INTERNATIONAL JOURNAL OF BUSINESS, 20(1), 2015 ISSN: 1083-4346

On the Estimation of Extreme Values for Risk Assessment and Management: The ACER Method

Kai Erik Dahlen[a], Per Bjarte Solibakke[b], Sjur Westgaard[c], Arvid Næss[d]

[a] *Molde University College and Norwegian University of Science and Technology, Department of Industrial Economics and Technology Management, Norway*
kai.e.dahlen@himolde.no
[b] *Molde University College, Norway*
per.b.solibakke@himolde.no
[c] *Norwegian University of Science and Technology, Department of Industrial Economics and Technology Management*
sjur.westgaard@iot.ntnu.no
[d] *Norwegian University of Science and Technology, Department of Mathematical Sciences*
arvid.naess@math.ntnu.no

ABSTRACT

In this paper we use an Average Conditional Exceedance Rate (ACER) method to model the tail of the price change distribution of daily spot prices in the Nordic electricity market, Nord Pool Spot. We use an AR-GARCH model to remove any seasonality, serial correlation and heteroskedasticity from the data before modelling the residuals from this filtering process with the ACER method. We show that using the conditional ACER method for Value-at-Risk forecasts give significant improvement over a standard AR-GARCH model with normal or Student's-t distributed errors. Compared to a conditional generalized Pareto distribution (GPD) fitted with the Peaks-over-Threshold (POT) method, the conditional ACER method produces slightly more accurate quantile forecasts for the highest quantiles.

JEL Classifications: *C4, Q4*

Keywords: *commodity markets, electric energy, electricity, risk analysis, volatility forecasting.*

I. INTRODUCTION

During the recent years we have seen a move from regulated to deregulated electricity markets. Following the deregulation of the British electricity market in the early 1990s, the Norwegian government ruled for a deregulation of the national electricity market in 1991. In 1993 Statnett Marked AS was established, and in 1996 the Nord Pool market was created as a common electricity market for both Norway and Sweden, making it the first market for trading power in the world[1]. Finland joined the Nord Pool market area in 1998, and western and eastern part of Denmark joined in 1999 and 2000 respectively. In 2002 the electricity and energy derivatives markets were separated into Nord Pool Spot and Nord Pool ASA (now NASDAQ OMX Commodities Europe). Today Nord Pool Spot runs the spot (day-ahead) market as well as the intraday market (Elbas) in the Nordic region and Estonia, and it is the largest market of its kind.

The spot market is an auction based day-ahead market where the participants place bids on hourly production and consumption, at a given price and volume, with deadline 12.00CET scheduled for delivery the next day. The system price is then calculated as the price where the supply curve meets the demand curve, without any regard for possible bottlenecks in the transmission grid. To deal with congestion in the transmission grid, the local transmission system operators (TSO) can divide their area into different bidding areas. A congested line from bidding area one to bidding area two can then be dealt with by raising the price in the second bidding area. This is done in order to lower demand and increase the production incentive in the relevant bidding area. Today Norway is divided into 5, Sweden into 4 and Denmark into 2 bidding areas. Finland and Estonia are not divided into any bidding areas.

The intraday market functions as a supplement to the spot market to secure balance in the supply and demand for the electricity market, with trading available up to one hour before delivery. With the increasing fraction of unpredictable wind power in the Nordic region (and Germany, also covered by Elbas), leading to more unpredictable supply, the importance of the intraday market is increasing.

In the Nordic market region, during a year with average precipitation, almost half of the electricity is produced by hydropower. In Norway hydropower counts for almost 98% of the total electricity production, while in Sweden and Finland the production is a mixture of hydro, nuclear and thermal power. In Denmark thermal power (mainly coal fueled) is the largest source of electricity generation with an increasing installed capacity of wind power.

Due to the difficulties and costs of storing electricity (electricity in itself is in practice un-storable, but resources for electricity generation, e.g., water in a reservoir, can be stored) and the observed price inelasticity of consumers (Fezzi and Bunn, 2010), the observed spot prices are highly volatile. Compared to other energy commodities the price changes of the electricity spot prices are often very large, and often originate from events such as unexpected power plant outages, transmission grid congestion and unexpected increases in demand (Geman and Roncoroni, 2006). Some stylized fact of the price changes in these spot data is that they display very heavy tails, significant serial correlation, seasonality and volatility clustering (Weron, 2006). Seasonality concerning the electricity sport price is for the Nordic market easiest to observe on a yearly basis. Cold winters generally give a significant increase in prices due to heavily increased demand. With the Nordic market being highly dependent on hydro power, the observed prices are generally higher during dry years.

Extreme price changes or spikes are observed in electricity markets around the world (Escribano et al., 2002), and has been studied extensively over the last years. We will follow earlier work on modelling these spikes as an error process (Contreras et al., 2003; Garcia et al., 2005; Swider and Weber, 2007) using Extreme Value Theory (EVT) (Byström, 2005; Chan and Gray, 2006). In this paper we are going to use a conditional extreme value approach, as suggested by McNeil and Frey (2000), to model the tails of the return distribution for Value-at-Risk (VaR) estimation purposes. They suggested that a generalized Pareto distribution (GPD) is to be fitted, with the use of the Peaks- Over-Threshold (POT) method, to a dataset of residuals from an AR-GARCH filtering process. We will, instead of the POT method, use the Average Conditional Exceedance Rate (ACER) method to estimate the tails of the distribution of these residuals. We will evaluate the method considering both in sample fit and out of sample forecast accuracy, and comparing these results with the performance of the POT method. The ACER method has been shown (Naess and Gaidai, 2009) to produce more accurate estimates of extreme quantiles than the POT method. The rest of this paper is outlined in the following manner. In Section II we give a brief introduction to the theory used in this paper, focusing on Extreme Value Theory (EVT), Value-at-Risk (VaR) and GARCH models. In Section III we present the conditional approach used in this paper. Section IV introduces our data set. Section V gives a short overview of forecast evaluation, before our empirical findings is presented in Section VI.

