

Forecasting Price Spikes in Electricity Markets

EFTHYMIOS STATHAKIS*

Democritus University of Thrace

THEOPHILOS PAPADIMITRIOU

Democritus University of Thrace

PERIKLIS GOGAS

Democritus University of Thrace

Electricity markets are considered to be the most volatile amongst commodity markets. The non-storability of electricity and the need for instantaneous balancing of demand and supply can often cause extreme short-lived fluctuations in electricity prices. These fluctuations are termed price spikes. In this paper, we employ a multiclass Support Vector Machine (SVM) model to forecast the occurrence of price spikes in the German intraday electricity market. As price spikes, we define the prices that lie above the 95th quantile estimated by fitting a Generalized Pareto distribution in the innovation distribution of an AR-EGARCH model. The generalization ability of the model is tested in an out-of-the-sample dataset consisting of 4080 hours. Furthermore, we compare the performance of our best SVM model against Neural Networks (NNs) and Gradient Boosted Machines (GBMs).

Keywords: electricity prices; extreme value theory; exponential GARCH; multiclass; support, vector machines.

JEL Classifications: G1, C19, C58, C53

1 Introduction

In wholesale electricity markets, settlement prices are determined at the intersection between the cumulative demand and supply on an hourly and daily basis. Due to the lack of cost effected storability, unexpected excesses or shortages in demand or supply are reflected almost instantaneously on the electricity prices, leading them into extreme levels. These extreme instantaneous changes in the price levels are termed “price-spikes” in view of their shape in the data series graph: a sequence of sharp upward and downward movements shaped like a spike.

© 2021 Efthymios Stathakis, Theophilos Papadimitriou and Periklis Gogas. Licensed under the Creative Commons Attribution - Noncommercial 4.0 Licence (<http://creativecommons.org/licenses/by-nc/4.0/>). Available at <http://rofea.org>.

Although these events are short-lived, they can result in significant losses for both producers and retailers; a spike forecasting model can mitigate the risk for both sides.

This paper deals with price-spikes that occur in the German electricity market and how they can be forecasted, giving a brief view of the market structure and the basic factors affecting the dynamics of the electricity prices. The German electricity market is the most deregulated, liberalized in EU, largest and most competitive in Europe. Yet, the most volatile, since the entry of the renewable energy sources (RES) in the energy grid, that are causing often imbalances between demand and supply. The German energy exchange market is divided in two centralized control organizations, the European Energy Exchange (EPEX) spot market where wholesalers can buy or sell power in daily, hourly, and quarterly basis, and the European Energy Exchange (EEX) market for derivatives. Focusing in the EPEX Spot market, electricity pricing is a more complex task than the pricing of other commodities: electricity prices are affected by more factors, not all of them controlled by the market forces. These factors can be exogenous, such as, the prices of fossil fuels imports, exchange rates, weather conditions, unexpected shortages or excesses in power supply, and the introduction of the RES that are highly depended from the weather conditions (Sensfuß et al. 2008, Fanone et al. 2013, He et al. 2013), or endogenous as the multiple daily, weekly, and yearly periodicities and the evolution of the price dynamics (autocorrelation and volatility clustering). In the supply-side, the marginal cost of electricity production grows exponentially when the total produced electricity output approaches the full capacity of the power plants (Caves et al., 2000; Geman & Roncoroni, 2006; Boogert & Dupont, 2008).

Empirical findings from various electricity markets suggest that electricity prices share some common characteristics such as mean reversion, long memory, complex seasonality patterns, price jumps, and price spikes. Based on these characteristics, a wide class of Econometric and Machine Learning (ML) models have been employed to describe the electricity market price spikes.

Jump-Diffusion models offer an analytical expression of the electricity price dynamics, and thus are often used for pricing and valuation of Options and Forward contracts. Barlow (2002) examine the ability of Jump-Diffusion models to capture the mean-reversion and the price jump behavior of the prices in the electricity markets of Alberta and California. Cartea & Figueroa (2005), Seifert & Uhrig-Homburg (2007), and Mayer et al. (2012), offer an extended version of Jump-Diffusion models by including a price spike term in their models. As Barlow (2002) stated, the main disadvantage of the Jump-Diffusion models is the difficulties arisen from the need to carefully select and calibrate all the parameters that represent the different price dynamics.

Markov Switching Regime (MSR) models is another class of econometric models that have been proposed to describe the electricity price dynamics. In MSR models, electricity price dynamics are assumed to be governed by different stochastic processes – some states and a transition mechanism that defines the probability to switch among these states. A common setting to model electricity prices is to use two or three regimes MSR models to capture the mean-reverting behavior of the prices, high-low volatility regimes, and price spikes. Huisman & Mahieu (2003), De Jong (2006), Haldrup & Nielsen (2006), Mount et al. (2006), Kanamura & Ohashi (2008), and Janczura & Weron (2010), proposed three regimes MSR models to capture the short-burst behavior of the price spikes. The three regimes of the MRS models represent the mean-reversion behavior of the prices, the other two representing the “Up” and the “Down” movement of the price spikes, provided that the “Down” regime comes immediately after the “Up” regime.

Statistical learning methods were also used to forecast the occurrences of price spikes. Christensen et al. (2012) use an Autoregressive Conditional Hazard (ACH) model and Probit classification model to forecast the occurrences of the price spikes in the half-hour Australian National Energy spot market. Hagfors et al. (2016) develop a Logit model to forecast the occurrences of positive and negative price spikes in the hourly EPEX-spot market.

Machine Learning (ML) methods are known for their ability to capture non-linear dynamics and complex patterns from the data. The requirement for minimal assumptions regarding the underlying processes and the excellent forecasting performance for both classification and regression tasks made the ML methods attractive for many scientific areas. Samitas and Armenatzoglou (2014) examine the ability of the Regression Trees (RTs) and a 2-Regime MSR to capture the extreme behavior of the electricity prices in New England’s spot market. The empirical findings show that the RTs outweighs MSR models in most cases. Dev & Martin (2014) and Sandhu et al. (2016) perform a similar task for the Australian National Energy and Ontario electricity Spot markets using NNs. Binary classifiers have been used to forecast the occurrences of price spikes in several electricity markets. Zhao et al. (2007) deploy a binary classifier based on SVMs to forecast the occurrence of the price spikes in the Australian National Energy Spot market. By the same token Amjady & Keynia (2008) and Fragkioudaki et al. (2015) employ Probabilistic NNs and Decisions Trees respectively for the PJM and Nord Pool Spot markets.

In this paper, a multiclass SVM model is used to forecast the occurrences of the positive and negative prices, along with normal prices in German intraday European Power Exchange (EPEX) Spot electricity market. Some of the advantages of the SVM models are that: a) they work well even with non-linear problems by choosing an appropriate kernel function and projecting the data into higher dimensions where we can efficiently approach any complex

problem, b) unlike NNs, the SVM is not solved for local optima, c) they perform relatively well to high dimensional data, d) the risk of over-fitting is minimized in the cross-validation process. The price spikes are identified under the assumption that the normal prices follow a mean-reverting process, while the price spikes lying in the tails of the price distribution follow a Generalized Pareto Distribution (GPD).

