Multiple mean reversion jump diffusion model for Nordic electricity spot prices

Matylda Jabłońska[11], Hasifa Nampala[1], Tuomo Kauranne[1]

[a]*Lappeenranta University of Technology, Department of Mathematics and Physics, Skinnarilankatu 34, 53850 Lappeenranta, Finland, matylda.jablonska[at]lut.fi*
[b]*University of Dar-es-Salaam, Mathematics Department, PO Box 35-62, Dar-es-Salaam, Tanzania, hnampala[at]yahoo.co.uk*
[c]*Lappeenranta University of Technology, Department of Mathematics and Physics, Skinnarilankatu 34, 53850 Lappeenranta, Finland, tuomo.kauranne[at]lut.fi*

Abstract

Electricity spot market prices are notoriously difficult to model, let alone predict, because of their extreme volatility. Such volatility is reflected in so-called price spikes that may increase the spot price by an order of magnitude as a matter of hours. Spot market price series are also subject to many other types of phenomena, such as periodicities at different scales, and to mean reversion. We introduce a model for electricity spot market prices that includes both spikes and mean reversion. The model is based on a jump diffusion process that is superimposed on a mean reverting Ornstein-Uhlenbeck model. Mean reversion takes place at several different time and price scales, so as to reproduce the observed behavior of spot market prices correctly. The parameters of the model are calibrated with the Nord Pool spot market hourly price series using a maximum likelihood approach. The simulated price series thus obtained very closely follows the statistical characteristics of the real price series.

Key words: electricity spot price, price spike, stochastic simulation, mean reversion jump diffusion model, multiple mean reversion

---

[1]Corresponding author: matylda.jablonska[at]lut.fi

1. Introduction

Nord Pool is the first international power exchange that operates along with a common electricity market for a group of European counties: Norway, Sweden, Finland, Denmark and part of Germany. Nord Pool price history presents a great laboratory of electricity price behavior. The key mission of the not-for-profit company Nord Pool Spot is to resolve the task of electricity trading, *i.e.* to match the generation of electricity with the demand precisely and at an optimal price for each hourly trading period of each day.

Electricity spot prices are very much dependent on external factors like demand level, generation possibilities, weather conditions, *etc.* They also have a lot of variability which seems to be unexplainable by those deterministic indicators. However, as very distinctive time series, electricity spot market price series have some distinguishing statistical features, such as mean reversion and spikes. Those can be studied and partially reproduced by proper models.

One of the main characteristics of Nord Pool spot trading is that always by noon of a given day market participants have to submit their bids for power generation and demand for all of the 24 hours of the next day. Therefore, the price behavior as a process is carrying more of a 24 hour pattern, rather than hour-by-hour relationships. Obviously, a part of this 24 hourly periodicity comes straightforward from the daily electricity demand pattern, while another part is determined by phenomena with longer time frames of impact.

This characteristic feature of electricity price has been confirmed by numerous other studies. The authors of **?** discovered that different price drivers get significantly different values over the day time. A study of New Zealand prices **?** has shown that five different behaviour schemes can be found within a one-day period, where prices are highly correlated within the groups and barely between them. The results showed that given half-hours of each day were more correlated with the respective trading periods of other days rather than with the adjacent hours. In **?** we find an analysis based on panel price representation – a panel with 24 cross-sectional hours.

Much as the stochastic models of finance have found their way to the power market, the most prominent model of *geometric Brownian motion* alone is not able to provide good results in modeling electricity prices since it does not capture spikes and mean reversion – the main distinguishing features. Hence, a modified version of geometric Brownian motion has been proposed, a *jump diffusion model*. In a comparative study **?** have evaluated the effectiveness of various diffusion models in describing the evolution of electricity prices in several markets. They conclude that geometrical mean-reverting jump diffusion models provide the best performance and that all models without jumps appeared inappropriate for modeling electricity prices, since they are not able to generate spikes. A general specification of a jump diffusion model that comprises many characteristics of electricity prices was proposed by **?**. The panel study of **?** suggest using a separate mean reversing model (including distinct mean reversion rates and levels) for each of the 24 hours.

In this study we utilize the mean-reverting jump diffusion model that has its origin in a classical Ornstein-Uhlenbeck process. We equip such a model with three separate reversion rates, with respect to different price regimes. Since the data we use are hourly prices, we build the model based on a 24 hour reversion time interval, to cater for Nord Pool trading characteristics as well as daily demand patterns that heavily influence price behavior. Moreover, we propose a non-constant mean reversion level based on a background variable, to capture the price annual seasonality. A related approach, but without jump diffusion, is presented in **?** where multiple mean reversion levels, corresponding to 'spiky' and 'non-spiky' regimes of behavior, are also used to good effect.