II. THEORETICAL BACKGROUND

In this section we will provide a brief introduction to the theoretical background needed for this paper.

A. Generalized Extreme Value Distribution

Extreme Value Theory (EVT) is a branch of statistics where the goal is to describe and estimate the unlikely, dating back to the early works of Fisher and Tippett (1928) and Gnedenko (1943). Dealing with the classic EVT, this problem is often presented as finding the distribution of $M_n = \max(X_1, \ldots, X_n)$, where the X_i' s are independent and identically distributed (iid). This leads to the famous Extremal Types Theorem, which states that if the distribution of (a renormalization of) M_N is non-degenerate, then it is distributed as

$$P(M_n \leq x) = \exp\left\{ -\left[1 + \xi\left(\frac{x-\mu}{\sigma} \right) \right]_+^{-1/\xi} \right\}, n \to \infty \tag{1}$$

Where $(x)_+ = \max(0, x)$ and the parameters ξ, σ and μ is the shape, scale and location parameters. The parameterization in (1) is named the Generalized Extreme Value (GEV) distribution, first suggested by Jenkinson (1955). When using the GEV distribution to model extreme events only the block maxima (e.g., monthly or yearly maxima) are used in the parameter estimation, all other observations are discarded.

B. Peaks-Over-Threshold Method

Instead of using only the block maxima, the POT method, as its name describes, is able to use all observations that are above a certain threshold, and is based on the following. If M_n is of the Generalized Extreme Value (GEV) family (i.e., non-degenerate), then the exceedances over a large enough threshold u, conditioned on exceedance of this level, are distributed as a Generalized Pareto (GP) distribution (Coles, 2001; Pickands, 1975). This means that the distribution of these exceedances can be approximated with

$$P\left(X-u>y \middle| X>u\right) \approx H(y)=1-\left(1+\frac{\xi y}{\sigma}\right)_{+}^{-1/\xi} \quad (2)$$

It should be noted that this is an asymptotic result, valid as $u \to \infty$. On a second note the POT method requires the threshold exceedances to be independent. This is of course not always fulfilled, as in many cases extreme observations tend to be followed by extreme observations. Selection of the threshold level u can also be problematic, as there is no way to actually check the asymptotic assumptions made are actually fulfilled. In the literature the threshold is often set at a given percentile (i.e., 90%) of the empirical distribution. In later years the use of EVT in finance has increased, with direct application found in Embrechts et al. (1997).

C. Average Conditional Exceedance Rate Method

Compared to the above methods the ACER method does not require the observations to be i.i.d., and there are no asymptotic arguments in the derivation of the method. When using the ACER method the first step is to construct an exceedance rate function that represents the exceedance rate over a given level, conditioned on a chosen number of previous non-exceedances. This is done by basically counting the number of exceedances over this level, preceded by a chosen number of non-exceedances. The number of previous non-exceedances is chosen by inspection of several of these exceedance rate functions in the same plot, choosing the one which depends on the fewest previous non-exceedances which is identical to the higher order exceedance rate functions in the tail. After choosing the desired empirical ACER function we assume the shape of the tail is dominated by a function of the following form:

$$\varepsilon\left(\eta\right)=-q\left(\eta\right)\left(1+a(\eta-b)^{c}\right)^{-1/\xi} \quad (3)$$

which can be considered as the tail of a sub asymptotic GEV distribution, where q and c are introduced as sub asymptotic parameters. In the case of $q = c = 1$, this simply represents the GEV distribution. The ACER method was originally developed for cases when $\xi \to 0$, or when the underlying distribution can be assumed to be a Gumbel distribution. When dealing with heavy tailed data this assumption is not valid as the underlying extreme value distribution will be a Fréchet distribution, so in this case we have $\xi > 0$. In this case the exceedance rate of level η is

$$\varepsilon(\eta)=q(\eta)\left(1+a(\eta-b)^{c}\right)^{-1/\xi} \approx q\left(1+a(\eta-b)^{c}\right)^{-1/\xi} \tag{4}$$

where it is assumed that in the tail the function $q(\eta)$ vary slowly compared to the rest of the expression, so that we can assume q to be constant. Using this assumption, the parameters of (4) can be estimated by minimizing the square error function

$$F(a,b,c,q,\xi)=\sum_{i=1}^{n} w_i\left[\log\left(\hat{\varepsilon}(\eta_i)\right)-\log(q)+\xi^{-1}\log\left(1+a(\eta_i-b)^{c}\right)\right]^2 \tag{5}$$

Introducing a weighted linear regression

$$\xi^{-1}=-\frac{\sum_{i=1}^{n} w_i(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n} w_i(x_i-\bar{x})^2} \tag{6}$$

for ξ, and

$$\log(\hat{q})=\bar{y}-\hat{\xi}^{-1}\bar{x} \tag{7}$$

For q, where $x_i=\log\left(1+\alpha(\eta_i-b)^{c}\right)$, $\bar{x}=\sum_{i=1}^{n} w_i x_i / \sum_{i=1}^{n} w_i$, $y_i=\log\left(\hat{\varepsilon}(\eta_i)\right)$, and $\bar{y}=\sum_{i=1}^{n} w_i y_i / \sum_{i=1}^{n} w_i$, the minimization problem is reduced to a problem in three variables. The parameters a, b and c is then estimated by minimizing (5) with the use of the sequential quadratic programming algorithm[2].