2 Methodology Explanation

The employed methodology has two parts. In the first part, we follow the Byström's (2005) two-step method to identify and analyze the price spikes from the intraday EPEX Spot market in Germany. In the second part, we tackle the multiclassification problem in order to forecast the occurrence of the price spikes and the normal prices.

Byström (2005) investigated the occurrences of price spikes under the framework of the Extreme Value Theory (EVT). In particular, the Peak-Over-Threshold (POT) method was selected to distinguish these short-lived extreme fluctuations from the normal prices. The price spikes in the POT method are treated as excesses over a sufficient high threshold. Under the condition that prices are IID and stationary, the resulted exceedances from the POT method follow a Generalized Pareto Distribution (GPD), while their occurrence follow a Poisson point-process.

To eliminate the time-dependent components from the electricity prices that violate the stationarity and IID assumption, an AR-GARCH model was employed by Byström as a pre-filtering method. The AR part of the proposed model is set to eliminate the linear autocorrelation effects in the error term and the multiple seasonalities from the electricity prices, while the GARCH terms are focused on the heteroscedasticity effects in the price-series, thus, eliminating non-linear processes in the error term.

In the proposed methodology, the Byström's scheme is improved by using the Exponential GARCH, proposed by Fong Chan & Gray (2006), that is able to capture the asymmetric effects of the price shocks in the conditional volatility.

3 Descriptive Statistics Analysis for the EPEX Spot Market

The data from the intraday EPEX ¹spot market covers the period between January 1st, 2013, and June 11th, 2014. In total, 12,648 hourly observations of electricity prices quoted in euros per MWh are used in our sample for this analysis. Before the development of the statistical analysis, a logarithmic transformation was applied to the data. The existence of negative prices in the data sample requires first to shift the prices from the domain \mathbb{R} to the domain $(0, +\infty)$, before applying the logarithmic transformation to the prices.

¹ European Power Exchange <https://www.epexspot.com/>

Table 1 reports some basic descriptive statistics for the shifted log-transformed electricity prices. The negative skewness and the high kurtosis indicate that electricity prices are not normally distributed. The non-normality assumption is also supported by the QQ plot in Figure 1, where the data noticeably deviate from the normal distribution. The Q and Q^2 values in Table 1 denote the statistical values of the Ljung-Box test for the shifted log-transformed, and the squared shifted log-transformed electricity prices. The extremely high values of the Ljung-Box statistics for 12th and 24th hours refer to the “off” and “on” peak periods within the day and the daily seasonality of electricity prices. The squared values of the Ljung-Box test, $Q^2(12)$ and $Q^2(24)$, can be considered an indication of the existence of heteroscedasticity in the electricity prices.

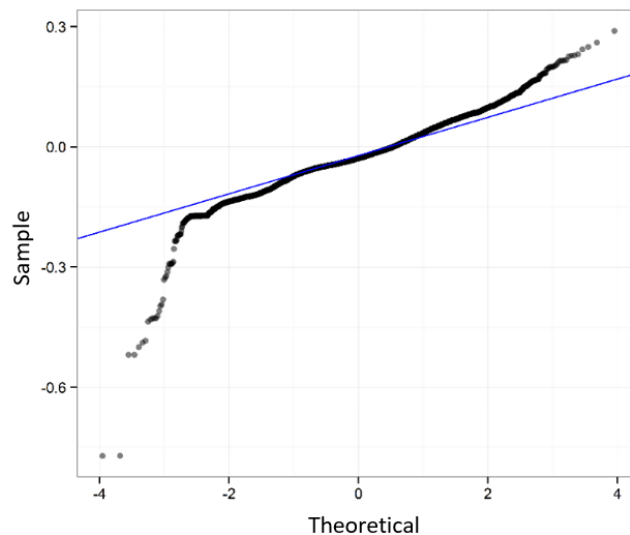
A clearer view regarding the existing seasonalities in the electricity prices is given in Figure 2, with the dominant cycles in the periodogram corresponding to 12 h (half day), 24 h (daily), and 168 h (weekly) periods.

The stationarity assumption is the last requirement that must be fulfilled before deploying the AR-EGARCH model, and eventually, the POT method. The Phillip-Perron (PP), the

Table 1. Descriptive Statistics for the Shifted log Transformed Hourly EPEX Spot Prices

Mean	Median	St. Dev	Skewness	Kurtosis	$Q(12)$	$Q(24)$	$Q^2(12)$	$Q^2(24)$
0.000	0.003	0.120	-97.79	14090	41306	62085	13325	14294

Figure 1. QQ Plot: Comparison of the Actual Data Distribution Against the Normal Distribution



Augmented Dickey-Fuller (ADF), and the Kwiatkowski–Phillips–Schmidt–Shin) (KPSS) are selected as well-known unit-root tests to examine whether the log-transformed electricity price series is stationary. According to the p -values in Table 2, the null hypothesis is rejected for the PP and ADF unit-root tests supporting the stationarity of the log-transformed series. The KPSS test is used under the null hypothesis that the log-transformed electricity prices are level stationary (log-prices oscillated around a constant mean). The reported p -value from the KPSS test rejects the hypothesis of stationarity at the level 5%, and thus, in this case, the time series assumed to be non-stationary.

Figure 2. Power Spectral Density for the Shifted log Transformed Electricity Price Series

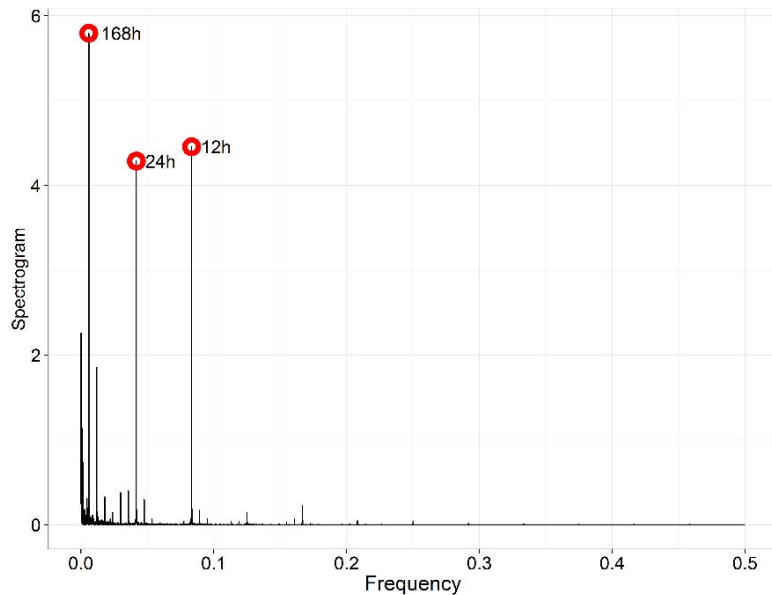


Table 2. ADF, PP, and KPSS Unit-Root Tests for the log-Transformed Prices

	ADF test		PP test		KPSS test	
	Critical Values	Test Statistic	Critical Values	Test Statistic	Critical Values	Test Statistic
1% level	-2.57	-12.36	-3.43	-66.74	0.74	0.95
5% level	-1.94		-2.86		0.45	
10% level	-1.62		-2.56		0.35	
	<i>p-values</i>					
		0.00		0.00		0.01

4 Time Dependency Modeling

Empirical findings from Section 3 indicate the presence of multiple seasonalities and heteroskedasticity in electricity prices. Eliminating time-dependencies from the electricity prices is a prerequisite for the Peak-Over-Threshold (POT) method that assumes that time series is an IID. Therefore, an AR-EGARCH model is employed as a filter method aiming to remove these time-dependencies from the electricity prices. The AR terms in the model aims to eliminate the 12 h, 24 h, and 168 h seasonalities, including the hourly dependencies from the electricity prices. A simple EGARCH(1,1) model is selected to eliminate the conditional volatility from the electricity prices (Nelson, 1991).