The paper is organized as follows. Section **??** gives more background on recent studies in the field and presents the characteristics of spot prices, their basic statistics as well as the features typical for electricity prices like mean reversion and spikes. Section **??** defines the multiple mean reversion jump diffusion model with the corresponding Matlab simulation algorithm. In Section **??** we collect simulation results and summary statistics. Finally, Section **??** concludes and gives suggestions for future work.

2. Spot price and price spikes statistical features

Within the last decade there have been a considerable number of studies carried out, attempting to apply classical and more modern time series approaches for modelling electricity spot prices. Even though many of them seem to give good fit to the data, they persistently fail in forecasting. For instance, even optimally chosen ARMA-GARCH models with improved information criteria are not able to produce reliable predictive distribution of forecasts, **?**. These models fitted with generalized Pareto distribution give only accurate estimates and forecasts for extreme quantiles, as can be found from **?**. Hence the authors suggested application of extreme value theory for risk management. ARIMA models, with trend and seasonality eliminated by differencing, prove to perform better when improved by predicted error correction **?**. In **?**, comparison of 12 different time series models is presented, showing that even the best fitting SNAR/SNARX models still meet a lot of limitations and can not be fully trusted in predictions. Also, some vector autoregressive models with exogenous effects and skewed t-distributed disturbance were proposed, for example the first order model in **?**. Its main advantage was to capture the intra-day time dependency of electricity price data.

Besides, some modern modelling approaches were proposed, with example of self-organizing maps and support vector regression networks, see **?**. They work very well for reconstructing seasonality in prices based on load as an explanatory variable. However, no out-of-sample forecasts were introduced.

The intra-day price behaviour was found tempting for many researchers. Clearly, electricity prices depend on electricity consumption, which itself is highly daytime-dependent. For instance, **?** performed a broad analysis of hourly price drivers, including not only the most commonly used temperatures, but also strategic effects, power plant dynamics, agent's learning, risk perceptions, market design implications and trading inefficiencies. They also proposed regime switching (in particular, specified as low and high price levels) to reveal discontinuities in price realizations.

In this work we suggest that when one aims at reconstructing electricity spot price features, one has to analyze the time series in two layers – one of regular behavior and another one of irregular behavior. The former oscillates around the so-called mean level, and thus drives the general mean reversion process. The latter is represented by price spikes. However, the classical regime switching suggests that the switches are occurring randomly, whereas we prove that in case of Nord Pool prices the high price state activation has connection with given price level. Also, we distinguish in total three different regimes – regular, spike and after-spike. These regular and non-regular price components require a separate statistical analysis but in the end they need to be combined together with one another in order to run the simulation properly.

2.1. General statistics on price evolution and spikes

Electricity spot prices are known to be very different from other financial time series. This can be noticed from a very first look at the price plot (see Figure **??**a) which shows the spiky character of the prices. Neither do price logarithmic returns follow the normal distribution required by most financial schemes and models. The distribution of both prices and returns are characterized by fat tails (see Figure **??**b), representing the sometimes extreme behavior of the spot price. Not only do they have the highest volatility of all financial markets, but also variability of the standard deviation itself is high. Electricity price can change up to twenty times higher within one hour only.

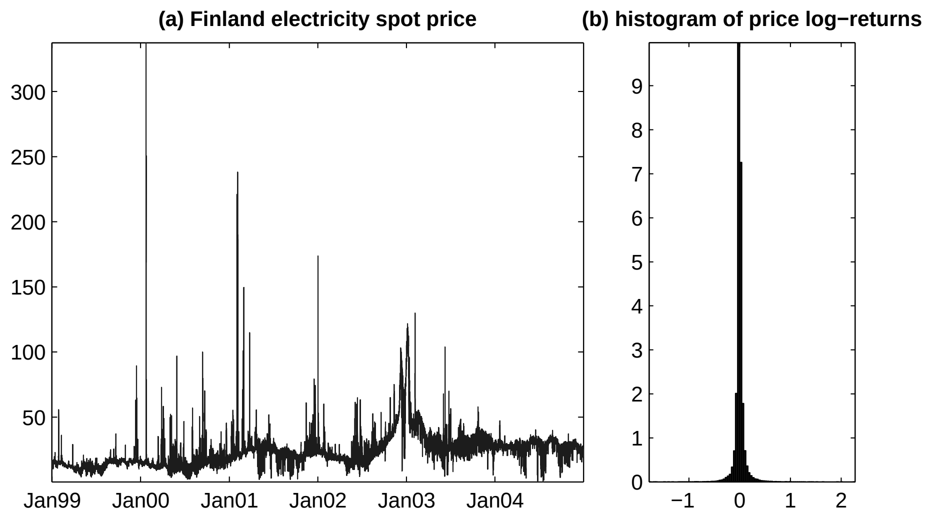

Figure 1: (a) Finland area hourly electricity spot prices. (b) Histogram of Finland price log-returns.

