Extreme Value Theory for Heavy-Tails in Electricity Prices

Florentina Paraschiv¹

Risto Hadzi-Mishev²

Dogan Keles³

Abstract

Typical characteristics of electricity day-ahead prices at EPEX are the very high volatility and a large number of extreme price changes. In this paper, we look at hourly spot prices at the German electricity market and apply extreme value theory (EVT) to investigate the tails of the price change distribution. Our results show the importance of delimiting price spikes and modeling them separately from the core of the price distribution. In particular, we get a realistic fit of the generalized Pareto distribution (GPD) to AR-GARCH filtered price change series, and based on this model accurate forecasts of extreme price quantiles are obtained. Generally, our results suggest EVT to be of interest for both risk managers and portfolio managers in the highly volatile electricity market.

1. Introduction

Finding realistic models for forecasting the tails of the electricity price distribution is often more important for risk managers involved in electricity markets than formulating the expectations about central tendencies. Unlike other types of commodities, electricity cannot be stored efficiently enough to enable inter-temporal balancing of load, which in combination with the relatively inelastic demand causes extremely large price changes. Extreme price changes can cause significant losses for the participants in this market. Thus, quantifying the size and the probabilities associated with these extreme price movements has attracted significant attention of both academics and practitioners. In this context, the goal of this paper is to find realistic models for forecasting the extreme tails of electricity price returns.

Extreme Value Theory (EVT) is a mathematical tool used to study the probabilities associated with extreme and, thus, rare events, and it describes the distribution of the sample maximum or the distribution of values above a given threshold. The application of EVT in modern risk management and its mathematical foundations are discussed in [18, 19 and 20]. These authors have empirically proven that models based on EVT are particularly well suited for the estimation of extreme quantiles of a distribution. In addition, the Bank for International Settlements [3] recognizes the EVT approach as accurate in delivering risk estimates. We test if the EVT- based model captures the extreme price behavior in the electricity market better than traditional time series models.

In this paper, we focus on modeling the EPEX Phelix day-ahead electricity prices. Extreme tail quantiles of the price change distribution are estimated by fitting both traditional time-series models (AR-GARCH) and an EVT-based model for the extreme tails. Our interest lies in modeling extremely

¹ Corresponding author: Florentina Paraschiv, University of St. Gallen, Institute for Operations Research and Computational Finance, Bodanstrasse 6, CH-9000, St. Gallen, Switzerland, florentina.paraschiv@unisg.ch

² Risto Hadzi-Mishev: University of St. Gallen, Switzerland, risto.hadzi-mishev@unisg.ch

³ Dogan Keles: Institute for Industrial Production (IIP), Karlsruhe Institute of Technology (KIT), Hertzstr. 16 - Building 06.33 (Uni-West), 76187 Karlsruhe, Germany, dogan.keles@kit.edu

large price increases. However, from a methodological point of view, modeling the negative tail of the distribution is completely analogous to modeling the positive tail.

A comprehensive review of the existing literature on modeling electricity prices is given in [26]. Studies which apply time series models for conducting forecasts of the extreme quantiles for energy markets can be found in [6], [1], [11], and [15]. Most of these authors employ GARCH models with various specifications for the innovations. [21] applied a quantile regression model and found that it performs better in out-of-sample forecasts of the electricity price distribution than pure time-series models like CaViaR and GARCH models. [7] and [24] show that EVT-based models are well suited for modeling the extreme tails of electricity prices. In this paper, we extend the existing literature on EVT-based models by a rigorous examination of the location of the threshold at which the tail becomes extreme. We found realistic forecasts of tail quantiles in electricity prices, which is of great importance for risk management purposes.

The rest of the paper is organized as follows: Section 2 briefly introduces the main characteristics of electricity prices. Section 3 focuses on the theoretical background of the EVT (POT) approach. Section 4 shows the application of EVT in estimating and forecasting the extreme tail quantiles of the electricity price change distribution. Section 5 concludes.

2. Characteristics of electricity prices

Exchange traded electricity contracts have certain characteristics that clearly distinguish them from financial assets or from other types of commodities. Unlike many commodities, electricity cannot be stored efficiently to enable inter-temporal smoothening of the load, which causes electricity prices to follow the same pronounced seasonal behavior as the demand [4], but modified by the availability of fluctuant renewable energy power production. In addition, market forces in this segment are relatively inelastic, at least on short run, which causes extremely large price movements (spikes) to occur when unanticipated variations in demand or supply happen [23]. We can summarize the main characteristics of electricity prices as follows: seasonality, mean reversion, spikes, volatility clustering, and negative prices. In particular, electricity prices exhibit one of the most pronounced seasonal behaviors among all commodities; in the long run they tend to revert to a long term mean while the price dynamic is characterized by relatively frequent jumps or spikes and volatility clustering. In addition, negative price bids are also allowed at EPEX. Given that these characteristics have been discussed extensively in the literature [4, 5, 12, 22], we will not elaborate them further in this paper.

3. Methodology

Within the EVT framework, there are two main types of models for modeling extreme values. The first group are the block maxima models. The main idea behind this group of models is to divide the data set into successive periods (blocks), for example weeks, months or years, and then to focus on the time series created only from the maximal values in each block. An alternative approach is the peaks-over-the-threshold (POT) method, which is generally regarded as more modern and more powerful tool for modeling extreme events, and hence it is our choice for this analysis [10],[24]. This approach deals with those observations in a certain data set that exceed a high threshold and models them separately from the rest of the observations. A more detailed discussion of the peaks-over-the-threshold (POT) method can be found in [17], [19] and [21] while here we only give a brief presentation of the method.