The mathematical form of the AR-EGARCH model is presented below:

$$l_t = \varphi_0 + \varphi_1 l_{t-1} + \varphi_{12} l_{t-12} + \varphi_{24} l_{t-24} + \varphi_{168} l_{t-168} + \varepsilon_t$$

$$\ln(\sigma_t^2) = \omega + \beta \ln(\sigma_{t-1}^2) + \gamma \frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \alpha \left[\frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right] \quad (1)$$

Besides the φ_0 parameter which denotes the intercept, the rest of the φ_i parameters, with $i = \{1, 12, 24, 168\}$, are used to capture the seasonal patterns in the electricity prices. The error term ε_t , in the first part of Equation 1, equals to $\sigma_t \eta_t$, where η_t is a Student's t distributed IID random variable defining the error, and σ_t is the conditional variance of ε_t . The second part of Equation 1 represents the EGARCH model of the conditional volatility. The ω parameter is the intercept of the conditional volatility, and α, β, γ are the coefficients related to the evolution of the current conditional volatility given the past conditional volatilities.

The reported AR coefficients in Table 3 suggest a positive correlation between the current and the related seasonal components.

Since the sum of the AR coefficients is less than unity, $\sum \varphi_i \approx 0.83$, this translates as the long-term tendency of electricity prices to return to the mean. The high value of the ν parameter (shape parameter for the Student's t distribution) suggests that the electricity prices distribution rather belongs to a "heavily" tailed distribution family. The positive value of the γ parameter in the EGARCH model implies that positive shocks in the electricity prices have a bigger impact in the conditional volatility than the negative ones.

Standardized residuals fulfil the IID requirement of the POT method better than the raw residuals and thus considered as more adequate for this analysis. Based on the Ljung-Box values in Table 1 and Table 3, the achieved reduction of the autocorrelation effects in electricity prices

Table 3. AR-GARCH Estimated Coefficients and Descriptive Statistics for the Standardized Residuals

Coefficients	Estimates	Robust Standard Errors
φ_0	0.000	0.000
φ_1	0.1075	0.009
φ_{12}	0.042	0.003
φ_{24}	0.3195	0.005
φ_{168}	0.3608	0.006
ω	-3.129	0.142
α	0.093	0.022
β	0.613	0.018
γ	0.701	0.029
ν	3.977	0.098
<i>Mean</i>	0.034	0.000
<i>Median</i>	-0.043	-0.002
<i>St. dev</i>	1.138	0.374
<i>Skewness</i>	-4.781	-104.7
<i>Kurtosis</i>	356.85	1163.78
$Q(12)$	186.39	328.88
$Q(24)$	402.85	587.12
$Q^2(12)^{**}$	115.48	2510.8
$Q^2(24)^{**}$	558.79	2710.4

after applying the AR-EGARCH model is ~99% for both the benchmark points (12 h and 24h). Similarly, the achieved reduction for the heteroskedasticity for the same benchmark points is 96% and 99%.

5 Peak-Over-Threshold method: Tail's quantile estimations

In POT method, the separation of the extreme fluctuations from the normal ones are made by defining a high threshold u , over which the exceedances can be fitted using the GPD. Considering that the residuals R from the AR-EGARCH model follow an IID and stationary process, then the exceedances over a threshold u are defined as $y = R - u$. The distribution of the exceedances above threshold u are assumed to follow an unknown excess distribution $F_u(y)$ that is given by

$$F_u(y) = Pr(R - u \leq y | R > u) = \frac{F_R(u + y) - F_R(u)}{1 - F_R(u)} \quad (2)$$

The threshold u acts as a cut point which separates the parent distribution F_R from the excess distribution F_u . Balkema and de Haan (1974) and Pickands (1975) show that if the threshold u is high enough then the F_u can be approximated by a GP distribution,

$$G_{\xi,\sigma}(y) = \begin{cases} 1 - \left(1 + \frac{\xi}{\alpha}y\right)^{-1/\xi} & \text{if } \xi > 0 \\ 1 - e^{-\frac{y}{\alpha}} & \text{if } \xi = 0 \end{cases} \quad (3)$$

The shape parameter ξ defines the asymptotic decay of the tail, that is, the ‘‘heaviness’’ of the tail, while the α parameter is the scale parameter for the GP distribution. The tail decay is determined as Pareto-type when $\xi > 0$, and exponential when $\xi = 0$. The cumulative distribution for the data points above the threshold u can be written as follows,

$$F_R(u + y) = (1 - F_R(u))F_u(y) + F_R(u) \quad (4)$$

The right side of the Equation 4, can be simplified by replacing the function $F_R(u)$ with $(n - N_u/n)$, where n denoting the total number of observations, and N_u the number of observations above threshold u , while the $u + y$ is rewritten as x . The final form of the simplified formula is presented below.

$$F_R(x) = 1 - \frac{n}{N} \left(1 + \xi \frac{x - u}{\alpha}\right)^{-1/\xi} \quad (5)$$

The estimation of the shape index ξ , and the scale parameter α can be done using the maximum likelihood method in Equation 5. The quantile q_p for a given probability can be now derived by inverting the $F_R(x)$:

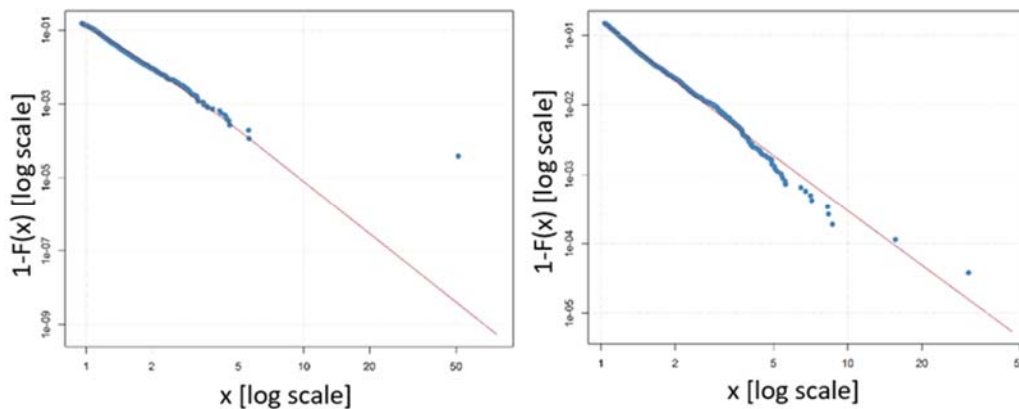
$$q_p = u + \frac{\alpha}{\xi} \left(\left(\frac{n}{N_u} p \right)^{-\xi} - 1 \right) \quad (6)$$

With the standardized residuals from the AR-EGARCH model being closer to an IID process, the POT method can finally employed in order to extract the extreme prices from the price series. The thresholds u_L (left tail) and u_R (right tail) that separates the negative and positive price spikes from the normal prices are selected using the Hill plots (Hill, 1975). The estimated GPD parameters, given the thresholds u_L and u_R , are presented in Table 4. The positive values of the shape indexes ξ_L and ξ_R indicate the existence of ‘‘heavy’’ tails on the both sides of the distribution. The goodness-of-fit of the modeled left and right tails are depicted in Figure 3.