As mentioned, what differentiates electricity spot prices from other types of financial time series is their high volatility over time. Therefore, a study of spikes along with general price statistics is very important. Before that it is necessary to formulate a spike definition that will be used within this paper. It has already been subject to some discussions of setting thresholds for specifying outlying price values. By definition, a spike is defined as a price that surpasses a specified threshold, **?**. However, there arise two main questions: how high the threshold should be and whether that would be one global value or rather some time-dependent parameter.

We can clearly see that the electricity spot prices neither follow any constant mean level nor have any constant trend. The general price path changes up and down over months. This gives first motivation for using a varying threshold value rather than just 'cut off' any global outliers. Specifying the boundary level itself is even more subjective. Some references suggest the use of fixed log-price change thresholds, *e.g.* log-price increments or returns exceeding 30% proposed by **?**, or varying log-price change thresholds, for instance, log-price increments or returns exceeding three standard deviations of total price changes as introduced by **?** and **?**. We have decided to use the methodology employed by **?**, *i.e.* base the threshold value on the original price values. This causes us to define a spike as the difference between a price exceeding a moving average μ_t in a specified window w by more than s times a moving standard deviation σ_t and the mean level. Considering the high frequency of data (hourly), we set $w = 96$ meaning 4 days and $s = 3$.

Having the spikes defined, we can extract them from our price series and verify data statistics. Table **??** collects basic distribution parameters for prices and spikes. Considering the spike definition we adopted implies that an outlier has an average value twice higher than neighboring prices (within the window). The count of spikes shows that they constitute less that 0.5% of all the hourly prices. However, their magnitude and unexpectedness cause them to have huge significance in electricity markets.

Table 1: Basic statistics for Finland area spot price and price spikes.

	number	mean	std	skewness	kurtosis
prices	52608	23.5990	13.0982	3.6055	35.2557
spikes	240	29.1706	39.9169	4.0015	22.4366

2.2. Spike sizes and densities

One needs to be aware of the fact that one of the most challenging questions in modelling electricity spot prices is how to model spikes. As a very distinguishing type of behavior, they should be studied in detail. In our work we start from studying spike sizes. Based on the definition proposed in Section **??** we get spike values ranging from 3.19 up to 296.16. Let us recall that a spike is found as a difference between the outlying price and a local mean value. As a result we get spikes a distribution as presented in

Figure **??**. Moreover, we compare the empirical normalized histogram with an exponential distribution having a parameter λ equal to the mean value of extracted spikes.

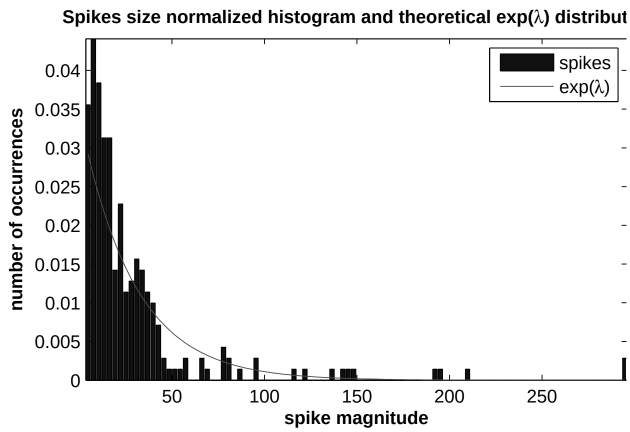

Figure 2: A normalized histogram of spike size and the corresponding theoretical $\exp(\lambda)$ distribution.

Even though there is some similarity in the distribution shapes, it seems advisable to use an empirical distribution for sampling spike sizes, rather than the exponential distribution. Therefore, we construct an empirical cumulative distribution function as shown by Eq. (**??**)

$$\hat{F}_n = \frac{\text{number of elements in the sample}}{n} = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x) \tag{1}$$

where $I(A)$ is the indicator of event A. It will be then used for sampling spike magnitudes with use of inverse CDF method.

Apart from spike sizes it is important to investigate what increases the chance of price to spike at a particular time point. One could suppose that to detect the probability of a price spike it would be enough to use the statistics of spike occurrence and intuitively divide the number of spikes by the total number of observations. Also, another option would be to study the histogram of how many spikes emanate from certain price intervals. However, it can be seen from the price plots that spikes are not uniformly distributed along the time axis but usually they seem to cluster and are more likely to occur from high price levels. To verify this hypothesis, we study spike densities with respect to the price level from which they occur. This shall give a relative view, *i.e.* the likelihood of having a jump from a certain price in respect to the probability of that price level itself. Even if there seem to be more spikes emanating from low prices, in absolute numbers lower prices are more common, and the relative probability of having a spike is still smaller than in the case of high prices.