These methods can be applied directly to the return series, but since we have observed significant GARCH effects it would be desirable to pre-filter the return series with a GARCH model to be able to accommodate for suddenly increasing or decreasing volatility.

D. Application to Simulated Time Series

Before using the method on the Nord Pool data we will try to motivate the use of the ACER method instead of the POT method by applying the two to simulated time series with known properties. By applying the two methods to simulated time series with known properties we can get an idea of the difference in performance for the two methods, and if the ACER method performs better argue that it should be used when estimating extreme quantiles. We will start by simply simulating i.i.d. innovations from a Student’s t distribution and comparing the estimated quantiles with the ones from the actual distribution. We simulate data sets consisting of 10 time series with 3,650 realizations each, giving us data sets with 36,500 observations. Five such data sets is simulated and used for comparing the performance of the ACER and POT methods. For the data sets simulated we have used ν = 4 degrees of freedom. Further we apply these methods to the simulated time series, and estimate the value that is expected to be exceeded once in 3,560,000 observations. The results from this estimation are found in Table 1. Presented are for the five simulated data sets the estimated value and percentage deviance from the real value (32.30), and the 95% confidence interval for this estimated value. From this table we see that while the ACER method estimate a value closer to the real value 4 out of 5 times, the width of the confidence intervals is far greater when using the POT method. The reason for the difference in the confidence intervals is likely due to the fact that the ACER method uses approximately 48% of all observations, compared to the approximately 10% of all observations used by the POT methods.

Table 1
The estimated tail quantile with 95% confidence interval for the t-model with v=4

η_{ACER}	CI_{ACER}	η_{POT}	CI_{POT}
32.99(2.1%)	[29.32, 37.80]	33.82(4.7%)	[23.75, 43.88]
27.78(14.0%)	[24.17, 31.59]	24.56(24.0%)	[18.35, 31.05]
31.68(1.9%)	[28.43, 34.72]	30.24(6.4%)	[21.87, 38.61]
29.76(7.7%)	[26.88, 33.31]	31.83(1.5%)	[22.46, 41.20]
31.80(1.5%)	[26.29, 34.25]	29.05(10.1%)	[21.15, 36.96]

E. GARCH Model

The autoregressive conditional heteroskedasticity (ARCH) model was introduced in Engle (1982) as a process with 0 mean and a non-constant variance conditional on the past. The ARCH(q) model can be written as

$$\varepsilon_t = \sigma_t z_t$$

$$\sigma^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \quad (8)$$

where $z_t \sim \mathrm{IID}(0,1)$, $\alpha_0 > 0$ and $\alpha_i \geq 0$ 0 for $i = 1, \ldots, q$. If we now assume that the error variance follows an autoregressive moving average (ARMA) model, we get the generalized autoregressive conditional heteroskedasticity (GARCH) model. The GARCH model was introduced by Bollerslev (1986), and can be written as

$$\varepsilon_t = \sigma_t z_t$$

$$\begin{aligned} \sigma^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-q}^2 \\ &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \end{aligned} \quad (9)$$

where $z_t \sim \mathrm{IID}(0,1)$, $\alpha_0 > 0$, $\alpha_i \geq 0$ for $i = 1, \ldots, q$ and $\beta_i \geq 0$ for $i = 1, \ldots, p$.

F. Value-at-Risk

To quantify the risk associated with these models we have here decided to use the Value-at-Risk (VaR) risk metric (Alexander, 2008). The VaR risk metric, for the next day, is defined for a probability α, as the value the loss will not exceed with probability α. The VaR is, for a stochastic variable X_t at time t, defined as

$$P\left(X_t < Y_\alpha\right) = \alpha \tag{10}$$

where $VaR_\alpha = -Y_a$ For the models used in this paper the VaR is straightforward to calculate. For the POT method the VaR is simply obtained by inverting the GPD for a given probability. This gives the VaR as

$$VaR_\alpha = \mu - \frac{\sigma}{\xi}\left(\alpha^{-\xi} - 1\right) \tag{11}$$

where μ, σ and ξ are the location, scale and shape parameters of the GPD respectively. For the ACER method the VaR is found just as easily by inverting the estimated ACER function for the desired exceedance level α. The VaR then becomes

$$VaR_\alpha = b + \left(\frac{1}{a}\left(\left(\frac{q}{1-\alpha}\right)^{\xi} - 1\right)\right)^{1/c} \tag{12}$$

where α, b, c and ξ are the parameters estimated by fitting the ACER method. It should be noted when using the conditional approach the VaR estimations need to be inserted into the equation for the conditional quantiles. As a last reminder it should be noted that for the VaR to be interpreted as the unlikely loss these two methods need to be fitted to the left tail of the return distribution. So using the VaR we now have a simple and straightforward method to quantify the risk in the Nord Pool spot market.

III. DATA

The data set used in this paper will be the daily spot price at the Nordic Power Exchange, Nord Pool. The data set spans, with daily observations, from January 1, 2000 to December 31, 2009 with a total of 3,653 observations. The spot price used is the price at 9:00 a.m. for the NO3 spot price region, which covers the middle of Norway. It should be noted that this price region was established on November 20, 2006, being integrated with the rest of Norway before that time. It is also worth mentioning that the spot prices in this market are actually a 1-day futures price where the prices for the following day are set by an auction at noon. After this auction the system price is calculated on basis of the bids for production and consumption that is received. The system operator can then deal with line congestion through different prices in the spot price areas. Plotted in Figure 1 are the daily prices observed over this ten year period, and in Figure 2 are the daily price changes, presented as logarithmic returns. Observable from these figures are large spikes in both the price and price change processes. These large price changes are important for the participants in this market. For a normal consumer the price changes may be of little interest, but for large consumers (e.g., power intensive industry) a sudden increase in the electricity price can in some extreme situations cause the need for a temporary shutdown. For the power producers the situation is reversed. They would want to produce when the price is high, and reduce their production should a negative price jump occur. There also seem to be some volatility clustering in the return series.