Table 4. Estimated Parameters for the GPD and Tail Quantile Estimations.

Generalized Pareto	Left tail	Right tail
Shape ξ	0.208	0.385
Tail index $\frac{1}{\xi}$	4.81	2.60
Scale α	0.281	0.346
Threshold u	1.09	1.19
Tail quantiles $P(0.95)$	1.17	1.95
Num. of Spikes	708	333

Figure 3. The Unconditional Probability of an Exceedance to Follow the GPD for the Left and the Right Tail of the Standardized Residuals Distribution.

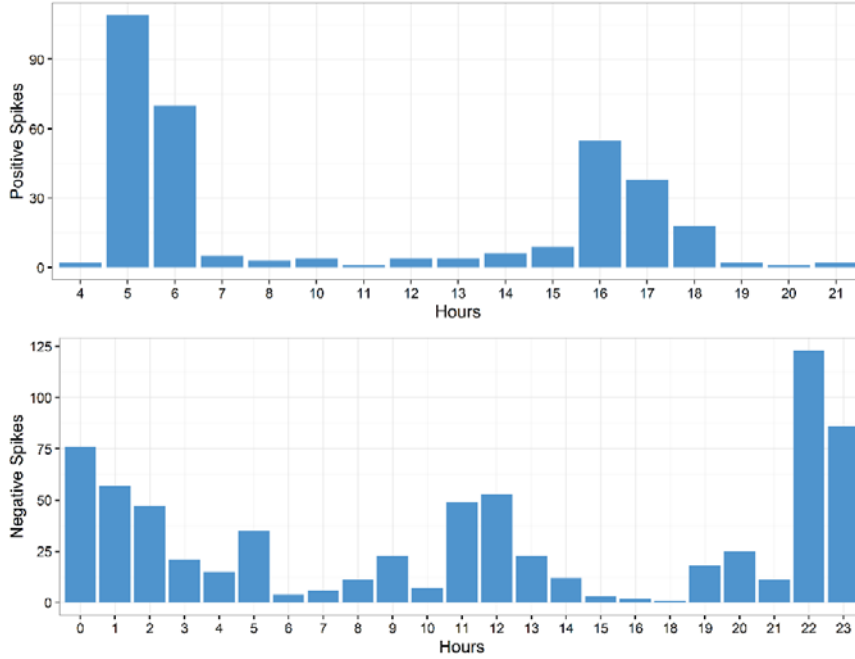


In both cases, the exceedances are closely following the GPD, except one value in the left tail which deviates significantly from the regression line.

The primary assumption in this paper is that the price spikes are lying above the 95th quantile of the GPD. The negative and the positive price spikes represent 5.6% and 2.6% of the sample, with the total number of negative spikes being 708 and the corresponding positive spikes being 333.

The occurrences of the negative and the positive price spikes seems to follow different patterns (Figure 4). Positive spikes are more likely to occur between 5 a.m. - 6 a.m., and 4 p.m. - 6 p.m. In contrast, negative spikes are more likely to occur after the sunset, between 7 p.m. and 5 a.m., when the economic and the human activity are in decline, and between 9 a.m. and 2 p.m.

Figure 4. Occurrences of Negative and Positive Price Spikes within the Day.



6 Support Vector Machines

Support Vector Machines is a supervised machine learning technique for binary classification developed by Cortes and Vapnik (1995). The key idea in the SVMs is to determine a linear separating hyperplane into the input space that preserves the maximum distance between the two classes. SVMs can expand to non-linear classification problems by projecting the data from the input space to a higher dimensional space (feature space) where the two classes are linearly separable. The projection of the data into higher dimensional spaces can be made explicitly or implicitly via kernel functions. The optimization problem in SVMs is expressed in convex form ensuring that the optimal solution can be reached. Also, SVMs can be extended to multi-classification tasks by splitting the multi-class problem into $\frac{k(k-1)}{2}$ binary classification problems.

Considering a set of input vectors $\mathbf{x}_i \in \mathbb{R}^n$ with $i = 1, \dots, I$, belonging to two classes (output targets) labelled $y_i \in \{-1, +1\}$. If the two classes are linearly separable in the input space, then the separator can be defined as

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} - b \quad (7)$$

$$\text{subject to } (y_i(\mathbf{w}^T \mathbf{x}) - b) \geq 1 - \xi_i \tag{8}$$

in such way that $y_i f(\mathbf{x}_i) > 0 \forall i$. In Equation 8, \mathbf{w} denotes the weight vector, b denotes the bias, and ξ is a non-negative slack variable to encounter the erroneously classified instances. The optimization problem can be defined as

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i y_i (x_i w + b) - 1 + \xi_i \tag{9}$$

The first term of the Equation 9 represents the maximal margin of the decision boundary, while the second term represents the minimal error between the output y_i and the prediction $f(\mathbf{x}_i)$. The parameter C in the same side of the equation represents the trade-off between the maximal margin and the minimization of the training error. A more convenient form of the Equation 9 is given by solving the dual Lagrangian problem

$$\mathcal{L}(a) = \max_a \left\{ \sum_{i=1}^I a_i - \sum_{i=1}^I \sum_{j=1}^I a_i a_j y_i y_j \langle x_i, x_j \rangle \right\} \tag{10}$$

subject to $0 \leq a_i \leq C, \forall i$,

$$\sum_{i=1}^I a_i y_i = 0 \tag{12}$$

where the parameter a_i is a non-negative Lagrangian multiplier, that satisfying the Karush–Kuhn–Tucker optimization conditions. The dot product $\langle \cdot \rangle$, in Equation 10, denotes the similarity measure between the points in the input space.

Depending only on the dot product to estimate the optimal solution, SVMs have the flexibility to generalize in different metric spaces. This property can be beneficial in cases where the two classes are not linearly separable in the input space. A convenient method is to use a non-linear mapping function, $\Phi: \mathbb{R}^n \rightarrow \mathbb{R}^m$, to project the data from the input space to a higher dimensional space where the two classes are linearly separable. The embedding of the non-linear mapping can be done easily by substituting \mathbf{x} with $\Phi(\mathbf{x})$ in Equation 10. In SVMs, the notion of similarity measures can be further generalized to arbitrary high dimensional spaces by introducing the Kernel methods. Kernel method provides a set of different measures that can compute the dot product without the exact knowledge of the dimensional space. A very common non-linear Kernel in SVMs is the Radial Basis Function (RBF),

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2\right), \gamma > 0 \quad (13)$$

where γ is an internal tuning hyperparameter of the RBF kernel. For more details see Hofmann, Schölkopf, and Smola (2008).

