limited influence on the VaR estimates. In Table

Models	$100 \times MSE_{\Sigma\alpha}(q_k - q_{k+1})$ for the WTI market									
	$q_k - q_{k+1}$	0.85-0.86	0.86-0.87	0.87-0.88	0.88-0.89	0.89-0.90	0.90-0.91	0.91-0.92	0.92-0.93	0.93-0.94
TE-POT _h		0.043	0.114	0.308	0.233	0.444	2.229	1.193	0.285	1.035
TE-POT _e		9.467	26.885	3.109	5.109	2.332	3.806	3.494	1.399	0.416
gACD-POT		0.184	0.085	0.642	0.162	1.224	1.006	0.401	0.761	0.552
gLogACD-POT		0.403	0.770	1.638	0.179	0.421	1.438	0.318	0.528	0.605

Models	$100 \times MSE_{\Sigma\alpha}(q_k - q_{k+1})$ for the Brent market									
	$q_k - q_{k+1}$	0.85-0.86	0.86-0.87	0.87-0.88	0.88-0.89	0.89-0.90	0.90-0.91	0.91-0.92	0.92-0.93	0.93-0.94
TE-POT _h		0.015	0.059	0.021	0.132	0.319	0.021	0.572	0.484	0.063
TE-POT _e		0.939	0.702	1.261	1.462	1.528	1.311	0.881	0.613	0.725
gACD-POT		0.268	0.122	0.002	0.535	0.505	1.047	0.973	0.691	1.214
gLogACD-POT		1.007	0.181	0.014	0.439	1.031	1.296	1.543	0.885	0.95

Table A.4: $MSE_{\alpha_j}(q_k - q_{k+1})$ estimates for the threshold selection for the WTI and Brent markets.

A.4, we display the values of the $MSE_{\Sigma\alpha}(q_k - q_{k+1})$ respect to the thresholds intervals $q_k - q_{k+1}$. We observe that most threshold dependent models are the TE-POT_e model.

This table shows that a threshold between the quantiles 0.92 and 0.93 may be the most justified for all models because of the fact that between these two quantiles the models showed a great stability.

Appendix B. Goodness of fit in sample

Residuals Analysis residual analysis for point process involves rescaling or thinning the original point process in order to obtain a new point process that is homogeneous Poisson. The common element of residual analysis techniques is the construction of an approximate homogeneous Poisson process from the data points and an estimated conditional intensity function $\hat{\lambda}_g(t | \mathcal{H}_t)$. Suppose we observe a one-dimensional point process $\{t_1, \dots, t_n\}$ on $[0, T)$ with conditional intensity $\lambda_g(t | \mathcal{H}_t)$. It is well known that the points

$$\tau_i = \int_0^{t_i} \lambda_g(t | \mathcal{H}_t) ds, \quad (\text{B.1})$$

for $i = 1, \dots, N(T)$ constitute a homogeneous Poisson process of rate 1 on an interval $[0, N(T)]$ which is therefore part of a transformed time axis. This new point process is called the residual process. If the estimated model $\hat{\lambda}_g(t | \mathcal{H}_t)$ is close to the true conditional intensity, then the residual process resulting from replacing $\lambda_g(t | \mathcal{H}_t)$ with $\hat{\lambda}_g(t | \mathcal{H}_t)$ in (B.1) should closely resemble a homogeneous Poisson process of rate 1. The resulting property