Figure 1
Daily electricity spot prices on Nord pool, NO3 price area, from January 1, 2000 to December 31, 2009

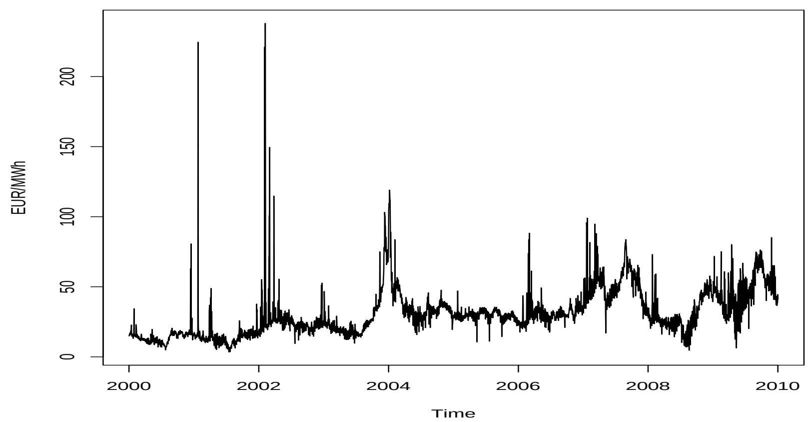
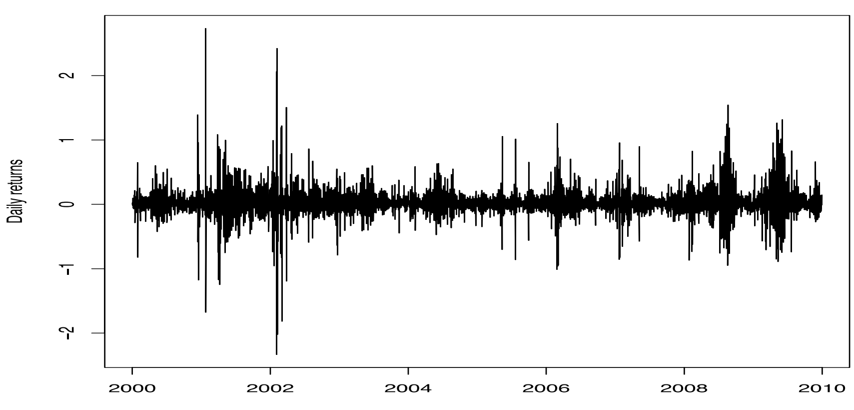


Figure 2
Daily electricity price changes on Nord pool, NO3 price area, from January 1, 2000 to December 31, 2009

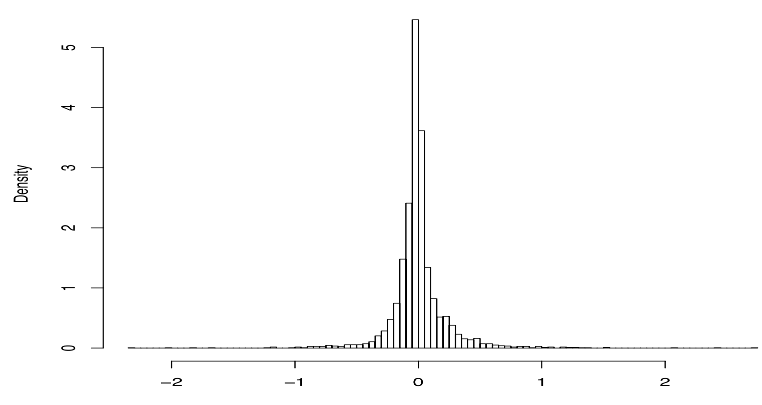


In Table 2 some descriptive statistics for our data set are presented along with test values for the Ljung–Box test (Ljung and Box, 1978) on both the return and squared return series. We observe a large excess kurtosis and positive skewness, meaning that both very small and very large price changes occur often compared to a normal distribution and large positive price changes are more common than the large negative price changes. From the Ljung–Box test, which should in the case of independent observations be Chi-square distributed with the chosen number of lags as degrees of freedom, we observe that there are significant serial correlation for all lags considered, which is what we expected. The p-value for the Ljung–Box statistic are for all tested lags equal to 0, clearly rejecting the null hypothesis of independent observations. For the squared returns we have significant serial correlation for all lags tested here, giving significant GARCH effects. Further the empirical 0.1%, 1% and 5% tail quantiles for both the left and right tail is calculated from the empirical cumulative distribution function. The difference between the right and left tail is not very large, and from the empirical distribution of the return series plotted in Figure 3 we only observe a small positive skewness.

Table 2
Descriptive statistics for the return series

Mean	Skew	Ex.Kurt	Q(1)	Q(2)	Q(7)	$Q^2(1)$	$Q^2(2)$	$Q^2(7)$
$3.00{\cdot}10^{-4}$	0.97	23.27	182.20	430.86	1505.67	265.50	606.49	863.32
Min	Max	$q_{0.001}$	$q_{0.01}$	$q_{0.05}$	$q_{0.95}$	$q_{0.99}$	$q_{0.999}$	
-2.334	2.732	-1.679	-0.689	-0.257	0.239	0.774	1.507	