6.1 Imbalanced Classes

The regularization factor C in error tolerant SVMs controls the margin width between the two classes and the distance of the misclassified points from the hyperplane. In case of imbalanced classes, where the data in two classes are skewed, the optimal hyperplane is also displaced over the majority class. SVMs offer a straightforward solution by introducing two different cost parameters for the two classes. The objective function of the optimization problem can be rewritten as follows:

$$\min_{\mathbf{w}, b, \xi_i} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + C_+ \sum_i \xi_i + C_- \sum_i \xi_i \right\} \quad (14)$$

subject to

$$\begin{aligned} \mathbf{w}^T \mathbf{x}_i - b &> 0 \text{ for } y_i = +1 \\ \mathbf{w}^T \mathbf{x}_i - b &< 0 \text{ for } y_i = -1 \end{aligned} \quad (15)$$

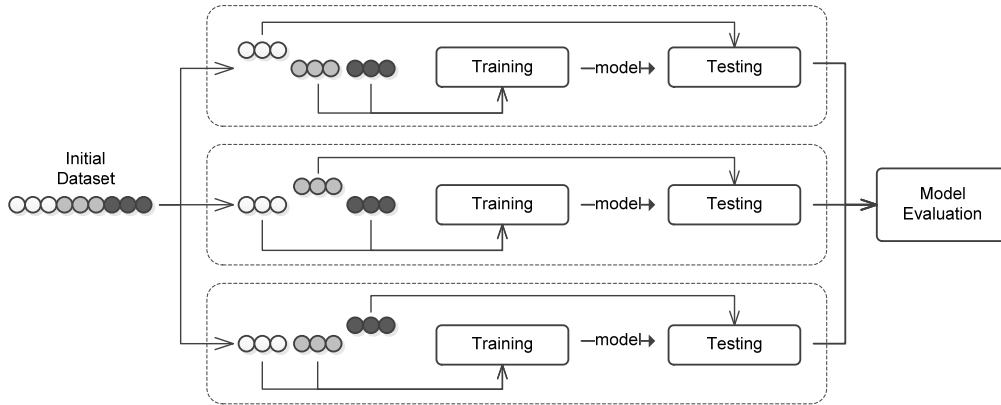
in such as $y_i f(\mathbf{x}_i) > 0 \forall i$. The C_+ and C_- are the different cost parameters for the two classes.

Assigning a bigger misclassification penalty to minority class leads to a solution where the errors of the minority class have bigger impact in the optimal solution.

6.2 Overfitting

In a machine-learning model, training may result in overfitting when the model produced is perfectly fitted to the data sample, instead of the true underlying relationship that describes the phenomenon. Usually, overfitting results in high performance on the training step and a significant accuracy drop on the testing step. Overfitting can be avoided using a k -fold cross validation procedure: the dataset is cut into k subsets of equal size, and the training-testing steps are repeated k times. In each turn, a different subset is used as the test dataset while the rest $k-1$ subsets are grouped together to form the training dataset. The procedure is called dataset folding. The model is evaluated by averaging the performance of the model on every fold. A graphical representation of a 3-fold cross validation procedure is depicted in Figure 5. The model's generalization ability is tested using an out-of-sample set (a subset that did not participate in the cross-validation procedure).

Figure 5. Overview of a 3-fold Cross Validation Scheme.



6.3 Performance Measures

For imbalanced classification problems, simple accuracy measures tend to benefit the majority class over the minority classes. In contrast, Precision, Recall, F1-Score, and G-Mean accuracy scores are employed as more appropriate methods for imbalanced classification problems. The confusion matrix in Table 5 describes the conditional state of every prediction made by a binary classifier. The vertical axis represents the actual states of the classes, while the horizontal axis represents the forecasted states. Given the conditional states, the Precision and Recall can be calculated as follows

Table 5. Confusion Matrix for Binary Classification Problem

	Classified as Positive	Classified as Negative
Real positive	<i>True positive (tp)</i>	<i>False negative (fn)</i>
Real negative	<i>False positive (fp)</i>	<i>True negative (tn)</i>

$$Precision = \frac{tp}{tp + fp} \tag{16}$$

$$Recall = \frac{tp}{tp + fn} \tag{17}$$

The F_β -score and the geometric mean (G-Mean) are usually adopted to integrate Precision and Recall to a single performance measure. The mathematical expressions of F_β -score and G-mean are defined as follows

$$F_\beta \text{ Score} = \frac{(\beta^2 + 1) \text{Precision Recall}}{\beta^2 \text{Precision} + \text{Recall}} \quad (18)$$

$$G - \text{Mean} = \sqrt{\text{Precision Recall}} \quad (19)$$

Where β parameter defines the relationship between precision and recall, when $\beta > 1$ the impact of recall in F_β -score measure is higher than precision, the opposite holds when $\beta < 1$. In the case where $\beta = 1$, the F_β -score can be considered as the harmonic mean between precision and recall.

7 Artificial Neural Networks

The basic idea behind Neural Networks (NNs) is to imitate how the brain neurons operate. NNs are organized in layers committed to perform different functions. Every layer is composed by a finite number of nodes, with each node being fully connected only to the nodes of the other layers. Schematically, the first layer denotes the input variables, and the last layer denotes the output of the decision function, (in our case the choice of class given an input vector). The middle layers, or hidden layers, act as a non-linear transformation mapping into the input space. Considering the number of nodes in every hidden layer and a proper non-linear mapping, NNs can approximate any non-linear function. The output of each hidden node z_j^l can be estimated as follows

$$z_j^l = \psi(\sum_i w_{ji}^l z_i^{l-1} + b_j^l) \quad (20)$$

In Equation 20, the value of the activation function z of the j^{th} neuron in the l^{th} layer is estimated as the weighted sum of the output from the previous activation functions z_i in the $(l-1)$ layer. The ψ function denotes the linear or non-linear transformation of the weighted sum. According to this formula, the objective in the training phase is to find those weights who reproduce better the output value. The training weights are derived through a ‘back-propagation’ iterative method (Rumelhart, Hinton, and Williams, 1986) by updating the weights of the nodes layer-by-layer until the estimated output minimizes the total errors.

8 Gradient Boosted Machines

GBMs is a supervised machine learning method that has been successfully implemented in many classification and regression tasks. For instance, Chang et al. (2016) use GBMs to assess the credit risk for different financial institutions, and Mustika et al. (2019) use multiclass GBMs to forecast the risk in the life insurance. In GBMs, the final model emerges as an ensemble of a large set of weak learners, usually decision trees of fixed size, which form a strong learner. The “Boosted” term implies an iterative learning process wherein each step a weak learner is added to slightly improve the overall performance of the model. In this paper, an extended version of GBMs developed by Chen & Guestrin (2016), called Extreme Gradient Boosted Machines (XGBoost), is used to perform the forecasting of positive and negative price spikes. The key difference between GBMs and XGBoost is that the latter introduces the regularization parameter $\Omega(f)$ that controls complexity of the model.

$$L(\theta) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (21)$$

$$\text{wher } \Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$$

The L function in Equation 21, is a differentiable loss function measuring the difference between the actual values y_i and the forecasted values \hat{y}_i . In the complexity control function $\Omega(f)$, γ is a regularization parameter that controls the number of leaf's attached in each weak learner $f(x)$, and λ is a penalize term for the weights that controls the smoothness of the weak learner.

