Outlier Treatment and Robust Approaches for Modeling Electricity Spot Prices

Stefan Trück[1], Rafał Weron[2] and Rodney Wolff [3]

ISI 2007, Lisbon
August 25, 2007

[1]Department of Economics, Macquarie University, Sydney
[2]Hugo Steinhaus Center, Wrocław University of Technology
[3]School of Economics and Finance, Queensland University of Technology, Brisbane

Overview

- Introduction
- Models for Spike Detection
- Empirical Results
- Conclusions

Introduction

Stylized Facts

Electricity is a flow commodity and subject to

- capacity limits
- transmission losses

There are **very limited arbitrage** possibilities

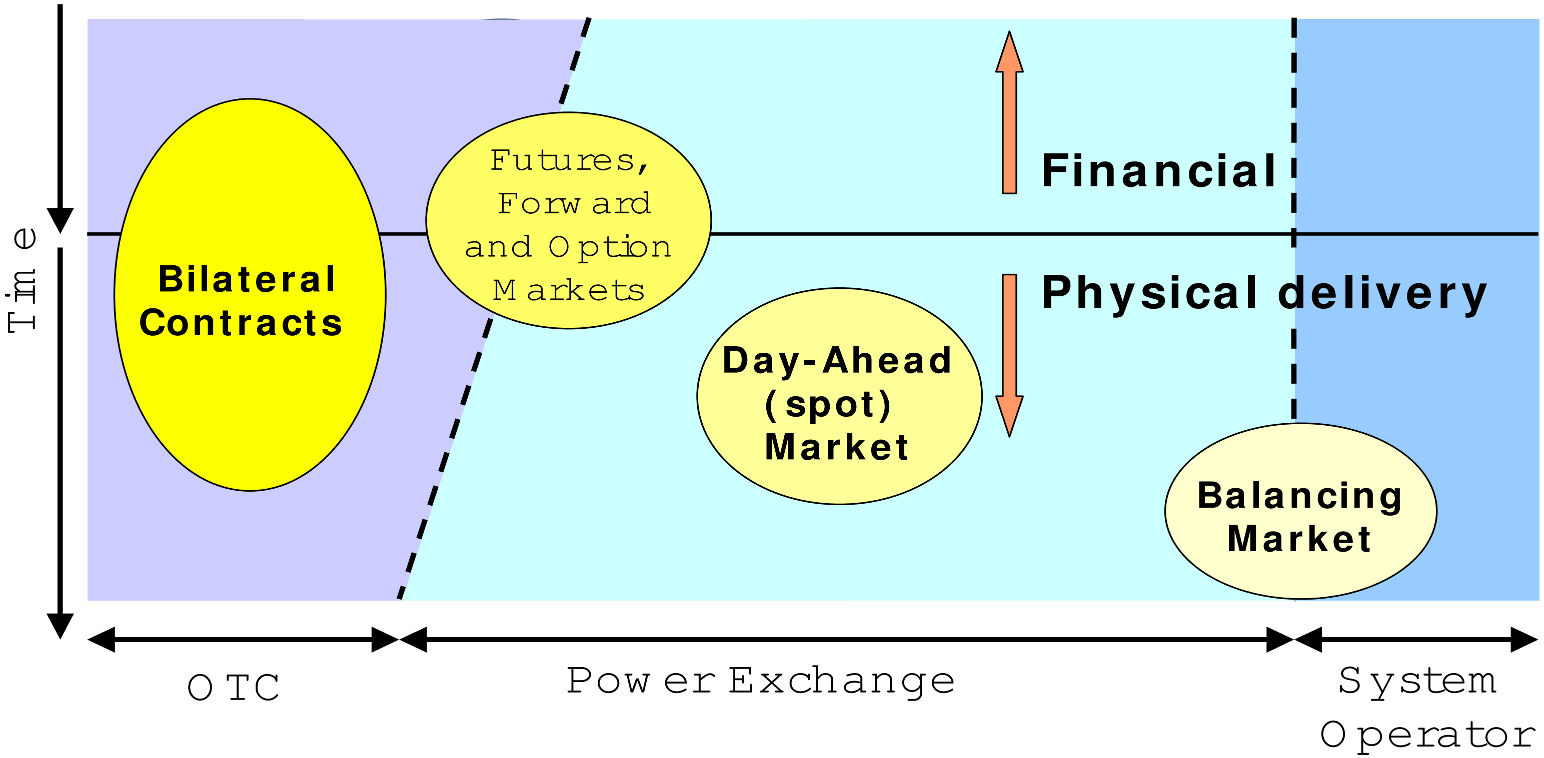
- across time
- across markets

Prices strongly depend on **short-term**

- demand and
- supply

Introduction

Wholesale Power Market Structure



Introduction

Spot Price Behavior

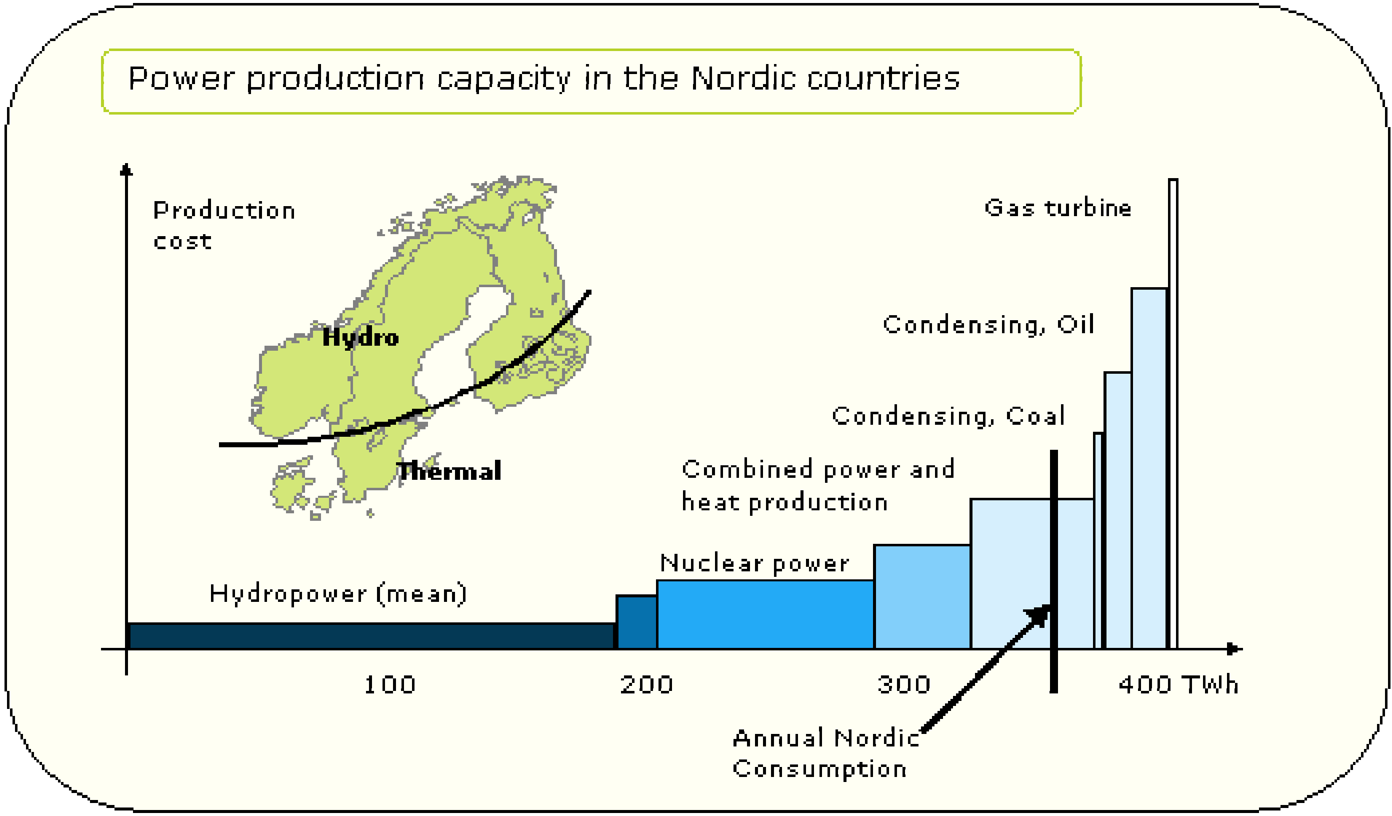
Electricity spot prices typically exhibit features like:

- Seasonality
- Mean-Reversion
- **Price jumps and spikes**
- High volatility

$\Rightarrow$ Classic financial spot price models have to be adapted for electricity prices.