We assume that F is the unknown distribution function of a random variable X and our interest lies in understanding the distribution (F_u) of the values (x) that are above a certain threshold u . In such setting F_u is called the excess distribution function and it is defined as:

$$F_u(y) = P(X - u \leq y \mid X > u), \quad 0 \leq y \leq x_F - u$$

where $y = x - u$ are the excesses and $x_F \leq \infty$ is the right endpoint of F . If we rewrite the F_u in terms of F we get the following expression:

$$F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)} = \frac{F(x) - F(u)}{1 - F(u)} \quad (1)$$

In practice, since there is a small number of extreme observations, it is very difficult to estimate F_u [14]. However, the Pickands-Balkema-de Haan Theorem ([2] and [25]) indicates that for a large class of underlying distribution functions the excess distribution function $F_u(y)$, for large value of u , can be approximated by the Generalized Pareto Distribution (GPD). Namely, if

$$u \rightarrow \infty \quad \text{then} \quad F_u(y) \approx G_{\xi, \beta}(y)$$

where $G_{\xi, \beta}$ is the Generalized Pareto Distribution and its distribution function is defined as:

$$G_{\xi, \beta}(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\beta}\right)^{-\frac{1}{\xi}}, & \text{if } \xi \neq 0 \\ 1 - e^{-\frac{y}{\beta}}, & \text{if } \xi = 0 \end{cases}$$

where $\beta > 0$, and the support is $y \geq 0$ when $\xi \geq 0$ and $0 \leq y \leq -\beta/\xi$ when $\xi < 0$. The parameter β is the scale parameter while the parameter ξ is called tail index and gives indication about the heaviness of the tail, the higher the value of ξ the heavier the tail. When the GPD parameters are estimated we can calculate the (unconditional) tail quantiles associated with certain probabilities p [7]:

$$\hat{q}_p = u + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{n}{N_u} p \right)^{-\hat{\xi}} - 1 \right). \quad (2)$$

where N_u is the number of observations which exceed the threshold u . [9] showed that unconditional tail estimates are better suited for a prediction over longer time horizons.

4. Extreme Value Theory for estimating and forecasting tail quantiles

In this paper we will use the McNeil and Frey [18] method to estimate the tail quantiles of the electricity price change distribution. It is well known that spikes occur in electricity prices due to fluctuant renewable energies, wind and photovoltaic in Germany, or due to power plant outages. As already mentioned, [18, 19, 20] bring clear empirical evidence that the GPD is a flexible and well suitable modeling approach for extreme tails of one distribution. The use of this modeling approach is further encouraged by Basel III financial regulations on risk management. We begin this section with a brief description of our data set. Return series will be further filtered of seasonality and volatility clustering by employing two AR-GARCH model specifications. In one third model version, we will model separately with EVT the extreme tail of filtered residuals from the center of the distribution, where normal price changes are concentrated. Finally, the modeling performance of the two simple AR-GARCH specifications and the AR-GARCH+EVT approach is compared.

4.1. Data

The data set consists of EPEX Phelix hourly electricity prices, for the sample period 01/08/2008–14/01/2014. This data set is obtained from Bloomberg and presented in Figure 1. Overall, we have 47'507 hourly prices. Weekends are included. As shown in Figure 1, the magnitude of the price changes in the EPEX Phelix is extreme. Thus, the electricity prices during some periods are as high as 496.26 EUR/MWh and in some hours they go as low as -500 EUR/MWh. Since we are dealing with such extreme price changes, we choose to use the simple net returns $(P_t - P_{t-1})/P_{t-1}$ instead of logarithmic returns. The same approach was used by [7] and [24]. The rationale behind this choice is that logarithmic returns tend to underestimate the magnitude of price changes. Namely, the simple net return for a price change from 50 EUR/MWh to 70 EUR/MWh is 40%, while the logarithmic return calculated on this change is 33.6%. Given that our goal is to model the extreme tails of the price change distribution, simple net returns are a better choice for this analysis. The drawback of this approach is that in most of the cases the prices are bounded from below, which makes the return distribution positively skewed. However, since our interest is to model the extremely large positive returns, the problem with the simple net returns is not a drawback for us.

Results in Tables 1, 2 and Figures 2, 3 in Appendix indicate that the EPEX Phelix hourly returns are stationary and are characterized by extremely high volatility (standard deviation of 41.5%). The skewness and kurtosis of the data together with the visual inspection of the QQ plot presented in Figure 2 clearly show the heavy tails of the price change distribution. The high value of the Ljung-Box statistics indicates strong autocorrelation in the return series, which is obviously connected to the daily seasonality of electricity prices. The same conclusions can be drawn by looking at the autocorrelation functions of the absolute returns presented in Figure 3. The Ljung-Box statistics indicate volatility clustering, which implies that the GARCH model is an appropriate choice in this case.

4.2. Estimation of AR-GARCH model

We have chosen a combination of AR and GARCH model⁴, due to the significant volatility clustering effect and strong seasonality in the data. As we have already seen, the seasonality of the data is particularly obvious at the 24th hour and in order to capture this effect in addition to the AR(1) element in the model we include also the AR(24)⁵ term. Furthermore, in order to capture the conditional volatility of the simple return series we combine the AR model with a GARCH (1, 1) specification and get the following specification:

$$r_t = \alpha_0 + \alpha_1 r_{t-1} + \alpha_{24} r_{t-24} + \varepsilon_t \quad (3)$$

$$\sigma_t^2 = \phi_0 + \phi_1 \varepsilon_{t-1}^2 + \phi_2 \sigma_{t-1}^2$$

where σ_t^2 is the conditional variance of ε_t and ε_t is equal to $\sigma_t Z_t$. For Z_t we use two options, one where Z_t are normally distributed ($N(0,1)$) i.i.d and the second option where Z_t are Student's t -distributed i.i.d. innovations with mean 0, variance of 1 and degrees of freedom ν . The idea of

⁴ It is worth mentioning that the hourly day-ahead electricity prices and hence the hourly returns are not typical times series since the prices for the next day are published the previous day (day ahead) in block of 24 hours. However, in this paper our interest lies in modeling the extreme tails and we use time series models to check for seasonality and volatility clustering.

⁵ With an aim to account for the weekly seasonality, we have estimated additional model that also included the AR(168) term in the mean equation. However, the additional lag did not affect significantly the results of the model.

estimating the model with t -distributed innovations is supported by the fact that electricity price dynamics are characterized by heavier tails than the normal distribution (see also [7] and [24]). We use the maximum likelihood estimation to fit both versions of the models to our data. Results are presented in Table 3.