9 Multiclass SVM Forecasting Performance

The findings from the Extreme Value Theory (EVT) analysis indicate the existence of “heavy” tails on both sides of the filtered electricity prices distribution. The derived positive, and negative price spikes and the normal cases are used to design an SVM multiclassification problem. The proposed SVM model aims to perform one hour-ahead forecasting using as input variables the past values of the log-transformed electricity prices, along with a set of explanatory variables. In addition to the explanatory variables, a set of dummy variables that simulate the 24 hours of the day are also included in the input variable set.

The available input variables set with the corresponding lags are reported in Table 6. The electricity demand and supply data are obtained from the four Transmission System Operators (TSOs): 50Hertz GmbH, Amprion GmbH, Tennet GmbH, and TransnetBW GmbH, which are responsible for the distribution of electricity and for balancing the demand and supply between the producers and consumers.

Table 6. List of Input Variables

Variable name	Lags
Electricity prices (log-prices)	24
Vertical load ($\times 4$)	24
Wind in-feed ($\times 4$)	24
Solar in-feed ($\times 4$)	24
Total Vertical load	24
Total wind in-feed	24
Total solar in-feed	24
Net exports	24
Dummy Hours	-

The input variable Vertical Load closely approximates the generated load by the conventional generation units (i.e. coal, gas, and nuclear), excluding the in-feeds from wind and solar power units (Wagner, 2014). Both Wind and Solar in-feed are related to the electricity production from the RES. Net Exports or Imports correspond to the total amount of the exchanged electricity between Germany and the neighbor countries. Additionally, total vertical load and total wind and solar in-feed are included in the input variable set.

9.1 Tuning and Evaluation of the SVM Hyperparameters

The RBF kernel is selected as a robust non-linear kernel function for the SVM model that is capable to efficiently model the complex relations between the data. In SVMs, the parameter space is defined by the RBF hyperparameter γ (Equation 13), and the regularization parameter C . The optimal set of parameters $C \times \gamma$ are determined through a Coarse-to-Fine grid search method in coordination with a 10-fold Cross-Validation procedure. The imbalanced classes problem introduces an extra layer of complexity in the final solution. SVMs provide a trivial solution that incorporates different weights for each class (Equation 14) that aim to mitigate the effect of the majority class in the determination of the classification hyperplane.

9.2 Feature Selection Procedure

The rule-of-thumb in Machine Learning methods dictates to keep 80% of the data sample for the training purposes, and the remaining 20% to test the generalization ability of the model in the unknown data. In the training phase, a total length of 1200 observations, between November 2, 2013, and December 22, 2013, are used to find the optimal set of the model's hypermeter, and in the same time, to select the best set of the input variables that gives the highest forecasting score.

Table 7. Highest Forecasting Scores for the Selected SVM Models

	Added variable	Lags	F_1 -Score	G-Mean	Precision	Recall
Step 1	Log-prices	7	0.70	0.70	0.81	0.62
Step 2	Vertical load (TransnetBW)	1	0.72	0.72	0.79	0.66
Step 3	Total Wind in-feed	3	0.77	0.78	0.76	0.78

The adopted feature selection method involves an inflated procedure where the best variable set is expanded by testing all the available input variables successively. In this procedure, the rejection criterion for a candidate variable is whether it increases the in-sample accuracy. Then, the generalization ability of the SVM models is tested using a total length of 240 unknown data, covering a period from December 23, 2013, to January 2, 2014.

The feature selection procedure starts by locating the optimal number of the lags for the shifted log-transformed electricity prices, and subsequently by expanding the input set with the explanatory variables. The final input set after the successive steps of the feature selection procedure is presented in Table 7, with the last step representing the SVM model with the highest forecasting scores.

The SVM model derived from the first step of the feature selection procedure resembles a seven-order AR(7) autoregressive process. The selected explanatory variables imply that the demand data from TransnetBW TSO area and the Total Wind in-feed have positively affected the overall forecasting accuracy of the SVM model. The overall F_1 -Score 0.77 and G-Mean score 0.78, indicates that the model can provide satisfactory forecasts regarding the occurrence of the price spikes.

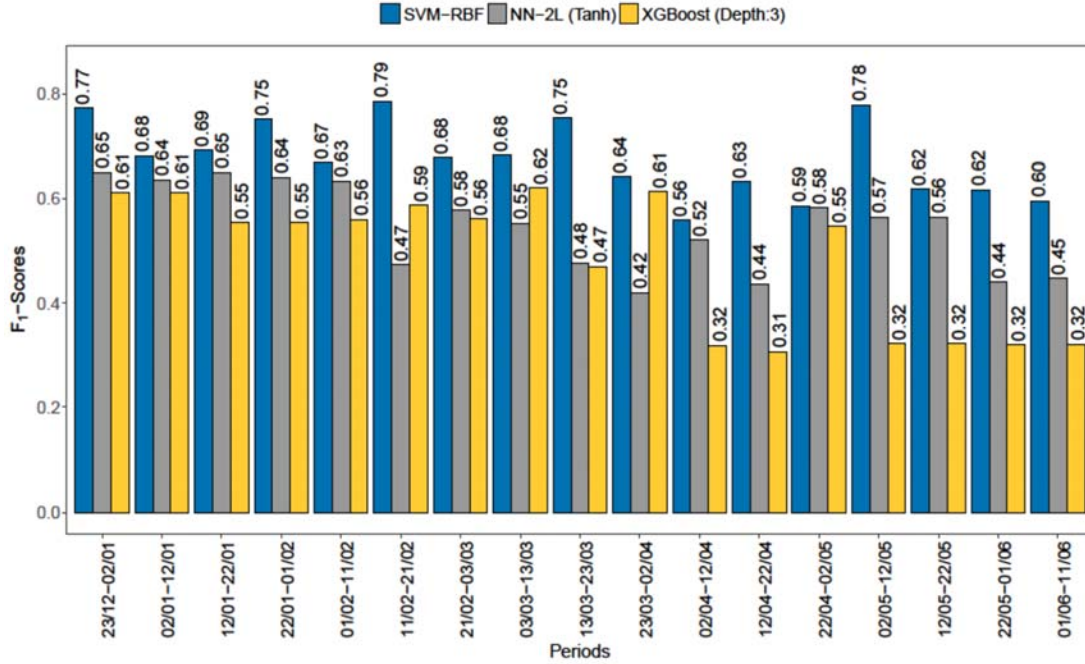
9.3 Rolling Window Evaluation

To further investigate the ability of the SVM model to sufficiently forecast the occurrences of the price spikes, a rolling window evaluation method is introduced. In every step, the models are re-trained and re-evaluated in the unknow data set by shifting both the train and the out-of-sample test set 240 hours. In total, the proposed SVM model is re-trained and tested in the Out-of-Sample data 17 times. The overall length of the Out-of-Sample data covers a period from January 2, 2014, to June 11, 2014, or 4080 hours.

9.4 Empirical Results

Besides the proposed multiclass SVM, two additional ML models are introduced to serve as benchmark models. A two-layer NN equipped with a hyperbolic tangent (Tanh) as non-linear activation function, and an Extreme Gradient Boosted machines (XGBoost) model. An inflated feature selection method similar to the SVMs adopted for NNs models. In the case of the XGBoost models, a different feature selection method was selected.