We have sliced the price range from 0 to 120 Eur/MWh by every 10 Eur/MWh and divided the number of spikes that occur within the specified cluster by the total number of hours having a price within the same range. As the result we get Figure **??** showing computed spike densities. We can see that the price levels are most likely to be based on spikes that are ranging from 50 to 70 Eur/MWh. Also, spike probability is zero below 10 Eur/MWh and drops back to zero when crossing 110 Eur/MWh.

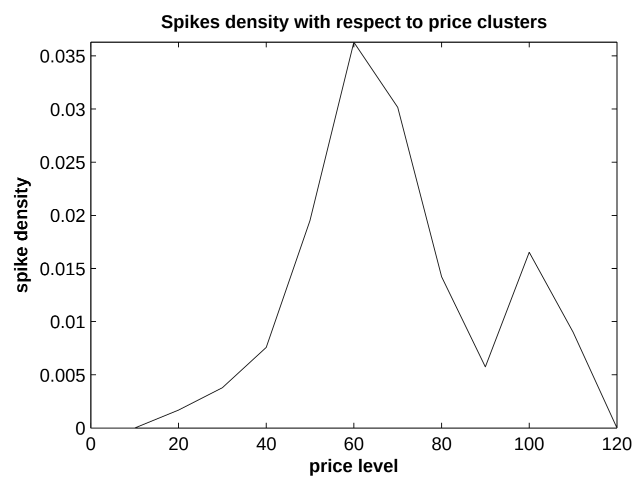

Figure 3: Spikes density within different price levels.

A classical approach for jumping processes uses the Poisson distribution (see **?**). In our work we will consider spikes as events coming from a Poisson distribution $P(\lambda_k|X_t \in I_k)$, where X_t is the current price level, I_k is a specified 'slice' of price range and λ_k is the spike density on the interval I_k.

We should also ensure that the jump process is Markov. We do use historical data to estimate the respective λ_k parameters, but while simulating each jump is not referring to the previous ones. Indeed, it is conditional on the current price level but memoryless with respect to historical jumps.

2.3. Multiple mean reversion rates

There have already been studies employing mean reversion jump diffusion models for electricity spot prices. The most common one is the Ornstein-Uhlenbeck process with a jump component in the form of Eq. (**??**).

$$dX(t) = \alpha(X^* - X(t))dt + \sigma(t)dW(t) + J(t)dN(t) \tag{2}$$

However, it has a weakness that arises from using only one constant reversion rate. In particular, if the reversion rate is low and the price jumps, it comes back to the mean level too slowly, losing the 'sharp' character of spikes. On the other hand, if the reversion rate is strong, it heavily influences the non-spiky price behavior. Therefore, we propose an idea of using different reversion rates for the regular and extreme price paths.

Electricity spot prices can be considered from two different points of view – depending on whether we consider them on a daily or on an hourly basis. They display different behavior as stochastic processes. Daily prices in each case have some spike persistence, *i.e.* high prices may occur on a few consecutive days. Note that daily hours are constructed as mean values of all hourly prices from a given day. From hourly price characteristics we know that if an extreme value occurs in peak hours it almost never persists over consecutive hours.

Moreover, as we mentioned before, Nord Pool spot bidding is carried out every day by noon for all the 24 hours of the next day. This causes price peaks in particular hours to become reflected in the same hour of the next day rather than next hour of the same day. To verify this phenomenon we have chosen three spikes in the Finland price series and study their inter- and intra-day behavior as presented in Figure **??**.

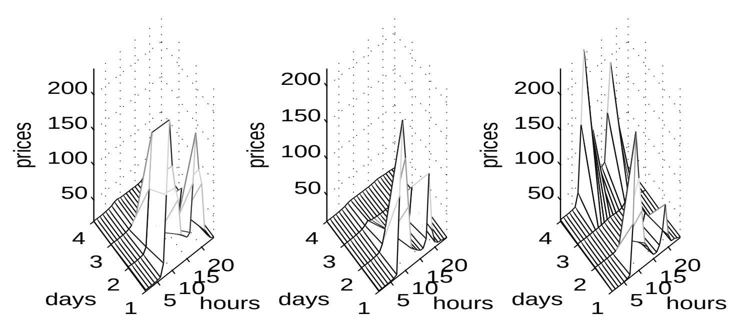

Figure 4: Three chosen Finland spikes in inter- and intra-day behavior.

Indeed, we can see that if a spike occurs in a particular hour it tends to have 'an echo' in the same hour of the following day, and two days after as well. Also, we suggest that prices in hours around the spikes are more related to the values on respective hours of previous days. Therefore, we propose that to capture these characteristics, we should mean-revert spikes separately with $\Delta t = 24$. Note that this separate reversion rate would be a lot stronger than the one for regular price behavior.