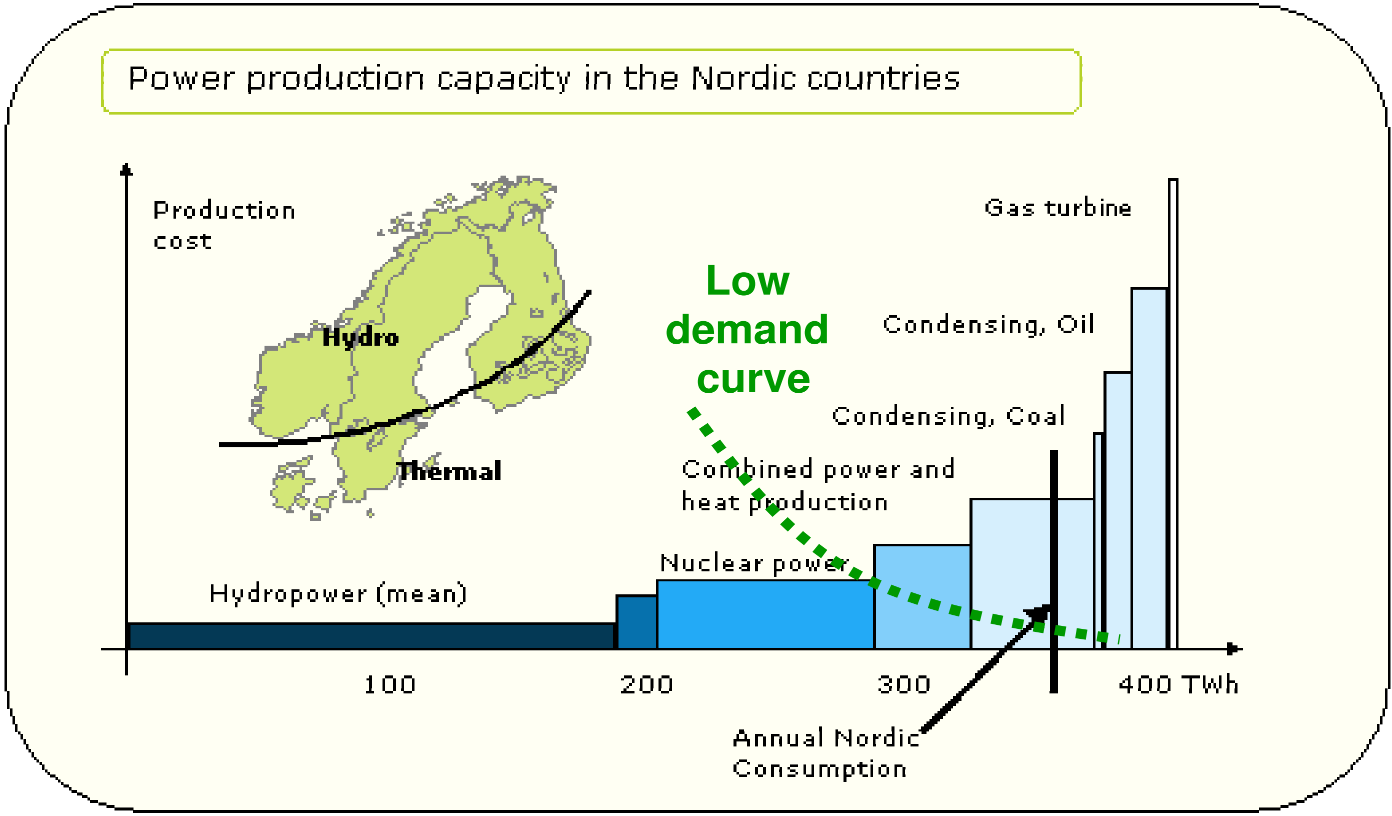
Introduction

Supply Stack and the Market Cross



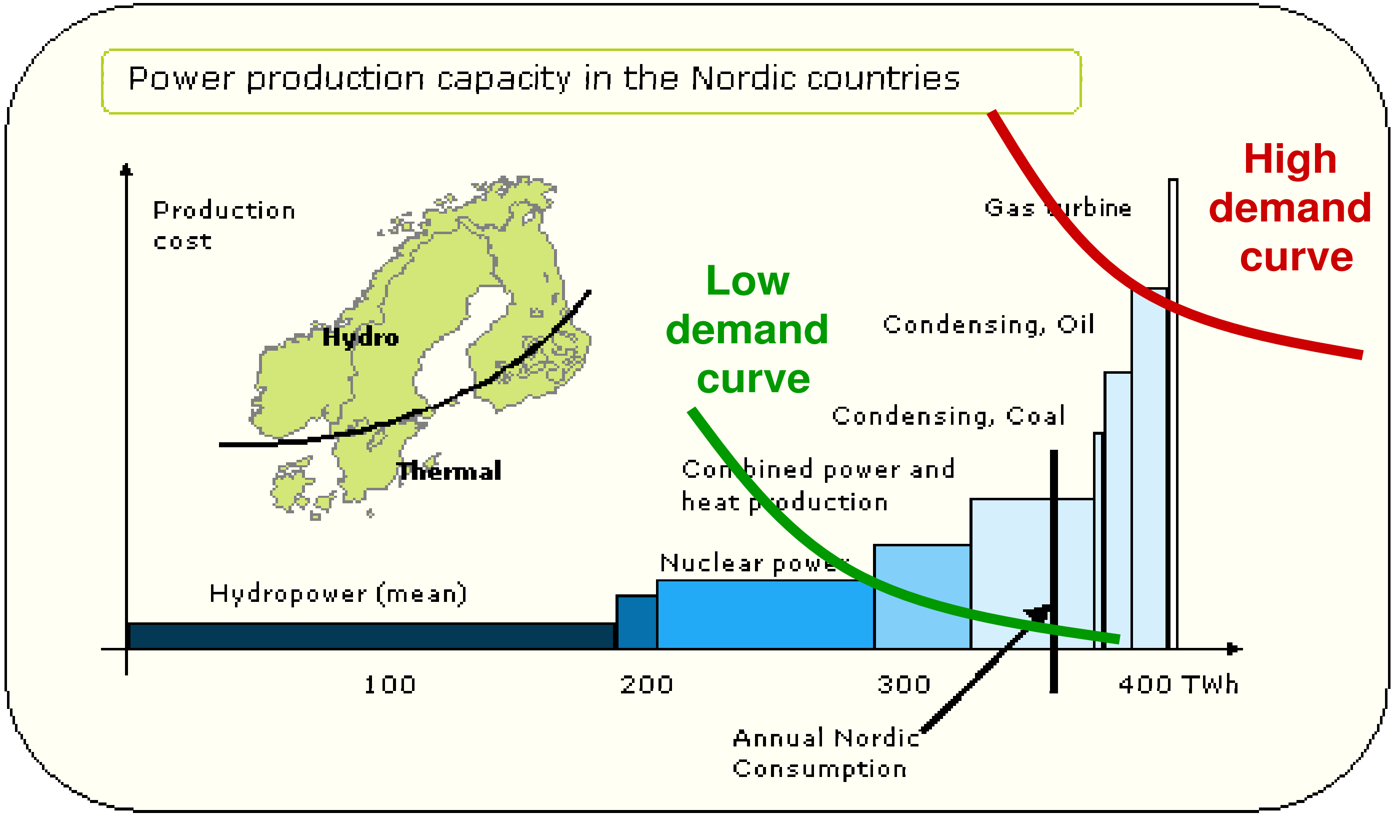
Introduction

Supply Stack and the Market Cross



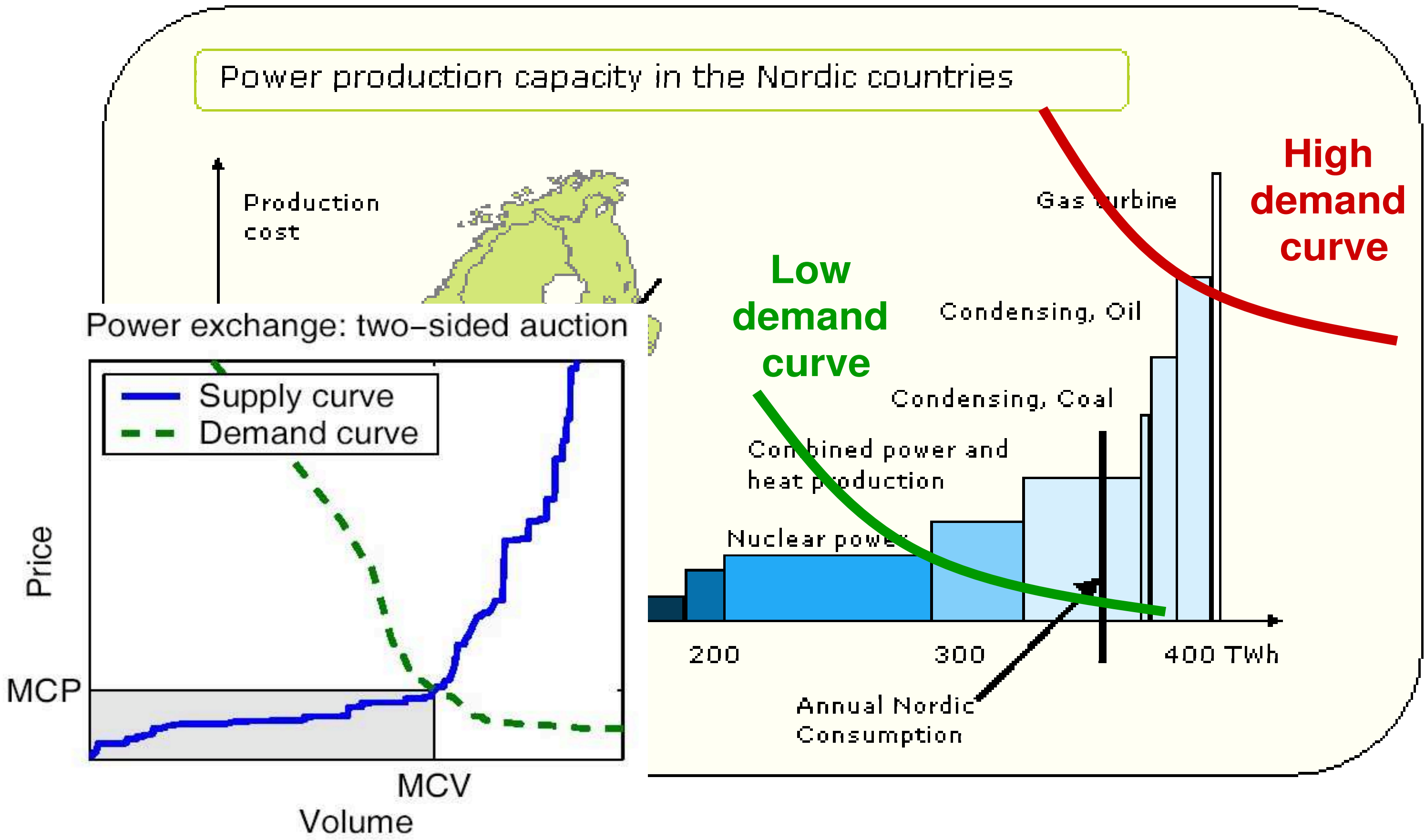
Introduction

Supply Stack and the Market Cross



Introduction

Supply Stack and the Market Cross



Introduction

Model Estimation

- the system (spot) price is generally split into a deterministic part $f(t)$ that comprises all kinds of seasonal behavior and a purely stochastic component S_t (Pilipovic, 1997; Lucia and Schwartz, 2002; Weron, 2006)
- before estimating the stochastic model for S_t, usually a seasonal model for the deterministic part $f(t)$ is calibrated
- preprocessing the data in terms of detecting price spikes might improve the estimation of the seasonal pattern (Bierbrauer, 2007)

$\longrightarrow$ **What is the influence of the chosen technique for spike / outlier detection on the further estimation process?**

Models for Spike Detection

Detecting Price Spikes

- fixed price thresholds (Lapuerta and Moselle, 2001)
- fixed price change thresholds, e.g., log-price increments or returns exceeding 30% (Bierbrauer et al., 2004)