Figure 6. Average F_1 -Scores for the 17 Test Periods



In XGBoost models, every input variable is checked and ranked for its contribution to the forecasting performance of the model. The final feature selection made by choosing the top fifteen input features with the most significant contribution in the forecasting score. The resulted F_1 -scores indicate that the SVM models surpasses both XGBoost and NN-2L based models in every test period. In particular, the F_1 -Scores for the SVM model range from a minimum of 0.56 to a maximum of 0.79. In comparison, the lowest and the highest achieved F_1 -scores for the NN-2L are 0.56 and 0.42, respectively. The XGBoost models produced the lowest F_1 -scores compared to the SVMs and the NNs models, with the achieved F_1 -scores range from 0.31 to 0.62.

A closer look at the per-class accuracies in Table 8 indicates that both NN-2L and XGBoost forecasting models are often incapable of forecasting the occurrences of price spikes. The underlined per class accuracies denote the times in which the models failed to pass the accuracy threshold of 0.50. In overall, the multiclass SVM model fails to cross the 0.50 limit only once, with the average per class accuracy be 0.86 for the normal prices, 0.75 for the positive spikes, and 0.74 for the negative spikes. In contrast, the NN-2L model fails to pass the 0.50 threshold 16 times for both positive and negative spikes. The XGBoost models failed to exceed the barrier of 0.50 threshold in all the 17 test-periods for both positive and negative price spikes.

Table 8. Per Class Accuracies Among 17 Periods.

Date	SVM			NN-2L (Tanh)			XGBoost (Depth:3)		
	Normal	Pos	Neg	Normal	Pos	Neg	Normal	Pos	Neg
23/12 - 02/01	0.88	0.71	0.70	0.88	0.60	0.66	1.00	<u>0.40</u>	<u>0.19</u>
02/01 - 12/01	0.89	0.67	0.69	0.92	0.79	<u>0.36</u>	0.99	<u>0.36</u>	<u>0.09</u>
12/01 - 22/01	0.89	0.86	0.64	0.93	0.59	<u>0.50</u>	1.00	<u>0.35</u>	<u>0.13</u>
22/01 - 01/02	0.93	0.81	0.75	0.86	0.64	0.65	0.99	<u>0.14</u>	<u>0.18</u>
01/02 - 11/02	0.79	0.71	0.82	0.82	0.67	0.68	0.99	<u>0.13</u>	<u>0.16</u>
11/02 - 21/02	0.92	0.73	0.84	0.85	<u>0.31</u>	<u>0.32</u>	1.00	<u>0.25</u>	<u>0.05</u>
21/02 - 03/03	0.90	0.92	0.37	0.89	0.63	<u>0.33</u>	1.00	<u>0.31</u>	<u>0.08</u>
03/03 - 13/03	0.88	0.87	0.58	0.70	0.83	0.73	0.96	<u>0.33</u>	<u>0.21</u>
13/03 - 23/03	0.93	0.71	0.70	0.86	<u>0.33</u>	<u>0.50</u>	0.99	<u>0.23</u>	<u>0.25</u>
23/03 - 02/04	0.84	0.83	0.75	0.93	<u>0.00</u>	<u>0.29</u>	1.00	<u>0.00</u>	<u>0.12</u>
02/04 - 12/04	0.82	<u>0.40</u>	0.76	0.81	<u>0.50</u>	0.62	0.98	<u>0.50</u>	<u>0.09</u>
12/04 - 22/04	0.82	1.00	0.65	0.91	<u>0.33</u>	<u>0.24</u>	0.99	<u>0.00</u>	<u>0.00</u>
22/04 - 02/05	0.80	0.67	0.82	0.86	0.67	0.54	0.99	<u>0.00</u>	<u>0.00</u>
02/05 - 12/05	0.90	1.00	0.89	0.91	<u>0.43</u>	<u>0.50</u>	1.00	<u>0.14</u>	<u>0.10</u>
12/05 - 22/05	0.88	0.57	0.80	0.86	0.67	0.54	1.00	<u>0.00</u>	<u>0.00</u>
22/05 - 01/06	0.83	0.67	0.85	0.91	<u>0.00</u>	0.54	1.00	<u>0.00</u>	<u>0.00</u>
01/06 - 11/06	0.81	0.67	0.92	0.89	<u>0.00</u>	0.53	1.00	<u>0.00</u>	<u>0.00</u>
Average	0.86	0.75	0.74	0.87	0.47	0.50	0.99	0.19	0.10

10 Conclusions

In this paper, we presented a multiclass SVM model to forecast the positive and negative price spikes, as well as the normal prices in the German electricity market. The price spikes were identified using a theoretical framework based on EVT. The derived price spikes dataset is highly imbalanced with the positive and the negative price spikes representing only 2.6% and 5.6% of the total sample.

The final input variable set derived from the feature selection procedure implies that demand data and Wind in-feed data positively affect the overall accuracy of the SVM model. This finding seems to be in line with the literature suggesting the contribution of RES in the increase in the electricity price volatility.

The results from the rolling window process show that the multiclass SVM model can provide satisfactory forecasts for all the three classes. In contrast, both NNs and XGBoost models failed to classify the positive and negative price spikes correctly.

References

- Amjady, Nima, and Farshid Keynia. 2008. "Day Ahead Price Forecasting of Electricity Markets by a Mixed Data Model and Hybrid Forecast Method." *International Journal of Electrical Power & Energy Systems* 30 (9): 533–46. <https://doi.org/10.1016/j.ijepes.2008.06.001>.
- Balkema, A. A., and L. de Haan. 1974. "Residual Life Time at Great Age." *The Annals of Probability*, 792–804. <https://doi.org/10.1214/aop/1176996548>.
- Barlow, M T. 2002. "A Diffusion Model for Electricity Prices." *Math. Fin.* 12 (4): 287–98.
- Boogert, Alexander, and Dominique Dupont. 2008. "When Supply Meets Demand: The Case of Hourly Spot Electricity Prices." *Power Systems, IEEE Transactions On*, 1–24. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4484956.
- Byström, Hans N E. 2005. "Extreme Value Theory and Extremely Large Electricity Price Changes." *International Review of Economics and Finance* 14 (1): 41–55. [https://doi.org/10.1016/S1059-0560\(03\)00032-7](https://doi.org/10.1016/S1059-0560(03)00032-7).
- Cartea, Álvaro, and Marcelo G. Figueroa. 2005. "Pricing in Electricity Markets: A Mean Reverting Jump Diffusion Model with Seasonality." *Applied Mathematical Finance* 12 (4): 313–35. <https://doi.org/10.1080/13504860500117503>.
- Caves, Douglas, Kelly Eakin, and Ahmad Faruqui. 2000. "Mitigating Price Spikes in Wholesale Markets through Market-Based Pricing in Retail Markets." *The Electricity Journal* 13 (3): 13–23. [https://doi.org/10.1016/S1040-6190\(00\)00092-0](https://doi.org/10.1016/S1040-6190(00)00092-0).
- Chang, Yung Chia, Kuei Hu Chang, and Guan Jhih Wu. 2018. "Application of EXtreme Gradient Boosting Trees in the Construction of Credit Risk Assessment Models for Financial Institutions." *Applied Soft Computing Journal* 73 (December): 914–20. <https://doi.org/10.1016/j.asoc.2018.09.029>.
- Chen, Tianqi, and Carlos Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." <https://doi.org/10.1145/2939672.2939785>.
- Christensen, T. M., A. S. Hurn, and K. A. Lindsay. 2012. "Forecasting Spikes in Electricity Prices." *International Journal of Forecasting* 28 (2): 400–411. <https://doi.org/10.1016/j.ijforecast.2011.02.019>.
- Cortes, C, and V Vapnik. 1995. "Support-Vector Networks." *Machine Learning* 297: 273–97. <http://link.springer.com/article/10.1007/BF00994018>.
- Dev, Priya, and Michael A. Martin. 2014. "Using Neural Networks and Extreme Value Distributions to Model Electricity Pool Prices: Evidence from the Australian National Electricity Market 1998–2013." *Energy Conversion and Management* 84: 122–32. <https://doi.org/10.1016/j.enconman.2014.04.012>.
- Fanone, Enzo, Andrea Gamba, and Marcel Prokopczuk. 2013. "The Case of Negative Day-Ahead Electricity Prices." *Energy Economics* 35 (May 2008): 22–34. <https://doi.org/10.1016/j.eneco.2011.12.006>.