Finally, we decide to consider also the price two days after a spike as a specific event as well, offering an individual reversion rate for it as well. It results from the study showing that price spike needs on average 2-3 days to relax.

One of the past studies (see **?**) uses linear regression to estimate the mean-reversion parameters: level, rate and volatility. They considered only the mean-reversion and the diffusion part while excluding the jumps, for reason mentioned before that spikes would probably need a stronger reversion rate. We use the same linear regression approach, also for treating spikes separately. We shall consider the dependent variable to be the price change between times t and $t-24$ or $t-24$ and $t-48$ accordingly, and the independent variable is the price level at time $t-24$ or $t-48$. The regression equation comes as in Eq. (**??**)

$$P_{t+24} - P_t = \rho_1 + \rho_2 P_t + \epsilon_t,\ \forall t = 1, 2, \ldots, n-24 \tag{3}$$

where ρ_1 and ρ_2 are regression parameters and ϵ is the observation error. Employing methodology from **?**, the mean reversion parameters are estimated as presented in Eq. (**??**)

$$mean\ reversion\ rate = -\rho_2. \tag{4}$$

The above equations bring results for reversion parameters of the non-spiky regime. To cater for spike reversion we use the same formulae as **??** and **??** but for the prices to be P_{t+24} regressed we set a constraint $t + 24 \in S$, where S is the set of indices of spikes in our price series. Finally, to estimate the rate for the price two days after spike we run the regression with respect to Eq. (**??**)

$$P_{t+48} - P_{t+24} = \rho_1 + \rho_2 P_{t+24} + \epsilon_t,\ \forall t = 1, 2, \ldots, n-48 \tag{5}$$

keeping the same restriction of $t + 24 \in S$.

Table **??** collects parameters for Finland and System price estimated using the method described above. We can clearly see, that all reversion rates are significantly different, which lends credibility to our approach. The non-spiky regime estimate α is substantially lower than the rate for spike first reversion, and still visibly weaker than the spike second reversion.

Table 2: Mean reversion rate estimates for Finland hourly spot price.

α	β_1	β_2
0.0291	0.8325	0.1095

The proposed three mean reversion rates can be related to treating three different regimes of prices. However, let us note that the regime switches are approached differently than in classical methodologies. As we have found, the probability of spike occurrence (spike density) is dependent on given price level, though does not relate to the previous spikes. Therefore, if we consider that regime 1 is the regular one, regime 2 is the spiky regime, and regime 3 is the after-spike situation, then the matrix of transition probabilities would come as Eq. (**??**).

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} = \begin{bmatrix} 1 - p_{Jt}|_{X_{t-1}=x} & p_{Jt}|_{X_{t-1}=x} & 0 \\ 0 & p_{Jt}|_{X_{t-1}=x} & 1 - p_{Jt}|_{X_{t-1}=x} \\ 1 - p_{Jt}|_{X_{t-1}=x} & p_{Jt}|_{X_{t-1}=x} & 0 \end{bmatrix} \tag{6}$$

where $p_{Jt}|_{X_{t-1}=x}$ stands for probability of having a spike at time t provided that price as moment $t-1$ equals x. The rows of the matrix sum up to one. Let us go through all the cases to understand the transitions:

- if the process is in regime 1
  - it can not switch to regime 3 (after-spike regime)
  - it can switch to regime 2 if there occurs a spike
  - it can stay in regime 1 if no spike occurs
- if the process is in regime 2
  - it can not switch to regime 1, as it has to go through the after-spike stage
  - it can stay in regime 2 if another spike occurs

- it can switch to regime 3 if there is no new spike

- if the process is in regime 3
  - it can not persist in the after-spike regime
  - it can switch to regime 2 if another spike occurs
  - it can relax to regime 1.

2.4. Mean reversion level

We know that electricity prices are strongly dependent on seasons, especially in countries like Finland where power production in winter is always a lot higher than in other seasons. Since we want to run simulation of pure, not deseasonalized prices, we propose including seasonality in the mean reversion level, changing it in this way to a non-constant series.

There are many factors driving local trends of electricity spot prices. The one considered as most influential is demand. The reason lies in the fact that the higher the demand is, the more more expensive production sources the country may need to activate. Especially in case of sudden increase on the demand side, more costly powerplants, like thermal or gas, may need to be started up (see Figure **??**).