Figure 3
Empirical distribution of the return series



IV. CONDITIONAL APPROACH

Following the conditional approach of McNeil and Frey (2000), and the use of a similar model on the electricity spot prices by Byström (2005) (following a similar study on stock returns (Byström, 2004), we want to model the data with the use of an AR-GARCH model, and then apply the Average Conditional Exceedance Rate method, as presented in Naess and Gaidai (2009), for estimation of the residual tail quantiles. The Average Conditional Exceedance Rate method is further generalized in Naess (2010) to accommodate the heavy tails, which are typical for financial data. To deal with the clear daily and weekly seasonality we include the AR(1) and AR(7) terms in the AR process. The intermediate terms are not included, as they do not improve the model fit in any way. For modeling the volatility a GARCH(1,1) process is chosen. This gives us a model that should be able to capture the serial correlation over the week and the observed heteroskedasticity. Our AR-GARCH model can be written as

$$\begin{aligned} r_t &= \alpha_0 + \alpha_1 r_{t-1} + \alpha_7 r_{t-7} + \varepsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \end{aligned} \tag{13}$$

where $\varepsilon_t = \sigma_t z_t$, and $z_t \sim \text{IID}(0,1)$ .In this paper we will assume z_t to be Normal or Student's t distributed, scaled to unit variance. The conditional quantiles, for these models, can be calculated as

$$q_{t,\alpha}^* = \alpha_0 + \alpha_1 r_{t-1} + \alpha_7 r_{t-7} + \sigma_t q_\alpha \tag{14}$$

where q_α is the standard α -quantile of the Normal or Student's t distribution. For the heavy tails observed in this type of data this standard AR-GARCH process will not be sufficient to model the tails of the return distribution accurately. The error distribution simply cannot be assumed to be Normal or Student's t, as we will observe later. Introduced by McNeil and Frey (2000), and applied to Nord Pool spot prices by Byström (2005), the Extreme Value Theory (EVT) approach has proven superior to the use of a standard AR-GARCH approach. We will use this approach, but instead of using a POT fitted Generalized Pareto distribution (GPD) for estimation of the tail quantiles of the standardized residual distribution, we will use the ACER method. The performance of this method will then be compared to the performance of the conditional model where the POT method is used. After using the ACER or POT method to estimate the tail quantiles of the residual distribution, the conditional quantiles for these models can be calculated as

$$q_{t,a}^* = \alpha_0 + \alpha_1 r_{t-1} + \alpha_7 r_{t-7} + \sigma_t q_\alpha \tag{15}$$

where q_α is the quantile of the residual distribution, associated with probability α , estimated by the POT or ACER method. The standardized residuals obtained by the Normal AR-GARCH filter can be observed in Figure 5. From this plot, and by calculation of the Ljung–Box test, these residuals are still slightly serially correlated, but most of the heteroskedasticity is removed. These conditional quantiles may also be regarded as the VaR(α) estimate when considering the distribution of the lower tail. For the in sample performance test we will estimate the conditional tail quantiles for all observations and then compare the number of empirical exceedances over these quantiles to what is expected. When dealing with the out of sample performance we start by estimating a model using the first five years and use this model to predict the tail quantiles for the next day. The model is then re-estimated with the last five years of the data, or a rolling window of length five years, for each day, and the tail quantiles for the next day is predicted. As with the in sample performance test the empirical exceedances over the predicted tail quantiles is compared to what is expected. The reason for the pre-filtering of the data set is that we want a GARCH process to model the volatility and to accommodate for sudden changes in the volatility. The use of an AR process for the autocorrelation is mainly because of the POT method's need of i.i.d. observations. There are no such i.i.d. requirements for the ACER method, so the use of an AR process is strictly not needed in this case.

V. EVALUATING FORECASTS

To evaluate our out-of-sample forecasts we will be using the unconditional coverage test of Kupiec (1995) and the conditional coverage test of Christoffersen (1998). The Kupiec (1995) test is a likelihood ratio test designed to test the unconditional coverage of the model. For this test we have a sequence of indicator variables I_t, taking the values

$$I_t = \begin{cases} 1 & \text{if } Y_t < Q_t \\ 0 & \text{if } Y_t \geq Q_t \end{cases} \tag{16}$$

where Q_t is the forecasted quantile and Y_t is the actual observation, both at time t. Under the null hypothesis of correct coverage, meaning $E[I_t] = p$, the likelihood ratio test statistic is

$$LR_{uc} = -2\log\left[\frac{(1-p)^{n_0}\, p^{n_1}}{(1-\pi)^{n_0}\, \pi^{n_1}}\right] \sim \chi_1^2 \tag{17}$$

Where n_0 and n_1 is the number of observations under and over the predicted quantiles respectively and $\hat{\pi} = n_1/(n_0 + n_1)$ is the maximum likelihood estimate of p. This test do only test if the observed number of quantile exceedances are close to the expected number of exceedances, not if there are any dependence between them.

Christoffersen (1998) proposed the conditional coverage test for joint coverage and independence of quantiles exceedances. The test statistic is

$$LR_{cc} = -2\log\left[\frac{\left(1-p\right)^{n_0} p^{n_1}}{\left(1-\pi_{01}\right)^{n_{00}} \pi_{01}^{n_{01}} \left(1-\pi_{11}\right)^{n_{10}} \pi_{11}^{n_{11}}}\right] \sim \chi_2^2 \tag{18}$$

where n_{ij} is the number of observations of type i followed by j, π_{01} is the fraction of observations of 0 followed by 1, and π_{11} are the fraction of observations of 1 followed 1.

VI. EMPIRICAL RESULTS

Using the models introduced, we now want to analyze the ten years of daily spot price data from Nord Pool. As an introduction we will start off with an unconditional tail fitting, that is, apply the POT and ACER methods directly to the return series. Further we will use the conditional approach as detailed in McNeil and Frey (2000), where both the ACER and POT methods will be used to fit the residuals produced by the AR-GARCH filtering.