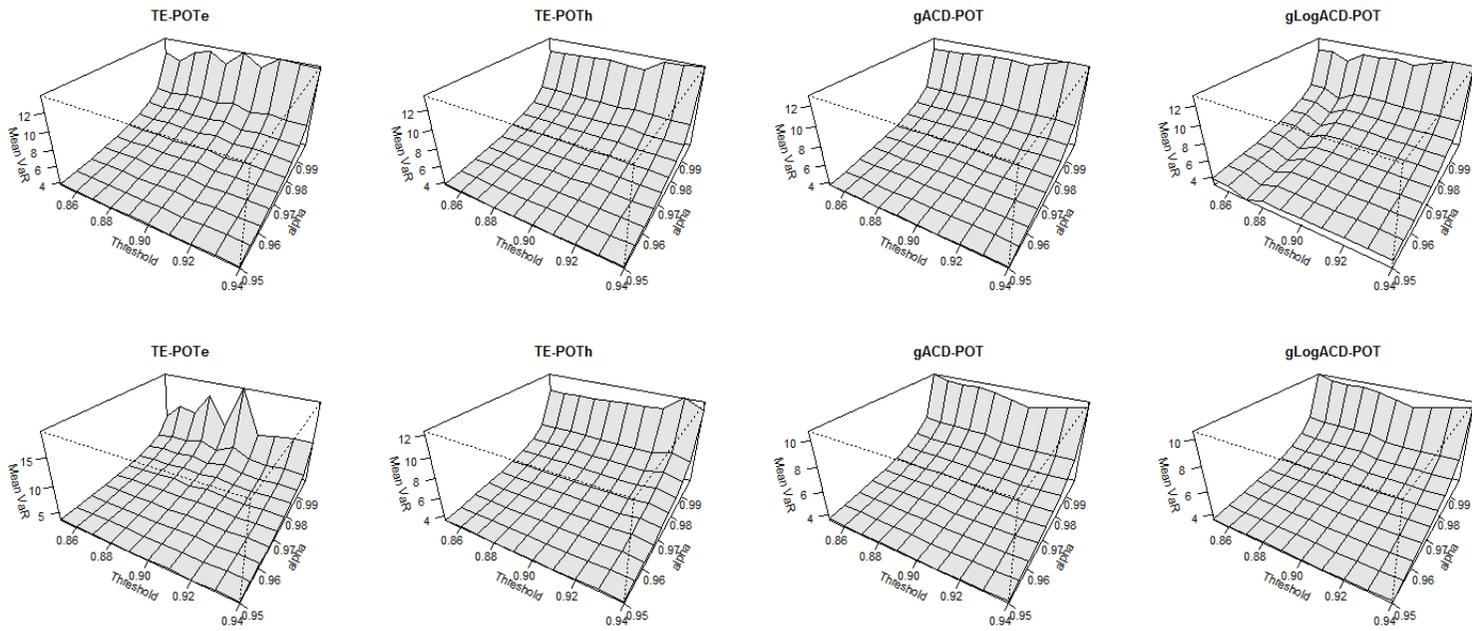


Figure A.1: The mean value of the average VaR estimates $\overline{VaR}(q_k, \alpha_j)$ for all the models for the WTI (top panel) and Brent (bottom panel) log-returns, for each VaR confidence level α_j , ($\alpha_j \in [0.95, \dots, 0.999]$ for $j = 1, \dots, 10$) and threshold-quantile q ($q_k \in [0.85, \dots, 0.94]$ for $k = 1, \dots, 10$).

of exponentially distributed durations enables us to test for the presence of a homogeneous Poisson process via a Kolmogorov-Smirnov test.

W-statistics In the case of the marks, we provide the W-statistics in order to assess our success in modeling the temporal behavior of the exceedances of the threshold u . The W-statistic is defined by

$$W = \xi^{-1} \ln \left(1 + \xi \frac{x - u}{\beta(t, y | \mathcal{H}_t)} \right).$$

This statistic states that if the GPD parameter model is correct, then the residuals are approximately independent unit exponential variables. In practice, the independence assumption can be checked via an ACF plot of the residuals.

Appendix C. Accuracy of VaR

Test of Unconditional Coverage (LR_{uc}): Christoffersen (1998) terms the sequence of VaR forecasts efficient with respect to the history \mathcal{H}_{t-1} if $E[I_t | \mathcal{H}_{t-1}] = \alpha$, where $I_t = \mathbb{I}(r_t < -VaR_t)$ with \mathbb{I} being the indicator function. Due to the fact that $I_t | \mathcal{H}_{t-1} \sim Ber(\alpha)$, $t = 1, 2, \dots, T$. Applying iterated expectations implies that I_t is uncorrelated (unconditional coverage) with any function of a variable in the information set available. This can be tested by means of a likelihood-ratio test

$$LR_{uc} = 2[\mathcal{L}(\hat{\alpha}; I_1, \dots, I_T) - \mathcal{L}(\alpha; I_1, \dots, I_T)] \sim \chi_1^2,$$

where \mathcal{L} is the log binomial likelihood. The maximum likelihood estimation $\hat{\alpha}$ is the ratio of number of violations, n_1 , to the total number of observations, $T = n_0 + n_1$.