As we can observe in Table 3, for both AR-GARCH models we get significant parameter estimates. The parameters of the variance equation in both models are positive but their sum is not significantly lower than one, which indicates that an infinite unconditional variance cannot be rejected for any of the two models. This is not surprising considering the extremely fat-tailed data series that we use.

AR-GARCH parameters	Model with Gaussian innovations		Model with Student's t innovations	
	Coefficient	Standard Error	Coefficient	Standard Error
Mean equation				
α_0	0,0191	0,00040	-0,0034	0,00038
α_1	0,0958	0,00153	0,2473	0,00410
α_2	0,3401	0,00117	0,3563	0,00237
Variance equation				
ϕ_0	0,0033	0,00002	0,0070	0,00023
ϕ_1	0,6066	0,00127	0,8098	0,02600
ϕ_2	0,3934	0,00084	0,1902	0,00680
Degree of Freedom				2,7465
Likelihood	2,05882E+14		3,64273E+14	

Table 3: Maximum likelihood estimates of the AR-GARCH parameters

Source: Bloomberg L.P. (2014) EEX Phelix hourly prices, own calculations.

In order to check how much of the autocorrelation has been removed from our return series by the AR-GARCH models, we look at the autocorrelation function of the standardized residuals (given in Figure 4), while in order to see how much of the heteroskedasticity has been removed, we will analyze the filtered residuals (given in Figure 5). Most of the autocorrelation and heteroskedasticity have been captured by our models, and both standardized residuals series of the EPEX Phelix price returns are independent-identically distributed series. Given that our standardized residuals are i.i.d., the next step in our analysis is to model them with EVT.

4.3. Extreme value theory for estimation of tail-quantiles

We will apply EVT, in particular the POT approach, for modeling the upper tail of the distribution of standardized residuals. The first step and one of the greatest challenges in this context is identifying the value of the threshold (u): the level beyond which returns become “extreme”. Instead of just fixing the upper tail at 10% of the observations, we applied several methods for the identification/location of the threshold. Once the threshold will be identified, we apply POT to describe the upper tail, as extensively discussed in Section 3. For short term predictions and for time series that exhibit volatility clustering, [18] proposed a two-stage method for modeling the conditional distribution against the current volatility and then fitting the GPD on the tails of the residuals. The idea is to apply the results from EVT to the standardized residuals of a GARCH model since it is assumed that these residuals are independent and identically distributed. More specifically, EVT is used to model the tails of the distribution of residuals, while the middle part is modeled with empirical distribution (kernel smoothed interior). Given the positive empirical evidence from the application of this method to financial time series, in the sequel we investigate the performance of this approach on the highly volatile electricity market and compare it with traditional time series models.

4.3.1. Determining the threshold

The theory tells us that u should be set high enough in order to satisfy the Pickands-Balkema-de Haan theorem but at the same time if we set u to very high level we will have only few observations to estimate the parameters of the GPD. Therefore, determining the threshold is of particular importance for the quality of our estimates. We will apply the three most widely used methods for this purpose in the literature: a) Hill plots, b) the so-called “Eyeball” method and c) Sample mean excess function. The first two methods should give us more general sense about the range from where it is safe to choose the threshold, while the third method should give us more precise information about its exact location.

4.3.1.1. Determining the threshold using the Hill estimator

The Hill estimator is a semi-parametric estimator for the tail index and has the following form:

$$\hat{\xi} = \frac{1}{i} = \frac{1}{C_t} \sum_{i=1}^{C_t} \log \frac{Z_{(i)}}{u} \quad (4)$$

where C_t is the number of observations in the tail, $2 \leq C_t \leq T$, T is the total number of observations, $Z_{(i)}$ indicates sorted data (in our case the standardized residuals) where the maximum is $Z_{(1)}$, second largest observation is $Z_{(2)}$ etc. The idea of this method is to calculate the tail index (or $\hat{\xi}$) for different number of observations in the tail (C_t) and to visually identify the region where the coefficient becomes stable [6]. It is worth mentioning that for each value for C_t in the equation for the Hill estimator, the value of u is automatically determined as the value above which we have only C_t standardized residuals. In this context, we have re-estimated the tail index almost 10'000 times for values of C_t from 2 to 10'000 (which is more than 20% of the standardized residuals) in order to identify the region in which the coefficient becomes stable. The results of this analysis are presented in Figure 6. As we can see from the chart, the tail index from the Hill estimator becomes fairly stable in the region above 1'000 C_t , but nevertheless it continues with a downward trend until 10'000 C_t . Thus, we should choose a threshold beyond which at least 1'000 observations are concentrated in the upper tail of the standardized residuals.

4.3.1.2. Determining the threshold using the “eyeball” method

This method is very similar to the Hill plot, the difference is that here we are estimating the tail index (ξ) and the scale parameter (β) using maximum likelihood estimation. Namely, we re-estimate the GPD parameters for different sizes of the upper tail of the standardized residuals. The idea is again to identify the region, or the size of the upper tail, in which the estimated parameters of GPD become stable. For this purpose, we have recalculated the GPD parameters for different sizes of the upper tail, ranging from 1% to 30% of the total number of observations for the standardized residuals, and the results are presented in Figures 7 and 8. The tail index (ξ) is relatively stable only in the region where the tail fraction is between 10% and 15% of the standardized residuals. On the other hand, the scale parameter (β) in Figure 8 becomes stable in the region where more than 15% of the standardized residuals are allocated to the upper tail. Having this in mind, we can conclude that the optimal level for the tail fraction would be around 15% of the observations which corresponds to an actual threshold (u) of 0.784.

4.3.1.3. Determining the threshold using the mean excess function

Detailed explanation of this approach can be found in McNeil et al. [20]. The sample mean excess plot for our standardized residuals is presented in Figure 9. The use of this plot is to help us determining the value of the threshold u . The mean excess function is linear in u [20], so the threshold should be set at the value of u in our analysis, such that our sample mean excess plot shows a linear trend thereafter. Figure 9 shows the value of 0.78 as an appropriate choice for our threshold (u). The value of 0.78 is very close to the result that we got from the “eyeball” method and is also in line with the result from the Hill plot. We therefore continue our analysis using 0.78 as our threshold u .