- variable price change thresholds, e.g., log-price increments or returns exceeding three standard deviations of all price changes (Cartea and Figueroa, 2005; Clewlow and Strickland, 2000)
- use of wavelet decomposition to filter out the spikes (Stevenson, 2001; Weron, 2006)

$\longrightarrow$ **Obviously, different definitions and techniques may lead to quite different results and identification of price spikes.**

Models for Spike Detection

Replacing the Outliers

The chosen technique for replacement will also affect parameter estimates, e.g., of the seasonal pattern if the estimation is conducted using the new series. Suggestions in the literature:

- replace the observed outliers by a chosen threshold (Shahidehpour et al, 2002)
- replace the extreme observations by the mean of the two neighboring prices (Weron, 2007)
- by one of the neighboring prices (Geman and Roncoroni, 2006)
- replace outliers by the median of all prices having the same weekday and month as the outlier (Bierbrauer et al, 2007)

Models for Spike Detection

Detrending (7-day Period)

- estimate trend by applying a moving median filter to vector of daily prices:

$$\hat{m}_t = median(x_{t-3}, .., x_t, .., x_{t+3}), \qquad t = 4, ..., 2188. \tag{1}$$

- estimate the seasonal component by computing the average w_k of the deviations $\{(x_{k+7j} - \hat{m}_{k+7j}), 3 < k + 7j \leq 2188\}$ for each $k = 1, .., 7$.
- estimate the seasonal component s_k as

$$\hat{s}_k = w_k - \frac{1}{7}\sum_{i=1}^{7} w_i, \tag{2}$$

- The deseasonalized (with respect to the 7-day period) data is then defined as $y_t = x_t - \hat{s}_t$ for $t = 1, ..., 2191$

Models for Spike Detection

Detrending (Long-Term Cycle)

- in a second step to take care of the trend in the data another moving average filter with different length can be applied
- a length of 31 days was chosen, hence the new moving median component for each observation was calculated by
  $\hat{m}_{2,t} = median(y_{t-15}, ..., y_{t+15}), t = 16, ..., 2176$
- consider the difference between the deseasonalized series and the moving median $y_t - m_{2,t}$ to identify outliers

Models for Spike Detection

Fixed and Percentage Price Thresholds

Probably the simplest technique to detect outliers is the use of **fixed** or **percentage** price thresholds:

- note that the choice of the levels themselves is non-trivial and rather arbitrary
- for the fixed price threshold, we chose to classify all prices beyond 75 EUR/MWh as extreme observations
- for the percentage price threshold, we considered the largest 1% of the observations as outliers
- we consider two different approaches:
  - use original observations
  - remove the weekly (7-day) seasonality

Empirical Results

Recursive Filter

Initially suggested in Clewlow and Strickland (2000), a recursive filter can be applied to identify price jumps in the sample distribution of daily returns:

- iterative procedure that is repeated until no more jumps can be identified
- first step: calculate the sample standard deviation $\hat{s}$ of the returns
- identify returns beyond a certain range – measured in multiples of $\hat{s}$ – as extreme returns, e.g. choose three standard deviations
- replace the outliers according to some specified rule
- after replacing the extreme observations, the next iteration is performed

Empirical Results

Wavelet Approximation

- Any function or signal (here: the spot price series) can be built up as a sequence of projections onto one father wavelet and a sequence of mother wavelets,

$$f(t) = S_J + D_J + D_{J-1} + \ldots + D_1, \tag{3}$$

  where 2^J is the maximum scale sustainable by the number of observations.
- at the coarsest scale the signal can be estimated by S_J; at each step, by adding a mother wavelet D_j of a lower scale $j = J-1, J-2, \ldots$, we obtain a better estimate of the original signal (lowpass filtering)
- we use the S_3 and S_5 approximations, roughly corresponding to weekly ($2^3 = 8$ days) and monthly ($2^5 = 32$ days) smoothing
- the chosen approximation (S_3 or S_5) is subtracted from the original price series: outliers are identified as the observations exceeding three standard deviations of the differences

Empirical Results

The Data

- data from the European Energy Exchange (EEX) in Leipzig, EEX exhibits high liquidity also in the spot market
- we consider the Phelix base *day* index, an equally weighted average of all 24 hourly spot prices for a particular day
- six years of daily prices from January 1, 2001 - December 31, 2006, totaling 2191 observations
- obvious positive trend in the data
- several spikes can be observed in the time series, e.g. spot prices peaked
  - in December 2001 with a daily average of 240 EUR/MWh
  - in January 2003 with 163 EUR/MWh and
  - in July and November 2006 with 301 and 162 EUR/MWh

Empirical Results

The Data

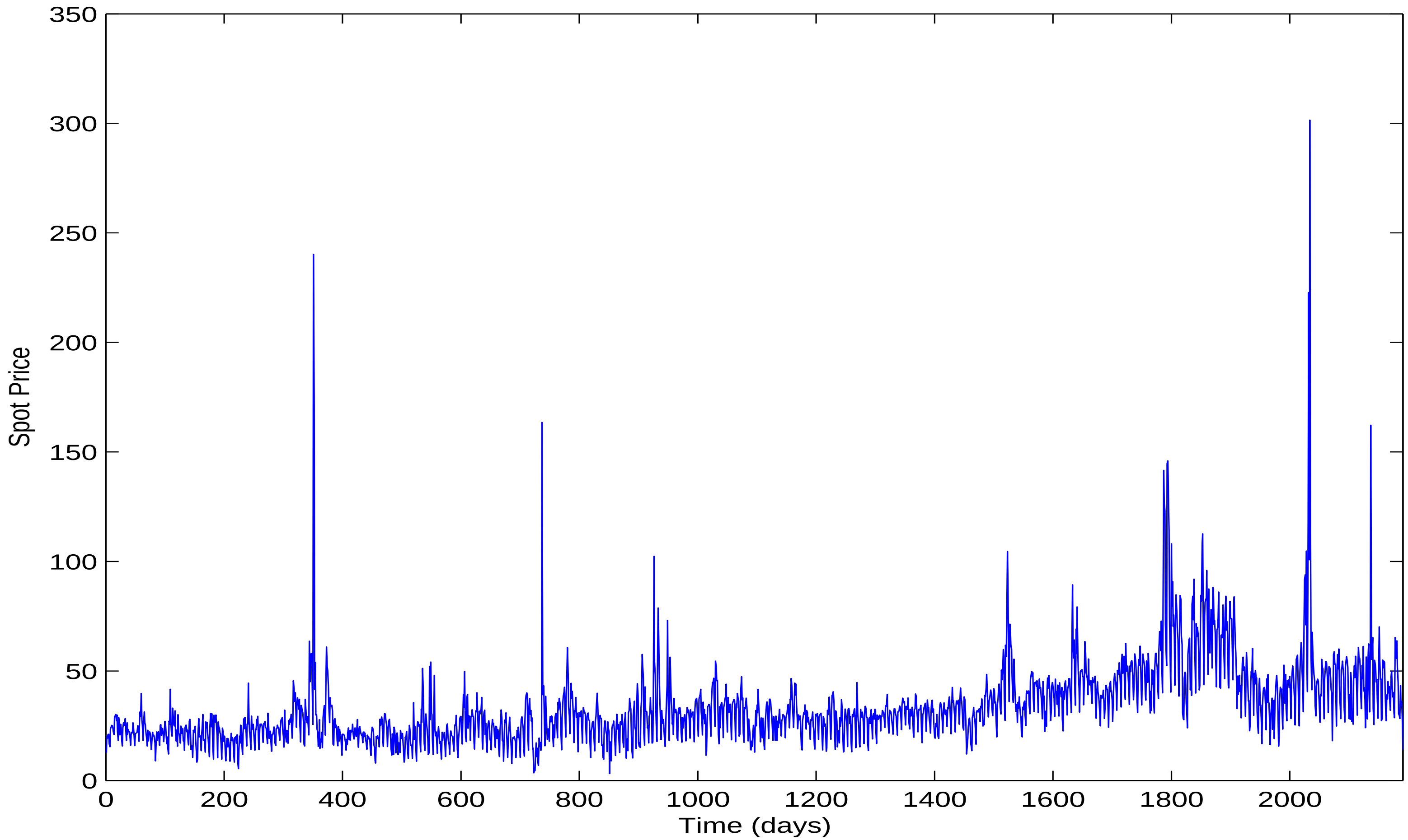


Figure 1: Spot price for Phelix Base (day) index from 01.10.2001-31.12.2006.