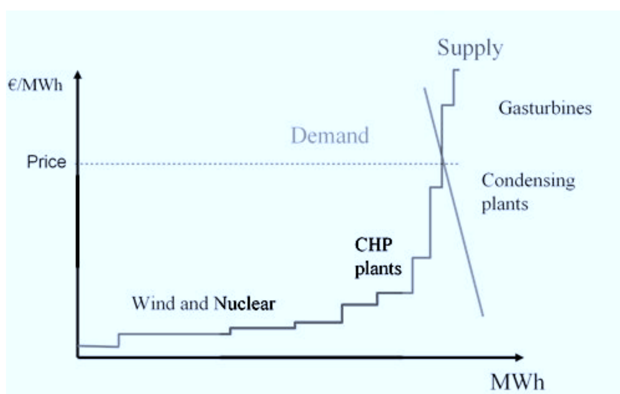

Figure 5: Cost of electricity with respect to source of production. (Source: Risø DTU)

It has been shown that price and load keep high correlation between each other and their volatility, even after removing the weekly and intra-day seasonalities, e.g. with use of Discrete Fourier Transform as presented in **?**. Therefore, we intend to use a variable that provides a lot of information about demand through high correlation, but also helps to avoid overfitting. The main one on the demand side in case of Nord Pool is temperature. The Nordic market, especially Finland, is quite vulnerable when it comes to weather influence. In particular, Finland receives substantial power transmission from Russia, and whenever the winter is cold in both countries, this flow may be limited and cause a necessity to increase local Finnish production which is more expensive. It can be seen from Figure **??**a that there is a slight mirror behavior between prices and temperatures.

Therefore, to give an example of a possible price mean level, we perform the usual linear regression of prices against temperatures. However, to avoid using a single estimate over the 6-year horizon (we know that there is also economic, as well as other factors that influence price trend) we use a moving 60-day window. To project a mean price on the following day, we run a regression of past 60 prices against temperatures from the past 60 days to obtain proper coefficients. The temperature data makes an average of notes from two Finnish cities: Helsinki and Sodankyla (this is based on geographically different location). As a result we get Figure **??**b showing the original price and the one projected from temperatures.

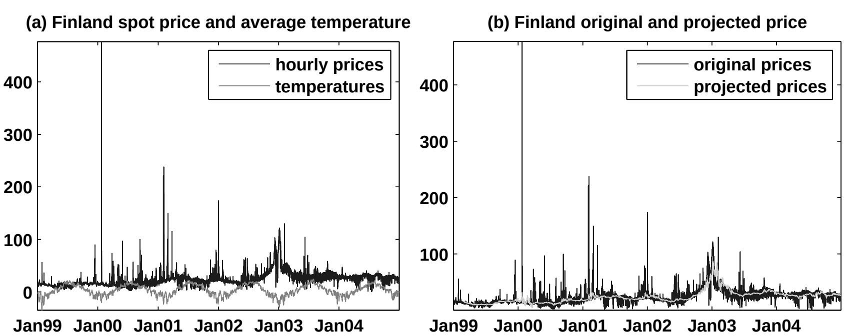

Figure 6: (a) Finland spot price and average daily temperature (average of Helsinki and Sodankyla cities). (b) Finland original and projected price, by temperature regression with horizon of 60 days.

We can see that the projected price follows nicely the original one, without overfitting. Of course, there is still possibility to include more explanatory variables in the regression model, but we do not investigate these any deeper, leaving it as a possibility for future work.

3. Spot market model and simulation algorithm

Having investigated the most prominent price and spike features we can move on to writing down the final model formula that will be used in the simulation.

3.1. The model

Summarizing the whole analysis so far, we intend to build a mean reversion jump diffusion model with three different reversion rates depending on the history of the process, *i.e.* separate mean reversion for non-spiky hours and times 24 hours and 48 hours after a spike, respectively. The mean reversion level is common for all reversion rates, but instead of using a single number it is a varying process related to Finland average daily temperature. The spikes shall be simulated as Poisson-distributed events with probabilities dependent on current price level, and with magnitudes sampled from the empirical distribution of spikes. Finally, the model is complemented with natural stochasticity by adding Wiener components with non-constant standard deviation initiated from the beginning of the interval of real data and later on continued as a moving standard deviation of the past 10 day horizon.