4.3.2. Estimation of the GPD parameters

The parameters are determined by fitting the GPD to the tail (above the threshold u) of our actual standardized residuals. In doing so, the maximum likelihood estimates for the tail and scale parameters are $\xi=0.3188$ and $\beta=0.6292$. As we can see, our tail index is larger than zero, which means that we are in the class of the Fréchet distribution. This is not surprising, given the heavy tails of our data series. The model fit for the upper tail of our standardized residuals is given in Figure 10.

4.3.3. Estimation of the tail quantiles

Since we have estimated the AR-GARCH model parameters with Gaussian and t -distributed innovations, we can now easily estimate the conditional tail quantiles assuming either normal or t -distribution. We do so by multiplying the estimates of σ_t with the standard quantiles of each distribution and adding this result to the conditional mean. The idea of this approach is to take into account the time dependence of our time series. In the case of EVT quantiles, we multiply the unconditional EVT tail quantiles from equation (2) with the estimates of σ_t , from the AR-GARCH model with normally distributed innovations, and then add this result to the conditional mean from the same AR-GARCH model. The mathematical representation of this procedure is the following:

$$\widehat{quantile}_{t,p} = \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-24} + \sigma_t q_p \quad (5)$$

where $\widehat{quantile}_{t,p}$ is the conditional tail quantile for a given probability level p , $\alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-24}$ is the conditional mean and σ_t is the conditional volatility. For the three different models we use three different versions of q_p . Namely, in the models with Gaussian and t -distributed innovations we use the standard quantiles of each distribution, while for the EVT model we use equation (2). The next step in our analysis would be to examine the accuracy of the three models. We do this by counting the number of actual returns over the whole sample that are larger than their respective estimated tail quantile (we call this the number of exceedances). Then, we compare the actual number of exceedances of our models with their theoretically expected number⁶ in order to assess the accuracy of the models. On that note, if a particular model works well, we should expect that the observed number of exceedances will be as close as possible to the theoretically expected. The results from this comparison are presented in Table 4.

⁶ For example, the theoretical number of exceedances of the 99% tail quantile over the time period of 47507h is $0.01 \cdot 47507 = 475$.

Probability	Expected	AR-GARCH (Gauss)	AR-GARCH t	Conditional EVT
0,95	2.374	2.262	1.001	2.389
0,99	475	1.134	209	460
0,995	237	914	117	232
0,999	47	614	32	45
0,9995	24	537	21	32
0,9999	5	397	4	6

Table 4: In-sample evaluation of estimated tail quantiles at different probabilities (number of exceedences)

Source: Bloomberg L.P. (2014) EEX Phelix hourly prices, own calculations.

Looking at the two AR-GARCH models, the results show that they both fail to capture the behavior of our positive extreme tail. However, it becomes apparent that the shape of the conditional distribution of the innovations plays an important role for estimating the extreme tail quantiles. The fat-tailed AR-GARCH model with t -distributed innovations gives much better results for the more extreme quantiles compared to the model with normally distributed innovations. The thin tailed model with Gaussian innovations significantly underestimates all tail quantiles, except the 95% tail quantile. Contrary to this, the AR-GARCH model with t -distribution systematically overestimates all tail quantiles, resulting in a number of exceedences that is always lower than theoretically expected. This result is not surprising given that AR-GARCH models are designed to capture the movement of the entire distribution. It is therefore reasonable to expect that these traditional time series models do not have a good performance in estimating the extreme tails, especially in the case of electricity data. But, despite this limitation, the AR-GARCH models are capable of modeling the time varying tail quantiles, meaning that in periods of relatively high (low) volatility the estimates for the tail quantiles are higher (lower) [7].

To overcome the shortcomings of traditional time series models, we combine the advantages of AR-GARCH models for estimating the time varying tail quantiles with the advantages of the unconditional EVT-based models, which allows us to model the tail quantiles separately from the rest of the distribution. As we can see from Table 4, the exceedences of the EVT-based model closely resemble the theoretically expected number of exceedences. This indicates that the EVT-based model can be a powerful tool for portfolio managers in estimating the worst-case scenarios in the context of stress testing. These results are also in line with the results from similar studies in the literature, particularly [7] and [24].

4.3.4. Seasonality effect on the model performance

As already mentioned, electricity prices are characterized by one of the most complicated seasonal behaviors among all exchange traded commodities. Thus, it is reasonable to expect that the performance of our models for estimating the extreme quantiles of the electricity price change distribution can be affected by the time of the year. We therefore analyze the models' performance by estimating the 99% tail quantile in different months of the year. The results from this analysis are presented in Table 5. As we can observe, our previous conclusions are not changed. Namely, the AR-GARCH model with Gaussian innovations again systematically underestimates the 99% tail quantile while the AR-GARCH model with t -distributed innovations systematically overestimates the 99% tail quantile in all months. At the same time, the exceedences of the tail quantile estimated with the conditional EVT model are fairly close to the theoretically expected. As we can see from the p -values, the empirically observed exceedences of this model are statistically different from the theoretically expected only in three months (July, September and December). Once again, this confirms our previous conclusion that the EVT-based model delivers significantly better results for modeling the extreme quantiles compared to the standard AR-GARCH models with normal- and t -distributed innovations.

4.4. Robustness of the models

For a robustness check, we reestimated our three models for the following five periods: 2005-2007, 2007-2009, 2009-2011, 2011-2013 and 2013 – March 2014. The results are presented in Table 6 and confirm our previous conclusions regarding the models' characteristics for estimating the extreme tail quantiles: the EVT-based approach delivers fairly precise estimates for all tail quantiles and in all subsamples.