- Fong Chan, Kam, and Philip Gray. 2006. "Using Extreme Value Theory to Measure Value-at-Risk for Daily Electricity Spot Prices." *International Journal of Forecasting* 22 (2): 283–300. <https://doi.org/10.1016/j.ijforecast.2005.10.002>.
- Fragkioudaki, Anna, Adamantios Marinakis, and Rachid Cherkaoui. 2015. "Forecasting Price Spikes in European Day-Ahead Electricity Markets Using Decision Trees." *International Conference on the European Energy Market, EEM 2015-Augus*. <https://doi.org/10.1109/EEM.2015.7216672>.
- Geman, Hélyette, and Andrea Roncoroni. 2006. "Understanding the Fine Structure of Electricity Prices." *The Journal of Business* 79 (3): 1225–61. <https://doi.org/10.1086/500675>.
- Hagfors, Lars Ivar, Hilde Hørthe Kamperud, Florentina Paraschiv, Marcel Prokopczuk, Alma Sator, and Sjur Westgaard. 2016. "Prediction of Extreme Price Occurrences in the German Day-Ahead Electricity Market." *Quantitative Finance* 16 (12): 1929–48. <https://doi.org/10.1080/14697688.2016.1211794>.
- Haldrup, Niels, and Morten Ørregaard Nielsen. 2006. "A Regime Switching Long Memory Model for Electricity Prices." *Journal of Econometrics* 135 (1–2): 349–76. <https://doi.org/10.1016/j.jeconom.2005.07.021>.
- He, Yang, Marcus Hildmann, Florian Herzog, and Goran Andersson. 2013. "Modeling the Merit Order Curve of the European Energy Exchange Power Market in Germany." *IEEE Transactions on Power Systems* 28 (3): 3155–64. <https://doi.org/10.1109/TPWRS.2013.2242497>.
- Hofmann, Thomas, Bernhard Schölkopf, and Alexander J. Smola. 2008. "Kernel Methods in Machine Learning." *The Annals of Statistics* 36 (3): 1171–1220. <https://doi.org/10.1214/009053607000000677>.
- Huisman, Ronald, and Ronald Mahieu. 2003. "Regime Jumps in Electricity Prices." *Energy Economics* 25 (5): 425–34. [https://doi.org/10.1016/S0140-9883\(03\)00041-0](https://doi.org/10.1016/S0140-9883(03)00041-0).
- Janczura, Joanna, and Rafal Weron. 2010. "An Empirical Comparison of Alternate Regime-Switching Models for Electricity Spot Prices." *Energy Economics* 32 (5): 1059–73. <https://doi.org/10.1016/j.eneco.2010.05.008>.
- Jong, Cyriel De. 2006. "The Nature of Power Spikes: A Regime-Switch Approach." *Studies in Nonlinear Dynamics & Econometrics* 10 (3). <https://doi.org/10.2202/1558-3708.1361>.
- Kanamura, Takashi, and Kazuhiko Ohashi. 2008. "On Transition Probabilities of Regime Switching in Electricity Prices." *Energy Economics* 30 (3): 1158–72. <https://doi.org/10.1016/j.eneco.2007.07.011>.
- Mayer, Klaus, Thomas Schmid, and Florian Weber. 2012. "Modeling Electricity Spot Prices: Combining Mean Reversion, Spikes, and Stochastic Volatility." *The European Journal of Finance* 4364 (December): 1–24. <https://doi.org/10.1080/1351847X.2012.716775>.

- Mount, Timothy D., Yumei Ning, and Xiaobin Cai. 2006. "Predicting Price Spikes in Electricity Markets Using a Regime-Switching Model with Time-Varying Parameters." *Energy Economics* 28 (1): 62–80. <https://doi.org/10.1016/j.eneco.2005.09.008>.
- Mustika, Widya Fajar, Hendri Murfi, and Yekti Widyaningsih. 2019. "Analysis Accuracy of XGBoost Model for Multiclass Classification - A Case Study of Applicant Level Risk Prediction for Life Insurance." In *Proceeding - 2019 5th International Conference on Science in Information Technology: Embracing Industry 4.0: Towards Innovation in Cyber Physical System, ICSITech 2019*, 71–77. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ICSITech46713.2019.8987474>.
- Nelson, Daniel B. 1991. "Conditional Heteroskedasticity in Asset Returns: A New Approach." *Econometrica*. <https://doi.org/10.2307/2938260>.
- Pickands, James. 1975. "Statistical Inference Using Extreme Order Statistics." *The Annals of Statistics* 3: 119–31.
- Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. 1986. "Learning Representations by Back-Propagating Errors." *Nature* 323 (6088): 533–36. <https://doi.org/10.1038/323533a0>.
- Samitas, Aristeidis, and Aggelos Armenatzoglou. 2014. "Regression Tree Model versus Markov Regime Switching: A Comparison for Electricity Spot Price Modelling and Forecasting." *Operational Research* 14 (3): 319–40. <https://doi.org/10.1007/s12351-014-0149-6>.
- Sandhu, Harmanjot Singh, Liping Fang, and Ling Guan. 2016. "Forecasting Day-Ahead Price Spikes for the Ontario Electricity Market." *Electric Power Systems Research* 141: 450–59. <https://doi.org/10.1016/j.epsr.2016.08.005>.
- Seifert, Jan, and Marliese Uhrig-Homburg. 2007. "Modelling Jumps in Electricity Prices: Theory and Empirical Evidence." *Review of Derivatives Research* 10 (1): 59–85. <https://doi.org/10.1007/s11147-007-9011-9>.
- Sensfuß, Frank, Mario Ragwitz, and Massimo Genoese. 2008. "The Merit-Order Effect: A Detailed Analysis of the Price Effect of Renewable Electricity Generation on Spot Market Prices in Germany." *Energy Policy* 36 (8): 3076–84. <https://doi.org/10.1016/j.enpol.2008.03.035>.
- Wagner, Andreas. 2014. "Residual Demand Modeling and Application to Electricity Pricing." *Energy Journa* 35 (2): 1994–9089. <https://doi.org/10.5547/01956574.35.2.3>.
- Zhao, Jun Hua, Zhao Yang Dong, Xue Li, and Kit Po Wong. 2007. "A Framework for Electricity Price Spike Analysis with Advanced Data Mining Methods." *IEEE Transactions on Power Systems* 22 (1): 376–85. <https://doi.org/10.1109/TPWRS.2006.889139>.