Empirical Results

Outlier Detection

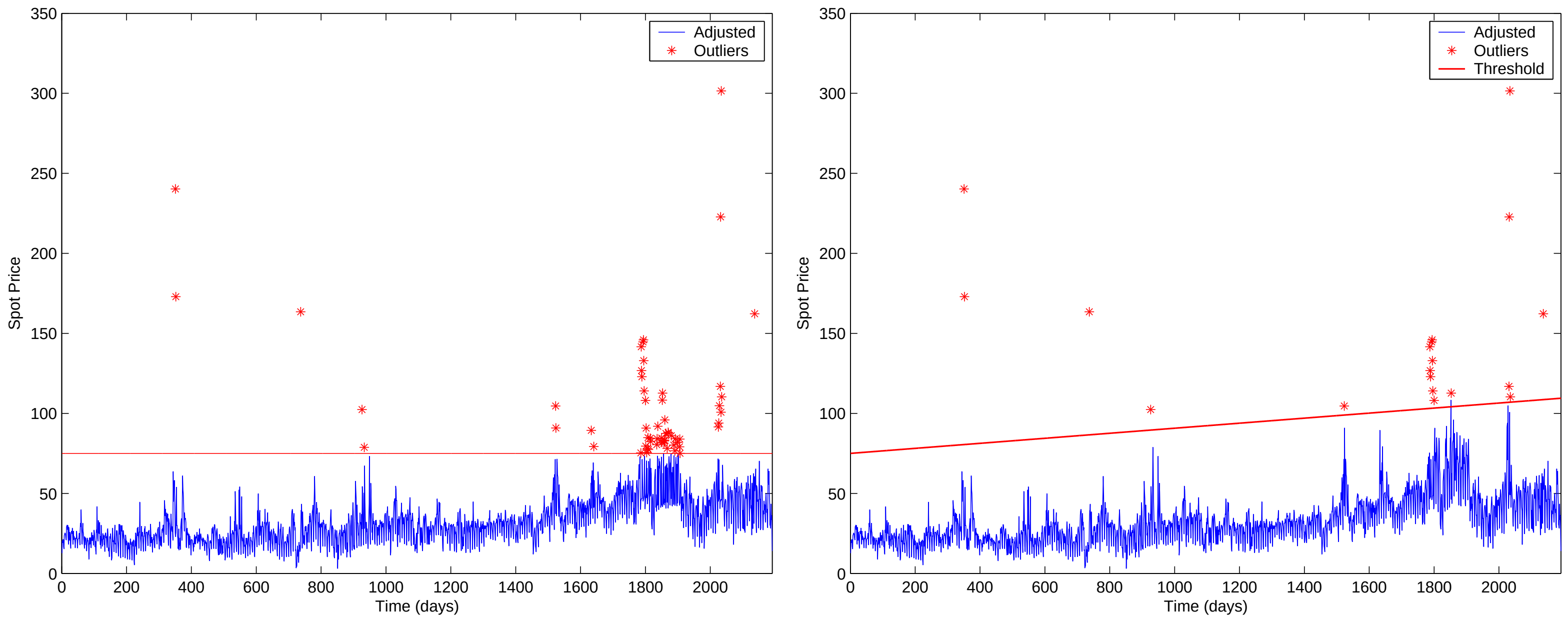


Figure 2: Time series after replacement of the spikes and original observations classified as outliers using fixed price thresholds for the original (*left panel*) and a detrended series (*right panel*).