Another interesting finding from this analysis is related to the fact that our maximum likelihood estimates for the tail index of the GPD indicate much heavier tails of the standardized residuals in the first three subsamples (period between 2005 and 2010) than in the last two subsamples (2011-2014). The reason for this becomes evident from the graphical representation of our data set, presented in Figure 12. As shown in there, the spikes in the price series occur more often and with higher magnitudes in the period 2005-2009 than afterwards. Having in mind that we want to estimate the extreme quantiles in the upper tail, it is reasonable to expect that our extreme tail will be much heavier in 2005-2009 period than afterwards. One potential reason behind this behavior of electricity price dynamic might be the increased participation of renewables [23], in particular the wind generation, since 2009 (Table 7). Namely, as Figure 13 indicates, there is a strong negative correlation between the tail index, representing the heaviness of the upper tail of electricity price change distribution, and the participation of wind production in total production of power in Germany. However, it is worth mentioning that the performance of the conditional EVT (GPD) model is not affected by this structural break.

Figure 13: Tail index of GPD of EEX Phelix and electricity produced by wind in Germany (2009-2013)

Source: Bloomberg L.P. (2014) EEX Phelix hourly prices, AG Energiebilanzene.V. [2014], own calculations

In order to further investigate the robustness of results regarding the performance of the three analyzed models, we investigate their performance in estimating the tail quantiles of prices in specific hours of the day. Time series are constructed from the daily returns of electricity prices for each hour of the day, separately. However, for simplicity we show estimation results for one hour with low load profile (hour 3) and two peak hours (hours 13 and 19). The graphical representation of the three daily return series and its descriptive statistics is presented in Figure 14 and Table 8, respectively. As one can see, the three daily return series exhibit much of the characteristics of our hourly returns. Namely, again there is extremely high volatility, volatility clustering and highly pronounced daily seasonality (evident from the Ljung-Box statistics which indicates strong autocorrelation in the first lag). This indicates that the current AR-GARCH model structure is also suitable for these daily returns with the only difference that the 24th lag will not be included (since we work here with daily data this time).

Next, the models are estimated for the three hours of the day. The results from the maximum likelihood estimation of the AR-GARCH parameters are presented in Table 9 while the in-sample estimation of the tail quantiles are presented in Table 10. The estimates in Table 9 indicate that the parameters of the variance equation in all models are again positive and their sum is not significantly lower than one, indicating that an infinite unconditional variance cannot be rejected for all of the models under consideration. The results presented in Table 10 for the three hours are very close to our previous conclusions and the number of empirically observed exceedances of the EVT-based estimates is consistently very close to the theoretically expected.

4.5. Forecasting of tail quantiles

In this section, we assess the day-ahead forecasting power of the different model versions. For this purpose, we use one-year rolling estimation sample and one year of test sample. We estimate the tail quantiles with our models for the first day (24 hours) of the testing period using the last 365 days (8'760 hours) from our estimation sample⁷. Then, we perform a forecast of the second day from our test period using again the last 365 days as estimation period. We re-estimate the AR-GARCH models each day and we create 24 hourly forecasts of the conditional mean and volatility. The volatility is then scaled up with the standard quantiles of the Gaussian or the t -distribution, for the AR-GARCH (Gaussian) or AR-GARCH- t model, respectively. In the case of the EVT approach for the tails, the forecasted conditional volatility from the AR-GARCH (Gaussian) model is scaled up with the estimated tail quantiles from equation (2), based on the daily re-estimated GPD parameters. However, as mentioned before, in the POT approach it is important to determine the appropriate threshold level (u). In order to avoid looking at charts for each day to determine the threshold in our forecasting process, for our rolling window analysis we fixed the tail fraction to the 14% upper most extreme observations. This approach is also in line with the recommendations in [17] and the methodology used by [7].

As one can see from Table 11, the performance of the models in our out-of-sample forecasts is consistent with our previous conclusions. Namely, the Gaussian AR-GARCH model systematically underestimates all tail quantiles except the 95% quantile, while the AR-GARCH- t model strongly overestimates the tail. Regarding the EVT-based approach, we see that the model only overestimated the 95% tail quantile significantly, resulting in lower number of exceedances compared to the theoretically expected. However, in all other quantiles the model performed very close to the theoretically expected number of exceedances, which is also confirmed by the high p -values of the unconditional coverage tests.

Probability	Expected	AR-GARCH (Gauss)	AR-GARCH t	Conditional EVT
0,95	438	282 (0.0%)	77 (0.0%)	296 (0.0%)
0,99	88	159 (0.0%)	15 (0.0%)	83 (61.8%)
0,995	44	131 (0.0%)	5 (0.0%)	49 (44.0%)
0,999	9	96 (0.0%)	0 (0.0%)	11 (46.7%)
0,9995	4	83 (0.0%)	0 (0.3%)	5 (77.2%)
0,9999	1	65 (0.0%)	0 (18.6%)	2 (30.5%)

The p -values of the unconditional coverage test are presented in parenthesis

Table 11: Out-of-sample forecast (24 hours) of the tail quantiles (number of exceedances)

Source: Bloomberg L.P. (2014) EEX Phelix hourly prices, own calculations

⁷ For example, we forecast the tail quantiles for 12.03.2013 using the sample of actual data for period 12.03.2012-11.03.2013, thereafter the tail quantiles for 13.03.2013 are forecasted based on the actual data for 13.03.2012-12.03.2013 and so on.

In order to further investigate the forecasting capabilities of the EVT-based approach, we analyze how the number of exceedances is distributed across the year. Table 12 summarizes the results. Since the other two models clearly underperform in forecasting the tail quantiles, here we present only the results for EVT-based approach. We observe that the exceedances are relatively close to the theoretically expected, except in July, August and September when the 99% tail quantile is overestimated.

Overall, the results of this analysis indicate that the EVT-based model delivers much more realistic estimates of the extreme tail quantiles compared to the classical financial time series models. These findings are particularly useful for the risk managers given the high volatility of electricity prices.

Conclusion

In this paper, we investigated day-ahead prices for electricity quoted at EPEX. The price change distribution in this market is very volatile and extreme spikes are common events. In our approach, we filtered the return series with an AR-GARCH model and then applied EVT to the standardized residuals. Estimates of both moderate and more extreme tail-quantiles are found to be more accurate than those from ordinary AR-GARCH models with normally- or t -distributed innovations. Our results converge to those of [7] and show the importance of delimiting extremes in electricity prices and modeling them separately from the core of the distribution.