A. Unconditional Approach

As mentioned, the unconditional approach will be to apply the POT and ACER method directly to the return series. To compare the performance of these two methods we will consider both in and out of sample fit. For the in sample fit the methods are simply used to estimate the tail quantiles of the return distribution and these quantiles will be compared to what we actually observe. For the out of sample test we will use the first five years of the data to estimate the model and then predict the tail quantiles for the following day. The rolling window is then moved to the next day and the model is re-estimated and the quantiles for the next day is calculated from the estimated parameters. This is repeated for the last five years of the data set.

For the POT method it is necessary for the data used to be i.i.d., which is from the Ljung–Box test results clearly not the case. We have both significant serial correlation and volatility clustering. To deal with this problem we will be declustering the data by extracting peaks with enough lags in between that it is reasonable to assume independence between the observations. For the ACER method there is no need to decluster the data as the correlation between lags is accounted for in the choice of ACER function.

In Table 3 the number of empirical exceedances over the estimated in sample quantiles is presented for both methods, along with the number of expected exceedances over these quantiles. The expected number of exceedances is calculated as $\left(1-p\right)\cdot n$, where n is the number of observations available and p is the associated probability in the left-most column. For the POT and ACER methods the actual number of exceedances over the estimated quantiles is in the POT and ACER columns respectively. We observe from this table that there is not a great difference between the numbers of exceedance over the estimated quantiles for the two methods. For the out of sample performance the number of exceedances over the estimated quantiles can be observed in Table 4. Again there is no great difference between the results of the two methods, but the ACER method is again slightly more accurate than the POT method. A problem with doing such out of sample prediction is that you use the last five years of the data with no more emphasis on what happened yesterday than five years ago. This means that your estimated quantiles and thus risk measures will take a long time to be able to incorporate a rise or fall in volatility. This again leads to the effect that most of the exceedances over the predicted quantiles will be observations from the periods with high volatility, while you ideally would like your exceedances to be uniformly distributed over the period in question.

Table 3
In sample performance of the unconditional methods

Probability	Expected	POT	ACER
.95	183	175	179
.99	37	38	38
.995	18	19	19
.999	4	3	4
.9995	2	3	2
.9999	0	1	0

Table 4
Out of sample performance of the unconditional methods

Probability	Expected	POT	ACER
.95	91	99	99
.99	18	29	24
.995	9	17	11
.999	2	2	2
.9995	1	1	1
.9999	0	0	0

B. Conditional Approach

To be able to accommodate for sudden changes in volatility, at least to some extent, the conditional approach is used. We start by filtering the data with an AR-GARCH process (We use a GARCH(1,1) process with AR parameters for the 1st and 7th lag) before fitting the POT and ACER methods to the standardized residuals (residuals standardized with the current volatility). The parameters for the AR-GARCH model with both normal and t-distributed errors are presented in Table 5, with the standard errors in brackets. For the model with normal distributed errors all parameters is significant at a 0.01 significance level, except ω, which is significant at a 0.1 significance level and μ which is non-significant. For the model with t-distributed errors all parameters is significant at 0.01 significance level, except μ, which is non-significant.

Table 5
Estimated AR-GARCH parameter values with standard errors in parentheses. Bold values are rejected at 0.1

Parameter	Value-N	Value-t
μ	$2.997\cdot10^{-3}(1.69\cdot10^{-3})$	$-1.628\cdot10^{-3}(1.14\cdot10^{-3})$
a_1	$-0.338(2.95\cdot10^{-2})$	$-0.283(1.57\cdot10^{-2})$
a_7	$0.338(2.33\cdot10^{-2})$	$0.425(1.45\cdot10^{-2})$
ω	$2.985\cdot10^{-4}(1.62\cdot10^{-5})$	$8.518\cdot10^{-4}(2.16\cdot10^{-4})$
α_1	$6.362\cdot10^{-2}(1.43\cdot10^{-2})$	$0.455(8.93\cdot10^{-2})$
β_1	$0.931(1.70\cdot10^{-2})$	$0.740(2.58\cdot10^{-2})$
ν	-	2.582(0.14)

Table 6
Descriptive statistics for the residual series

Mean	Skew	Ex.Kurt	Q(1)	Q(2)	Q(7)	$Q^2(1)$	$Q^2(2)$	$Q^2(7)$
$2.90\cdot10^{-4}$	4.44	97.72	6.70	102.91	$7.20\cdot10^{-3}$	$8.60\cdot10^{-3}$	$8.60\cdot10^{-3}$	$3.67\cdot10^{-1}$
Min	Max	$q_{0.001}$	$q_{0.01}$	$q_{0.05}$	$q_{0.95}$	$q_{0.99}$	$q_{0.999}$	
-8.761	7.704	-6.345	-2.424	-1.364	-1.366	-2.670	5.640	

Descriptive statistics and Ljung–Box test results for the residual series (Residual series after pre-filtering with the normal AR-GARCH model) can be found in Table 6. For the residual series we still have positive skewness and high excessive kurtosis. We observe that while there are still significant serial correlation it has be greatly reduced, and there are no significant GARCH effects in the residual series. It is observed that it is possible to remove slightly more of the serial correlation by including more AR-terms, but the difference is minimal so the model with less parameters is preferred. In Figure 4 the empirical distribution of the residuals is plotted and in Figure 5 the residual series is plotted. It is observed from these plots, and from the results of the Ljung–Box test, that this residual series is much less serial correlated and the observations can be reasonably considered independent.