We are considering (see Eq. (**??**)) as an extension of Ornstein-Uhlenbeck and the other commonly used mean reverting jump diffusion models, to include a second drift of the spike regime.

$$dM_t = \gamma(t, M_t)dt + \sigma_t dW_t + J(t, M_t)dN_t \tag{7}$$

where, given a price threshold M^* beyond which the prices are regarded spiky

$$\gamma(X) = \begin{cases} \alpha(X^* - X_t), & \forall_{X \leq M^*} \\ \beta(Y^* - X_t), & \forall_{X > M^*} \end{cases} \tag{8}$$

Hence, if the prices that surpass the threshold are denoted Y, we construct the final simulation equation as presented below with the following components: Eq. (**??**) representing regular non-spiky regime mean reversion and Wiener increments, Eq. (**??**) and Eq. (**??**) standing for mean reversion provided that there was a spike 24 or 48 hours before respectively and, finally, Eq. (**??**) representing the jump components necessary to create proper price changes.

$$\Delta X(t_k) = \alpha(X^*_{t_k} - X(t_{k-24}))\Delta t \,|_{(v(t_{k-24})=0)} + \sigma(t_k)\Delta W(t_k) \tag{9}$$
$$+ \beta_1(X^*_{t_k} - Y(t_{k-24}))\Delta t \,|_{(v(t_{k-24})=1)} \tag{10}$$
$$+ \beta_2(X^*_{t_k} - Y(t_{k-24}))\Delta t \,|_{(v(t_{k-48})=1)} \tag{11}$$
$$+ J(t_{k-48})v(t_{k-48}) + J(t_{k-24})v(t_{k-24}) + J(t_k)v(t_k) \tag{12}$$

where

- X_{t_k} is the spot price in time t_k
- $\Delta X(t_k)$ is the simulated price change
- α is the mean reversion rate for non-spiky regime
- $X^*_{t_k}$ is the time-dependent mean reversion level
- v_{t_k} is the binary indicator of a spike event at time t_k
- $\sigma(t_k)$ is the standard deviation for Wiener increments
- $\Delta W(t_k)$ is the Wiener process increment
- β_1 is the mean reversion rate between a spike and the price 24 hours later
- β_2 is the mean reversion rate between the price 24 hours after the spike and 48 hours after the spike
- $J(t_k)$ is the jump size in time t_k

3.2. Simulation algorithm

Having estimated the parameters in the previous sections we move on to a Matlab simulation of the stochastic jump diffusion model with three mean reversion rates and find out how it replicates the historical Nord Pool data. We consider two sets of price series, the spiky set, where we include the price series only at times when the spikes occur and the non-spiky series where the spike prices are replaced by the mean value of the analysis window.

Simulation algorithm

1. Split the price range into intervals and compute spike occurrence probabilities on each subinterval I_q
2. Compute mean reversion rates and levels
3. Initialize the first 240 observations from the original data $psim_{1:240} = p_{1:240}$
4. For the current time point k, verify the current price level
5. Generate the value of the event variable v from a Poisson distribution $P(\lambda_q)$, with a parameter dependent on current price level I_q; if Poisson number equals zero set $v_k = 0$, else set $v_k = 1$
6. Generate a jump size based on an empirical jump distribution
7. Then
   (a) If there was a spike 48 hours ago (*i.e.* $v_{k-48} = 1$), compute the current price change $dpsim_k$: restore the $psim_{k-24}$ value and revert it with reversion rate β_2 to the respective mean reversion level; add a Wiener increment and a new jump value if $v_k = 1$
   (b) Else, if there was a spike 24 hours ago (*i.e.* $v_{k-24} = 1$), compute the current price change $dpsim_k$: restore the $psim_{k-24}$ value and revert it with reversion rate β_1 to the respective mean reversion; add a Wiener increment and a new jump value if $v_k = 1$
   (c) Else compute the current price change $dpsim_k$: revert the current price value with reversion rate α to the respective mean reversion level; add a Wiener increment and a new jump value if $v_k = 1$
8. Then
   (a) If there was a spike 25 hours ago (*i.e.* $v_{k-25} = 1$), set the current price $psim_k$ as the previous simulated price $psim_{k-1}$ decreased by the previous price change $dpsim_{k-1}$ (carrying information about a mean-reverted spike) and increased by the current price change $dpsim_k$
   (b) Else, if there was a spike 1 hour ago (*i.e.* $v_{k-1} = 1$), set the current price $psim_k$ as the previous simulated price $psim_{k-1}$ decreased by the previous price change $dpsim_{k-1}$ (carrying information about a spike) and increased by the current price change $dpsim_k$
   (c) Else set the current price $psim_k = psim_{k-1} + dpsim_k$
9. If current price drops below zero ($psim_k < 0$), set $psim_k = |psim_k|$
10. Return to step 4 until the current time reaches the simulation horizon length.

4. Simulation results

In this section we present simulation results in two approaches: a long-run simulation to verify how general price process statistics can be reconstructed, and a short-run out of sample simulation to investigate how the micro-structure of the simulated path follows the original data.

4.1. Long-run simulation

The simulation described in Section **??** is now run to make an attempt at reconstructing the spot price dynamics of Finland. We start the investigation of results from Figure **??**a by showing the simulated price path. For a better overview of the outcome we also present, in Figure **??**b, the original price against the simulated one, together with the underlying mean reversion level estimated from the regression model based on the average temperature in Finland. We can immediately see that the dynamics of simulated prices look similar when it comes to spikes. They are not uniformly distributed along the time axis but cluster especially within higher price regions.