For the identification of the threshold value we have tested different methods proposed in the literature and compared the results. The findings converge towards a threshold that delimitates the 10—15% upper most extreme values from the core of the distribution. This is consistent with the threshold location suggested in the literature.

Our results also suggest that the heaviness of the tail of electricity price changes at EPEX was reduced over time due to the continuous increase in the wind infeed. With almost zero marginal cost of production, renewable energies reduce the volatility of prices, as shown in [23]. Further research should address the question whether the shape of the extreme tails of electricity prices is described by market fundamentals.

The out-of-sample evaluation of multi-period tail-quantile forecasts strengthens the already strong in-sample support for (conditional) EVT-based risk management in electricity market. This indicates that the EVT-based model is a powerful tool for portfolio managers for estimating the worst-case scenarios in the context of stress testing. Accurate “value-at-risk” forecasts for power portfolios are of great interest for traders, brokers, distributors, consumers or producers.

Appendix

Figure 1: Hourly prices-EPEX Phelix for period 01/08/2008-16/01/2014

Source: Bloomberg L.P. (2014) EPEX Phelix hourly prices

Stationarity test	Null Hypothesis of the test	Version of the test	Test statistic	Critical value at 5% *	Result
Augmented Dickey-Fuller test	The time series have a unit root	with Intercept	-31,479	-2,861	Stationary
		with Intercept and Trend	-31,507	-3,410	Stationary
		without Intercept or Trend	-31,475	-1,941	Stationary
Phillips-Perron test	The time series have a unit root	with Intercept	-219,280	-2,861	Stationary
		with Intercept and Trend	-219,285	-3,410	Stationary
		without Intercept or Trend	-219,281	-1,941	Stationary
Kwiatkowski-Phillips-Schmidt-Shin test	The time series is stationary	with Intercept	0,197	0,463	Stationary
		with Intercept and Trend	0,026	0,146	Stationary

* Critical values for Augmented Dickey-Fuller test and Phillips-Perron test are based on MacKinnon (1996) one-sided p-values while the critical values for Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test are asymptotic critical value based on KPSS (1992, Table 1)

Table 1: Stationarity tests on the hourly simple net returns of EPEX-Phelix price

Source: Bloomberg L.P. (2014) EPEX Phelix hourly prices, own calculations using EViews 7.

Mean (%)	Standard Deviation (%)	Kurtosis	Skewness	$Q(24)$	$Q^2(24)$
2,64%	41,5%	2.350	36	2.657	257

Kurtosis is excess kurtosis, $Q(24)$ and $Q^2(24)$ are the Ljung-Box tests for autocorrelation at 24 lags in the return series and the squared return series.

Table 2: Descriptive statistics of the EPEX Phelix hourly price changes (simple net returns)

Source: Bloomberg L.P. (2014) EPEX Phelix hourly prices, own calculations.

Figure 2: QQ Plot of simple net returns of EPEX Phelix prices

Source: Bloomberg L.P. (2014) EPEX Phelix hourly prices, own calculations

Figure 3: Autocorrelation function of simple net returns of EPEX Phelix prices

Source: Bloomberg L.P. (2014) EPEX Phelix hourly prices, own calculations

AR-GARCH with Gaussian innovations

AR-GARCH with t-distributed

innovation **Figure 4: Autocorrelation function of the standardized residuals**
Source: Bloomberg L.P. (2014) EPEX Phelix hourly prices, own calculations

AR-GARCH with Gaussian innovations

AR-GARCH with t-distributed

innovations

Figure 5: Filtered residuals and filtered standard deviation

Source: Bloomberg L.P. (2014) EPEX Phelix hourly prices, own calculations

Figure 6: Hill plot

Source: Bloomberg L.P. (2014) EPEX Phelix hourly prices, own calculations

Figure 7: Estimated tail index for different fractions of the standardized residuals allocated in the upper tail

Source: Bloomberg L.P. (2014) EPEX Phelix hourly prices, own calculations

Figure 8: Estimated β for different fractions of the standardized residuals allocated in the upper tail
Source: Bloomberg L.P. (2014) EPEX Phelix hourly prices, own calculations

Panel a) Sample mean excess plot

Panel b) Sample mean excess plot (zoomed)

Figure 9: Sample mean excess plot for the standardized residuals

Source: Bloomberg L.P. (2014) EPEX Phelix hourly prices, own calculations

Figure 10: Upper tail of the standardized residuals

Source: Bloomberg L.P. (2014) EPEX Phelix hourly prices, own calculations

Figure 11: Semi parametric empirical CDF

Source: Bloomberg L.P. (2014) EPEX Phelix hourly prices, own calculations

Months	Expected	AR-GARCH (Gauss)	AR-GARCH t	Conditional EVT
Jan	41	116 (0,0%)	22 (0,1%)	48 (25,3%)
Feb	34	81 (0,0%)	12 (0,0%)	39 (37,4%)
Mar	37	101 (0,0%)	11 (0,0%)	31 (31,3%)
Apr	36	83 (0,0%)	17 (0,0%)	34 (74,5%)
May	37	72 (0,0%)	18 (0,1%)	38 (84,5%)
Jun	36	74 (0,0%)	18 (0,1%)	32 (54,6%)
Jul	38	62 (0,0%)	12 (0,0%)	22 (0,7%)
Aug	45	87 (0,0%)	19 (0,0%)	38 (33,0%)
Sep	44	59 (2,1%)	12 (0,0%)	25 (0,3%)
Oct	45	116 (0,0%)	13 (0,0%)	42 (71,6%)
Nov	43	140 (0,0%)	26 (0,5%)	51 (23,0%)
Dec	44	143 (0,0%)	29 (1,9%)	60 (1,7%)

The p-values of the unconditional coverage test are presented in parenthesis

Table 5: Number of exceedences (in-sample) of the 99% tail quantile estimates per month

Source: Bloomberg L.P. (2014) EPEX Phelix hourly prices, own calculations.