After filtering the return series the POT and ACER methods are applied to the series of standardized residuals. As this series is much closer to i.i.d. than the return series, and observations over the chosen threshold seem to be independent of each other, there is no need to decluster the data in the same way that was done with the unconditional method. Nevertheless is should be noted that you still only get to use observations over the chosen threshold, which in this case will be less than 10% of the data. Using the POT method to fit a generalized Pareto distribution (GPD) to the data, with the threshold u selected from inspection Coles (2001), we get the parameters presented on the left hand side of Table 7. Here λ is the empirical estimate of $P(X > u)$. Inverting (2) for the desired probabilities give us the POT estimated quantiles, which in turn is inserted into (15) to get the conditional quantiles. For the ACER method the same procedure is used, and the parameters for the extrapolated ACER function can be found on the right hand side of 7.

Figure 4
Empirical distribution of the standardized residual series

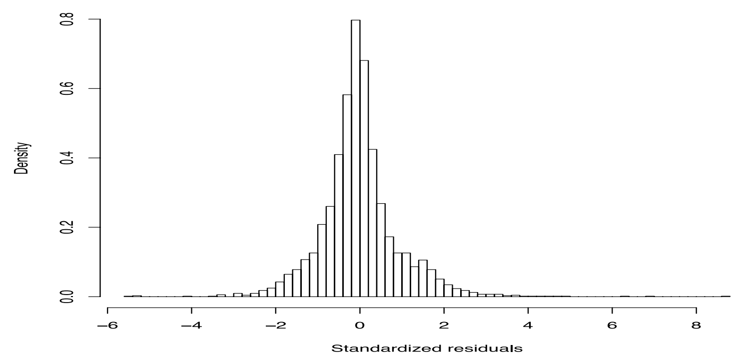


Figure 5
Plot of the standardized residuals after filtrating the return series with an AR-GARCH process

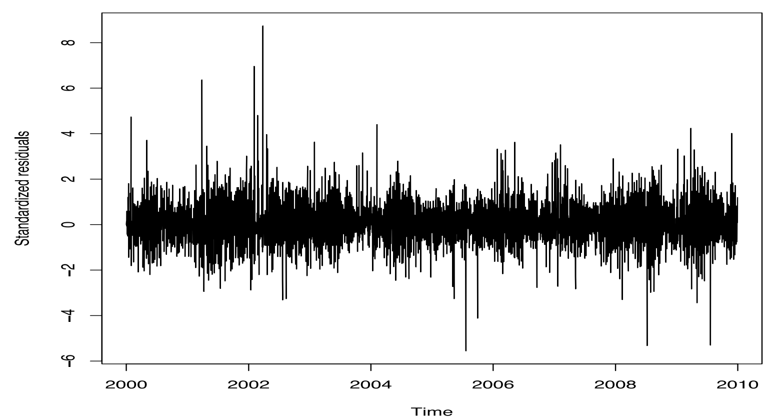


Table 7
Left: GDP parameters from POT in sample fit. Right: parameters from the ACER method in sample fit

Parameter	Value	Parameter	Value
σ	0.5318	ã	0.254
ξ	0.3118	b	0.010
λ	$9.32 \cdot 10^{-2}$	c	1.181
u	1	q	0.46
		ξ	0.334

Table 8
Number of exceedances over predicted in sample right quantiles for different methods

Probability	Expected	AR-GARCH-N	AR-GARCH-*t*	C-POT	C-ACER
0.95	182	227	257	187	182
0.99	37	129	39	34	35
0.995	18	118	7	19	18
0.999	4	86	2	4	4
0.995	2	72	1	3	1
0.999	0	56	0	1	0

When comparing the methods in sample performance the same procedure as for the unconditional approach is used, that is counting the number of exceedances over the estimated tail quantiles. The number of exceedances over these quantiles can be found in Table 8.

For the in sample performance of these two methods the same procedure as for the unconditional approach is used. In Table 8 the number of exceedances over a given quantile for the different methods is presented. We see that for the standard AR-GARCH model with standard Normal distributed errors the quantiles are severely underestimated for all quantile levels, and for the same model with t distributed errors the quantiles for the lower levels are severely overestimated. We also observe that the AR-GARCH model, where the POT method has been applied to the standardized residuals, clearly is able to estimate the extreme quantiles much better than just a standard AR-GARCH model. This is the same as was found by Byström (2005).

To compare the performance of the two methods it is important to assess the out of sample fit for the two methods. To do this we are going to start by estimating a model using the first five years of the data and then predict the conditional quantiles for the next day. The model is then re-estimated using what is now the last five years of the data, and again the next day conditional quantiles is predicted. This gives us a period of five years for the out of sample prediction. In Table 9 the number of exceedances over the predicted out of sample quantiles is presented. We see from this that the performance of the conditional POT and the conditional ACER method is quite similar. In table 10 the result from the unconditional and conditional coverage tests are presented, along with the percentage of out of sample observations below the estimated quantiles. From this we can clearly see that the AR-GARCH model with normal distributed errors is rejected for all quantiles, both for the unconditional and conditional coverage tests. The AR-GARCH model with Student's-t distributed errors is a significant improvement over the normal, but still the model is rejected by the conditional coverage test for the 95% quantiles. For the conditional EVT models there does not seem to be much difference. Both models pass the unconditional and conditional coverage tests. Looking at both Tables 9 and 10 the difference in performance of the two methods seem marginal, but also the ACER method seem to produce more accurate results in the most extreme quantiles.