Empirical Results

Outlier Detection

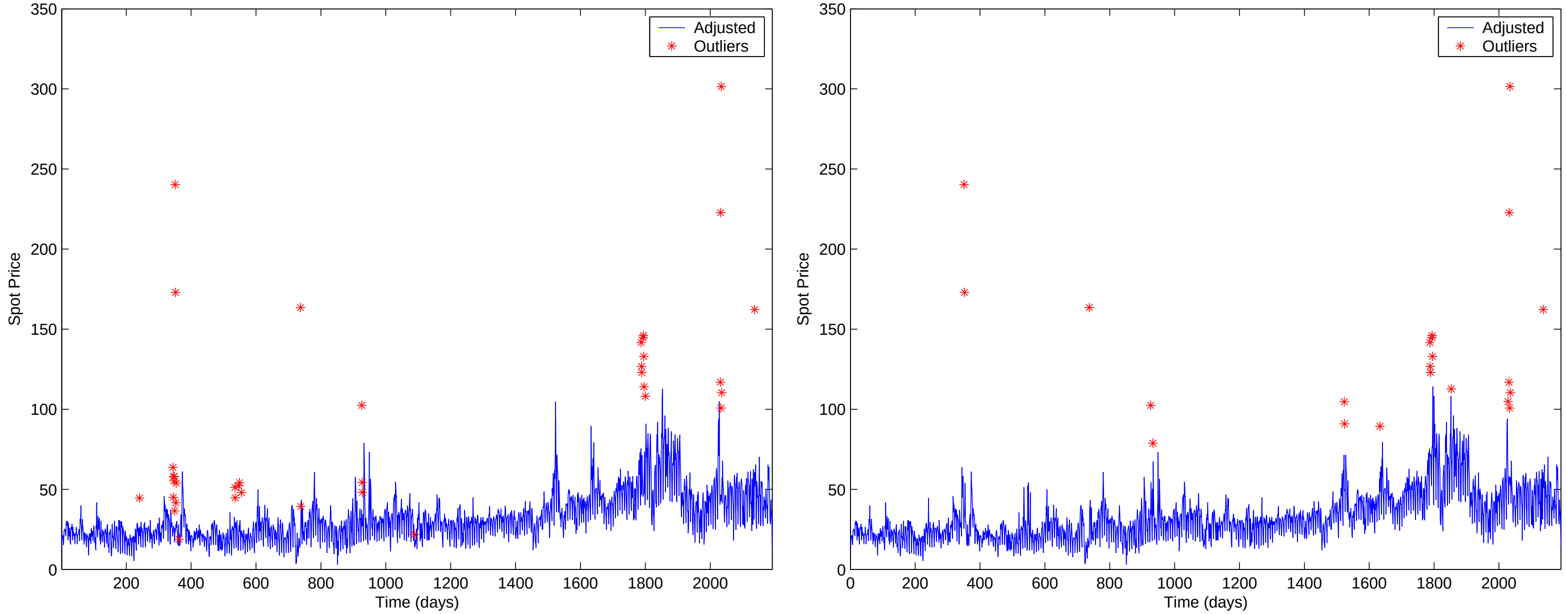


Figure 3: Time series after replacement of the spikes and original observations classified as outliers using the recursive filter technique (*left panel*) and percentage price thresholds for the detrended series (*right panel*).

Empirical Results

Outlier Detection

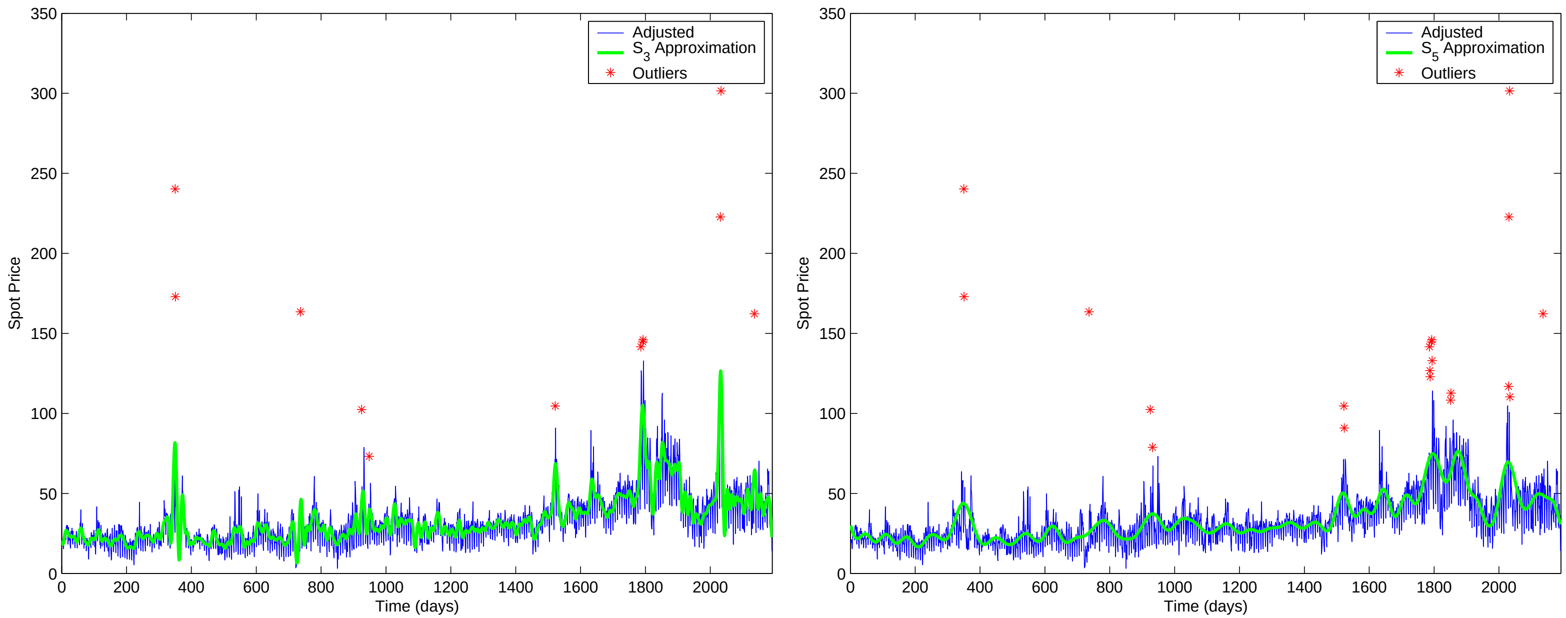


Figure 4: Time series after replacement of the spikes, wavelet approximations using S_3 (*left panel*) and S_5 (*right panel*) and original observations classified as outliers using the wavelet filter technique.

Empirical Results

Spikes and Preprocessed Series

Method	Preprocessing			#spikes	Max	Mean	Std	Skew	Kurt
	Trend	Year	Week						
Original	—	—	—	0	301.54	33.56	19.01	3.93	37.16
Fixed Threshold	—	—	—	50	79.49	32.01	13.43	0.83	3.52
Fixed Thres. Detrend.	✓	—	✓	19	108.25	32.64	14.90	1.24	5.08
Recursive Filter	—	—	✓	37	112.65	32.47	14.92	1.32	5.51
Percentage Thres.	✓	—	✓	22	114.06	32.70	14.99	1.24	5.13
Wavelet Approx. S_3	✓	✓	$\sim$	12	132.91	33.10	16.23	1.70	7.81
Wavelet Approx. S_5	✓	✓	—	20	114.06	32.78	15.14	1.27	5.18

Table 1: Number of detected spikes and descriptive statistics of the series after removing the spikes. Preprocessing indicates whether the trend, the annual and/or the weekly seasonal components have been removed from the original data before identifying the spikes.