Panel a) Model comparison for the period 01.01.2005-31.12.2006

Probability	Expected	AR-GARCH (Gauss)	AR-GARCH t	Conditional GPD ($u=0,71, \xi=0,31$)
0,95	874	840	284	915
0,99	175	422	61	161
0,995	87	333	37	81
0,999	17	217	9	23
0,9995	9	185	3	20
0,9999	2	140	1	3

Panel b) Model comparison for the period 01.01.2007-31.12.2008

Probability	Expected	AR-GARCH (Gauss)	AR-GARCH t	Conditional GPD ($u=0,61, \xi=0,32$)
0,95	871	830	339	909
0,99	174	400	75	172
0,995	87	322	42	92
0,999	17	218	14	21
0,9995	9	197	8	14
0,9999	2	149	4	5

Panel c) Model comparison for the period 01.01.2009-31.12.2010

Probability	Expected	AR-GARCH (Gauss)	AR-GARCH t	Conditional GPD ($u=0,6, \xi=0,36$)
0,95	867	725	369	905
0,99	173	351	90	166
0,995	87	281	55	84
0,999	17	198	18	24
0,9995	9	170	10	13
0,9999	2	133	2	1

Panel d) Model comparison for the period 01.01.2011-31.12.2012

Probability	Expected	AR-GARCH (Gauss)	AR-GARCH t	Conditional GPD ($u=0,6, \xi=0,19$)
0,95	871	810	314	870
0,99	174	371	43	167
0,995	87	283	20	93
0,999	17	177	4	20
0,9995	9	147	2	9
0,9999	2	105	0	3

Panel e) Model comparison for the period 01.01.2013-11.03.2014

Probability	Expected	AR-GARCH (Gauss)	AR-GARCH t	Conditional GPD ($u=0,4, \xi=0,26$)
0,95	515	493	193	514
0,99	103	271	30	113
0,995	52	218	12	58
0,999	10	141	1	7
0,9995	5	123	0	3
0,9999	1	83	0	0

Numbers in the tables are number of exceedances of the model estimated for the specific tail quantiles. For each period we have estimated the AR-GARCH parameters separately and in the case of conditional GDP model we have separately determined the appropriate threshold (u) and estimated the tail index (ξ).

Table 6: Robustness check-Evaluation of the models for different periods (number of exceedances, in-sample)
Source: Bloomberg L.P. (2014) EPEX Phelix hourly prices, own calculations.

Figure 12: EPEX Phelix hourly electricity prices 01.01.2005-11.03.2014
Source: Bloomberg L.P. (2014) EPEX Phelix hourly prices, own calculations

Figure 14: Daily return series for EPEX Phelix prices for hour 1,13 and 19 (01.08.2008-16.01.2014)
Source: Bloomberg L.P. (2014) EPEX Phelix hourly prices, own calculations

	2009	2010	2011	2012	2013
Coal	42,6	41,5	42,8	44,0	45,2
Nuclear	22,6	22,2	17,6	15,8	15,4
Natural Gas	13,6	14,1	14,0	12,1	10,5
Oil	1,7	1,4	1,2	1,2	1,0
Renewable energies	15,9	16,6	20,2	22,8	23,9
- Wind	6,5	6,0	8,0	8,1	8,4
- Hydro power	3,2	3,3	2,9	3,5	3,2
- Biomas	4,4	4,7	5,3	6,3	6,7
- Photovoltaic	1,1	1,8	3,2	4,2	4,7
- Waste-to-energy	0,7	0,7	0,8	0,8	0,8
Other	3,6	4,2	4,2	4,1	4,0

Table 7: Electricity production in Germany by source (in %)
Source: AG Energiebilanzene.V. [2014]

Hour	Mean (%)	Standard Deviation (%)	Kurtosis	Skewness	Q (1)	Q(1) P-val
01	7,64%	74,7%	421	17	27	0
13	6,33%	47,7%	45	5	84	0
19	3,50%	31,6%	47	4	91	0

Kurtosis is excess kurtosis, $Q(1)$ is the Ljung-Box test for autocorrelation at the first lag in the return series and $Q(1)$ P-val is the P-value of the test.

Table 8: Descriptive statistics for the daily returns of EPEX Phelix for hours: 1h, 13h and 19h

Source: Bloomberg L.P. (2014) EPEX Phelix hourly prices, own calculations

AR-GARCH parameters	Hour 1				Hour 13				Hour 19			
	Gaussian		Student's t		Gaussian		Student's t		Gaussian		Student's t	
	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Mean equation												
α_0	0,0286	0,00380	0,0148	0,0035	0,0584	0,00512	-0,0194	0,0047	0,0454	0,00480	-0,0083	0,00397
α_1	-0,5560	0,00820	-0,2702	0,0215	-0,3903	0,01377	-0,029	0,0119	-0,3378	0,02515	-0,1257	0,01797
Variance equation												
ϕ_0	0,0175	0,00062	0,0330	0,0052	0,0472	0,00197	0,0005	0,0003	0,0041	0,00037	0,0016	0,00051
ϕ_1	0,7585	0,01020	0,8909	0,1263	0,9925	0,03253	0,00541	0,0017	0,2878	0,00904	0,0659	0,01577
ϕ_2	0,2415	0,00450	0,1091	0,0235	0,0075	0,00343	0,99459	0,0026	0,7122	0,00539	0,9341	0,01079

Table 9: Maximum likelihood estimates for the AR-GARCH parameters of hours: 1h, 13h and 19h

Source: Bloomberg L.P. (2014) EPEX Phelix hourly prices, own calculations

Probability	Hour 1				Hour 13				Hour 19			
	Expected	AR-GARCH (Gauss)	AR-GARCH t Dist.	EVT (Tail Fraction 10%)	Expected	AR-GARCH (Gauss)	AR-GARCH t Dist.	EVT (Tail Fraction 5%)	Expected	AR-GARCH (Gauss)	AR-GARCH t Dist.	EVT (Tail Fraction 4%)
0,95	99	85	46	102	100	93	34	99	100	113	47	98
0,99	20	42	12	19	20	41	5	20	20	43	6	21
0,995	10	31	7	11	10	34	3	13	10	35	2	9
0,999	2	20	2	2	2	23	0	3	2	21	1	3
0,9995	1	20	2	1	1	19	0	0	1	19	1	1
0,9999	0	17	0	0	0	14	0	0	0	11	0	0