Table 9
Number of exceedances over predicted out of sample right quantiles for different methods

Probability	Expected	AR-GARCH-N	AR-GARCH-*t*	C-POT	C-ACER
0.95	91	62	75	94	92
0.99	18	36	20	17	17
0.995	9	29	5	11	10
0.999	2	20	0	4	2
0.995	1	18	0	0	1
0.999	0	18	0	0	0

Table 10
Out of sample model diagnostics. First column is the desired quantiles. The second column is the out of sample quantiles estimated by different models. The third and fourth columns are the *p*-values of the unconditional and conditional coverage tests respectively. Bold values are rejected at 5% confidence

Model	Quantile	Exceedances	LR_{uc}	LR_{cc}
AR-GARCH-N	0.95	0.9660	0.0009	0.0038
	0.99	0.9802	0.0002	0.0005
	0.995	0.9841	0.000	0.000
	0.999	0.9890	0.000	0.000
AR-GARCH-*t*	0.95	0.9589	0.0721	0.0246
	0.99	0.9890	0.6852	0.7298
	0.995	0.9973	0.1324	0.3204
	0.999	1.000	NA	0.1611
C-POT	0.95	0.9485	0.7688	0.3431
	0.99	0.9907	0.7661	0.8077
	0.995	0.9940	0.5467	0.7754
	0.999	0.9978	0.8354	0.3768
C-ACER	0.95	0.9490	0.8514	0.3260
	0.99	0.9907	0.7661	0.8077
	0.995	0.9945	0.7749	0.9035
	0.999	0.9989	0.8985	0.9886

VII. CONCLUSION

For risk management purposes it is important for both producers and consumers of electricity to be able to describe the future distribution of electricity prices. We have in this paper looked at daily electricity prices in the Nordic market, modelling the tails of the distribution of future price changes. We have shown that our AR-GARCH model with ACER fitted residuals produce much more accurate estimates of the extreme quantiles of the return distribution than standard GARCH with either normal or Student's-t distributed errors. We also compare our conditional ACER method to the conditional GPD, which is already proven to be superior to the normal and Student's-t GARCH when it comes to modelling electricity spot price returns. Although the difference in the performance of these two methods is marginal, the conditional ACER method seems to produce more accurate out of sample forecasts for the most extreme quantiles.

ENDNOTES

1. http://www.nordpoolspot.com/How-does-it-work/
2. NAG toolbox for Matlab, e04wd function:
   http://www.nag.co.uk/numeric/MB/manual_21_1/pdf/ E04/e04wd.pdf

REFERENCES

Alexander, C., 2008, *Market Risk Analysis Volume I, II, III and IV*, John Wiley & Sons Ltd.

Bollerslev, T., 1986, "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics 31(3), 307 – 327*.

Byström, H.N.E., 2004, "Managing Extreme Risks in Tranquil and Volatile Markets Using Conditional Extreme Value Theory," *International Review of Financial Analysis 13, 133 – 152.*

Byström, H.N.E., 2005, "Extreme Value Theory and Extremely Large Electricity Price Changes," *International Review of Economics and Finance 14(1), 41 – 55*.

Chan, K.F., and P. Gray, 2006, "Using Extreme Value Theory to Measure Value-at-Risk for Daily Electricity Spot Prices," *International Journal of Forecasting 22(2), 283 – 300*.

Christoffersen, P.F., 1998, "Evaluating Interval Forecasts," *International Economic Review 39, 817–840*.

Coles, S., 2001, *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag: London.

Contreras, J., R. Espinola, F.J. Nogales, and A.J. Conejo, 2003, "Arima Models to Predict Next-Day Electricity Prices," *Power Systems, IEEE Transactions on 18(3), 1014– 1020*.

Embrechts, P., C. Klüppelberg, and T. Mikosch, 1997, *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin.

Engle, R.E., 1982, "Autoregressive Conditinal Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica 50(4), 987–1008*.

Escribano, Á., J.I. Pena, and P. Villaplana, 2002, "Modeling Electricity Prices:
International Evidence," *in* 'EFMA 2002 London Meetings.

Fezzi, C., and D. Bunn, 2010, "Structural Analysis of Electricity Demand and Supply Interactions," *Oxford bulletin of economics and statistics 72(6), 827–856*.

Fisher, R.A., and L.H.C. Tippett, 1928, "Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of A Sample," *Mathematical Proceedings of the Cambridge Philosophical Society 24(02), 180–190.*

Garcia, R.C., J. Contreras, M. Van Akkeren, and J.B. Garcia, 2005, "A GARCH Forecasting Model to Predict Day-ahead Electricity Prices," *Power Systems, IEEE Trans- actions on 20(2), 867–874*.

Geman, H., and A. Roncoroni, 2006, "Understanding the Fine Structure of Electricity Prices," *The Journal of Business 79(3), 1225–1261*.

Gnedenko, B., 1943, "Sur la Distribution Limite du Terme Maximum d'une série aléatoire," *Annals of Mathematics*.

Jenkinson, A.F., 1955, "The Frequency Distribution of the Annual Maximum (or minimum) Values of Meteorological Elements," *Quarterly Journal of the Royal*

Meteorological Society 81(348), 158–171. *http://dx.doi.org/10.1002/qj.49708134804*

Kupiec, P., 1995, "Techniques for Verifying the Accuracy of Risk Measurement Models," *Journal of Derivatives 2, 183–184*.

Ljung, G.M., and G.E. Box, 1978, "On A Measure of Lack of Fit in Time Series Models," *Biometrika 65(2), 297–303*.

McNeil, A.J., and R. Frey, 2000, "Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach," *Journal of Empirical Finance 7(3-4), 271 – 300*.

Naess, A., 2010, "Estimation of Extreme Values of Time Series with Heavy Tails," *Department of Mathematical Sciences, NTNU, Working Paper*.

Naess, A., and O. Gaidai, 2009, "Estimation of Extreme Values From Sampled Time Series," *Structural Safety 31(4), 325–334*.

Pickands, J., 1975, "Statistical Inference Using Extreme Order Statistics," *The Annals of Statistics 3(1), 119 – 131*.

Swider, D.J., and C. Weber, 2007, "Extended ARMA Models for Estimating Price Developments on Day-ahead Electricity Markets," *Electric Power Systems Research 77(5), 583– 593*.

Weron, R., 2006, *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*, Wiley.