Empirical Results

Estimating a Seasonal Pattern for the Preprocessed Series

Split system prices P_t into a deterministic part $f(t)$ that comprises all kinds of seasonal behavior and a purely stochastic component S_t:

$$P_t = f(t) + S_t.$$

- we are not interested in specifying a model for the stochastic component S_t, but mainly in the estimated seasonal pattern for the differently preprocessed data.
- to keep results comparable, the same seasonal pattern (linear trend, dummy variables for days and months) is estimated for all preprocessed series:

$$f(t) = \alpha + \beta \cdot t + d \cdot D_{day} + m \cdot D_{mon} \tag{4}$$

Empirical Results

Parameter Estimates for the Seasonal Pattern

Parameter	Original	Fixed	Fixed Detr	Filter	Percent	Wave S_3	Wave S_5
Constant	**23.573***	**22.249**	**22.683***	**22.546***	**22.998***	**22.908***	**22.660***
Trend	**0.0158***	**0.0137***	**0.0150***	**0.0154***	**0.0151***	**0.0156***	**0.0152***
Tue	**2.388***	0.223	1.192	1.211	0.741	1.212	0.880
Wed	0.759	**1.344***	**1.725***	1.217	1.381	1.414	1.436
Thu	1.156	0.949	1.2097	1.260	1.120	0.994	0.937
Fri	**-2.372***	**-1.371***	-1.271	-0.919	**-1.854***	-1.705	**-1.750***
Sat	**-10.277***	**-8.199***	**-8.765***	**-8.580***	**-9.135***	**-9.614***	**-9.226***
Sun	**-16.818***	**-14.750***	**-15.279***	**-15.016***	**-15.687***	**-16.146***	**-15.767***

Table 2: Parameter estimates for constant, trend and weekly seasonal pattern depending on the outlier detection technique (* indicates significant parameter estimates at the 5% level).

Empirical Results

Parameter Estimates for the Seasonal Pattern

Parameter	Original	Fixed	Fixed Detr	Filter	Percent	Wave S_3	Wave S_5
Feb	0.210	1.097	1.315	1.059	1.243	0.869	1.491
Mar	-1.129	0.128	-0.351	-0.313	-0.613	-0.666	-0.356
Apr	**-5.551***	**-2.601***	**-4.446***	**-4.764***	**-4.514***	**-4.896***	**-4.267***
May	**-8.862***	**-5.928***	**-7.791***	**-8.128***	**-7.846***	**-8.240***	**-7.609***
Jun	**-5.242***	**-2.661***	**-4.106***	**-4.677***	**-4.399***	**-4.595***	**-3.936***
Jul	-0.091	-1.430	**-2.198***	**-3.676***	**-2.888***	-1.180	-1.839
Aug	**-7.021***	**-3.708***	**-5.755***	**-6.236***	**-5.840***	**-6.505***	**-5.612***
Sep	**-4.247***	-0.898	**-2.972***	**-3.343***	**-3.061***	**-3.549***	**-2.834***
Oct	**-5.206***	**-1.829***	**-3.929***	**-4.313***	**-4.017***	**-4.512***	**-3.792***
Nov	-0.638	-0.797	**-2.901***	**-3.327***	**-2.137***	-1.228	-1.914
Dec	-1.093	-1.144	**-2.528***	**-4.078***	-1.802	-1.787	-1.459

Table 3: Parameter estimates for yearly seasonal pattern depending on the outlier detection technique (* indicates significant parameter estimates at the 5% level).

Empirical Results

Prediction of Deterministic Prices

Determine price forecasts for the deterministic part of system prices based on the estimated seasonal pattern:

- starting with a model including all possible independent variables, use backwards stepwise regression to exclude non-significant variables from the model
- obtain a model including only significant variables for each preprocessed series
- the estimated seasonal pattern is used to give (deterministic) price forecasts for each month in 2007

Empirical Results

Prediction of Deterministic Prices

Method	Original	Fixed	Fixed Detr	Filter	Percent	Wave S_3	Wave S_5
Jan	**54.5209**	49.0125	52.4774	52.9667	52.3237	53.5435	52.2009
Feb	**55.0786**	49.5304	53.0487	53.5786	52.9056	54.1691	52.7876
Mar	**55.5115**	49.9425	53.5171	54.0819	53.3647	54.6760	53.2511
...	...	...	...	...	...	...	...
Jul	**57.3837**	51.4427	52.9390	52.0750	52.1647	56.3713	54.9484
...	...	...	...	...	...	...	...
Nov	**59.5912**	53.4480	54.4707	54.6911	55.0923	58.6951	57.1431
Dec	**57.6905**	51.7984	53.0882	52.2966	55.5464	56.8890	55.4473
MAD		**4.16**	**2.14**	**2.03**	**2.11**	**0.59**	**1.66**

Table 4: Average price forecasts for the months January - December 2007 based on the estimated seasonal pattern for the different outlier detection techniques.

Conclusions

1) there is currently no commonly accepted definition of a price spike in the literature
2) different approaches including fixed or percentage thresholds, recursive filter techniques or wavelet approximations for outlier detection are suggested
3) for our dataset, depending on the chosen technique a different number of observations is classified as spikes - ranging from 12 to 50 observations
4) the remaining (preprocessed) series show substantial differences for standard deviation, skewness and kurtosis
5) estimation of the seasonal pattern yields substantial difference in terms of significance of the variables and also in parameter estimates for each series
6) using the estimated seasonal models for (deterministic) price forecasts shows differences up to 10% in average daily price levels depending on the chosen approach for outlier detection

Conclusions and Outlook

7) is it possible to define the optimal deseasonalization technique? what should optimal mean?

8) does the order of performing calculations, i.e. deseasonalize then remove the spikes or vice versa, matter? and how much?

9) what is the impact of a given estimation procedure for spikes and seasonal pattern on parameters of stochastic models for S_t?

10) do (Markov) regime switching models provide better means of identifying the spikes?