Test of Independence (LR_{ind}): Christoffersen (1998) suggests a test of independence by modeling the number of violations I_t as a binary first order Markov chain with transition probability matrix

$$\Pi = \begin{bmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{bmatrix}, \quad \pi_{ij} = P(I_t = j | I_{t-1} = i),$$

as the alternative hypothesis of dependence. The joint likelihood, conditional on the first observation is given by

$$L(\pi^*; I_2, \dots, I_T | I_1) = (1 - \pi_{01})^{n_{00} + n_{10}} \pi_{01}^{n_{01} + n_{11}},$$

where n_{ij} represents the number of transitions from state i to state j . The maximum-

likelihood estimators under the alternative hypothesis are

$$\hat{\pi}_{01} = \frac{n_{01}}{n_{00} + n_{01}} \quad \text{and} \quad \hat{\pi}_{11} = \frac{n_{11}}{n_{10} + n_{11}}.$$

Under the null hypothesis of independence, we have $\pi = \pi_{01} = \pi_{11}$, from which the conditional binomial joint likelihood is defined as

$$L(\pi; I_2, \dots, I_T | I_1) = (1 - \pi_{01})^{n_{00}} \pi_{01}^{n_{01}} (1 - \pi_{11})^{n_{10}} \pi_{11}^{n_{11}}.$$

Similar to the unconditional coverage test, the likelihood ratio test is given by

$$LR_{ind} = 2 [\mathcal{L}(\hat{\pi}^*; I_2, \dots, I_T | I_1) - \mathcal{L}(\hat{\pi}; I_2, \dots, I_T | I_1)] \sim \chi_1^2.$$

Conditional Coverage (LR_{cc}): Christoffersen (1998) suggests combining the unconditional coverage test and the test of independence in order to test correct conditional coverage, because π^* is unconstrained. Then, we have

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

We can jointly test for independence and correct coverage using the conditional coverage test.

Ljung-Box test (BT): we implement a test statistics proposed by Berkowitz et al. (2009) for the autocorrelations of de-meanned violations $Hit_t(\alpha) = I_t - \alpha$, which form a martingale difference sequence. This is a Ljung-Box statistic, which is a joint test of whether or not the first m autocorrelations of $Hit_t(\alpha)$ are zero by calculating

$$LB_{VaR}(m) = T(T+2) \sum_{k=1}^m \frac{\gamma_k^2}{T-k}$$

where T is the sample size, γ_k is the sample autocorrelation at lag k and $LB_{VaR}(m)$ is asymptotically chi-square with m degrees of freedom.

Dynamic quantile test (DQ_{hit}): A relevant VaR model should also feature a sequence of VaR violations which are not serially correlated. Engle and Manganelli (2004) suggest the Dynamic Quantile (DQ), which can jointly test the hypothesis that $E[Hit_t(\alpha)] = 0$ and that $Hit_t(\alpha)$ is uncorrelated with the variables included in the information set, where $Hit_t(\alpha) = I_t - \alpha$.

Both tests can be done using the following artificial regression

$$Hit_t = \mathbf{X}\beta + \mathbf{u}, \quad \begin{cases} -\alpha, & \text{with probability } 1 - \alpha \\ 1 - \alpha, & \text{with probability } \alpha \end{cases},$$

where, under the null hypothesis, $H_0 = \beta = \mathbf{0}$, i.e, the regressors should have no explanatory power. Considering that the regressors are not correlated with the dependent variables under the null hypothesis, invoking a suitable central limit theorem Engle and Manganelli (2004) deduce the test statistic

$$DQ = \frac{\hat{\beta}\mathbf{X}'\mathbf{X}\hat{\beta}'}{\alpha(1-\alpha)} \sim \chi_{p+2}^2,$$

where p is the number of explanatory variables \mathbf{X} . In the empirical application, we use the dynamic quantile hit (DQ_{hit}) test, whose regressor matrix \mathbf{X} contains a constant and one lagged hit.