Table 10: In-sample evaluation of tail quantiles for each hour at different probabilities (number of exceedences)

Source: Bloomberg L.P. (2014) EPEX Phelix hourly prices, own calculations

Months	Expected	Number of exceedences	Unconditional coverage test (p-value)
Jan	7	9	56%
Feb	7	6	80%
Mar	8	9	82%
Apr	7	6	65%
May	7	8	83%
Jun	7	11	15%
Jul	7	3	6%
Aug	7	0	0%
Sep	7	4	19%
Oct	7	7	89%
Nov	7	9	52%
Dec	7	11	17%

Table 12: Number of exceedences of the forecasted 99% tail quantile per month (conditional EVT model)

Source: Bloomberg L.P. (2014) EPEX Phelix hourly prices, own calculations

References

- [1]. **Aloui, C.** (2008) Value-at-risk analysis for energy commodities: long-range dependencies and fat-tails in return innovations. *The Journal of Energy Markets* 1,1, 21-40.
- [2]. **Balkema, A. A. and de Haan, L.** (1974). Residual life time at great age. *Annals of Probability*, 2:792-804.
- [3]. **Bank for International Settlements:** Range of practices and issues in economic capital frameworks. Basel Committee on Banking Supervision. <http://www.bis.org/publ/bcbs152.pdf>.
- [4]. **Blöchliger, L.** (2008). Power Prices – A regime-switching spot/forward price model with Kim Filter. PhD thesis, University of St. Gallen.
- [5]. **Burger, M., Klar, B., Müller, A. and Schindlmayr, G.** (2003). A spot market model for pricing derivatives in electricity markets. *Journal of Quantitative Finance*, 4 (1), 109-122. <http://mspcdip.mathematik.unikarlsruhe.de>.
- [6]. **Bunn, D., Andresen, A., Chen, D., and Westgaard, S.** (2013) Analysis and forecasting of electricity price risks with quantile factor models. Working paper series NTNU.
- [7]. **Bystrom, N.E.** (2005). Extreme value theory and extremely large electricity price changes. *International Review of Economics & Finance* 14, 41–55.
- [8]. **Danielsson, J.** (2011). Financial risk forecasting—the theory and practice of forecasting market risk, with implementation in R and Matlab. London : John Wiley & Sons Ltd.
- [9]. **Danielsson, J. and de Vries, C.** (1997). Beyond the sample: extreme quantile and probability estimation. Mimeo. Iceland University and Tinbergen Institute Rotterdam.
- [10]. **Embrechts, P., Kluppelberg, C., and Mikosch, T.** (1999). Modelling extremal events for insurance and finance. Applications of Mathematics. Springer. 2nd ed.
- [11]. **Füss, R., Adams, Z., and Kaiser, D. G.** (2010), The predictive power of value-at-risk models in commodity futures markets. *Journal of Asset Management* 11, 4, 261-285.
- [12]. **Geman, H. and Roncoroni, A.** (2006). Understanding the fine structure of electricity prices. *Journal of Business*, 79 (3), pp.1225-1261. <http://papers.ssrn.com>
- [13]. **Ghosh, S. and Resnick, S.** (2010). A discussion on mean excess plots. *Scholastic Processes and their Application*, 120 1492-1517.
- [14]. **Gilli, M. and Kellezi, E.** (2006). An application of extreme value theory for measuring financial risk, *Computational Economics* 27(1), 1-23.
- [15]. **Giot, P. and Laurent, S.** (2003) Market risk in commodity markets: a VaR approach. *Energy Economics* 25, 5, 435-457.
- [16]. **Hung, J.C., Lee, M.C., and Liu, H.C.** (2008) Estimation of value-at-risk for energy com-modities via fat-tailed garch models. *Energy Economics* 30, 3, 1173-1191.
- [17]. **Kupiec, P.** (1995). Techniques for verifying the accuracy of risk management models. *Journal of Derivatives*, 3, 73-84.
- [18]. **McNeil, A.J. and Frey, R.** (2000). Estimation of tail related risk measure for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 34 ,7, 271-300.
- [19]. **McNeil, A.J.** (1999). Extreme value theory for risk managers. Mathematical Department of ETH Zurich. <http://www.math.ethz.ch/~mcneil/ftp/cad.pdf>.
- [20]. **McNeil A.J., Frey, R. and Embrechts P.** (2005). Quantitative Risk Management: Concepts, Techniques and Tools. Princeton and Oxford : Princeton University Press.
- [21]. **Nyström, K. and Skoglund, J.** (2002). Univariate extreme value theory, GARCH and measures of risk. Working Paper, Swedbank.
- [22]. **Ockenfels A., Grimm V. and Zoettl, G.** (2008). Electricity market design - the pricing mechanism of the day-ahead electricity spot market auction on the EEX. http://cdn.eex.com/document/38615/gutachten_eex_ockenfels_e.pdf.
- [23]. **Paraschiv, F., Erni, D. and Pietsch R.** (2014) The impact of renewable energies on EEX day-ahead electricity prices. *Energy Policy*, 73, 196-210.
- [24]. **Paraschiv, F.** (2013). Price dynamics in electricity markets. In *Handbook of Risk Management in Energy Production and Trading* (pp. 47-69). Heidelberg: Springer Science, DOI:10.1007/978-1-4614-9035-7_3.
- [25]. **Pickands, J. I.** (1975). Statistical inference using extreme value order statistics. *Annals of Statistics*, 3:119-131.
- [26]. **Weron, R.** (2014). Electricity price forecasting: A review of the state-of-the art with a look into the future. *International Journal of Forecasting*, 1030-1081.