The biggest misalignment in price behavior can be seen in winter 2002-2003 when the general price level is high and, therefore, our simulation considers that time as more likely to have spikes. The second visible difference comes in year 2004 when the real prices are stable, hardly spiking at all, whereas the current simulation finds that price level as one with a significant spike probability.

Besides the general visual investigation, we are interested in a more detailed statistical comparison of the generated prices and spikes with respect to the true ones. Table **??** provides such an assessment. We can easily see that the mean value as well as standard deviation of prices are very close. Also skewness differs not much from the true one, and kurtosis is as significantly high as the original one. When it comes to spikes, the distribution parameters are all very close.

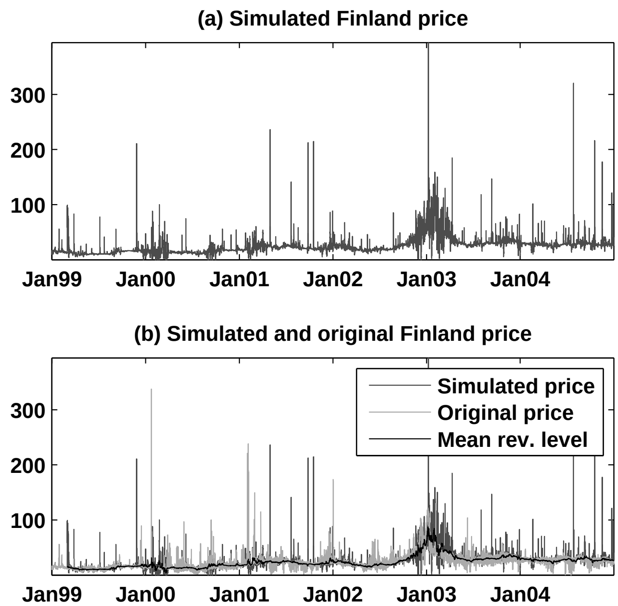

Figure 7: (a) Simulated Finland price. (b) Simulated and original Finland price with mean reversion level estimated from 60-day temperature regression.

Table 3: Basic statistics for Finland area spot price and price spikes.

	number	mean	std	skewness	kurtosis
orig prices	52608	23.5990	13.0982	3.6055	35.2557
sim prices	52608	23.8024	13.5725	4.3122	49.2521
orig spikes	240	29.1706	39.9169	4.0015	22.4366
sim spikes	222	29.3215	39.7939	4.1183	23.4566

Moreover, we compare distribution histograms of both original and simulated prices and spikes, as presented in Figures **??** and **??**. The plots are very much alike.

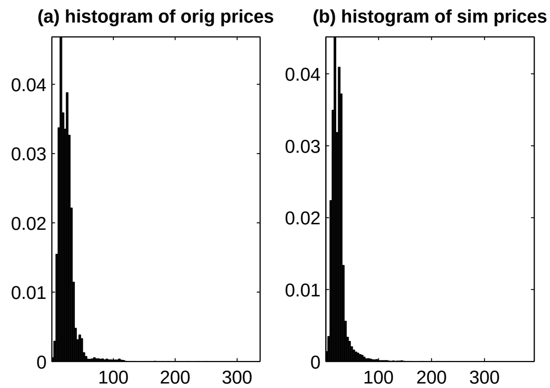

Figure 8: Histogram of Finland prices: (a) original and (b) simulated.

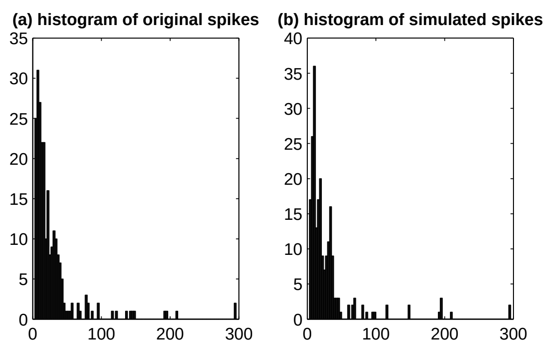

Figure 9: Histogram of Finland price spikes: (a) original and (b) simulated.

Finally, we verify the distribution of price logarithmic returns. As mentioned before, we find electricity spot prices having fat-tailed log-returns. One of the aims of this work is to simulate a process which resembles this feature. Figure **??** provides the answer. We can see that the returns obtained from the proposed simulation are even more fat-tailed than the true ones. There we also analyze the results in a loglog scale (see Figure **??**c) with comparison to theoretical shape of that plot for normally distributed numbers. Both original and simulated returns do show similar fat-tailed behaviour.

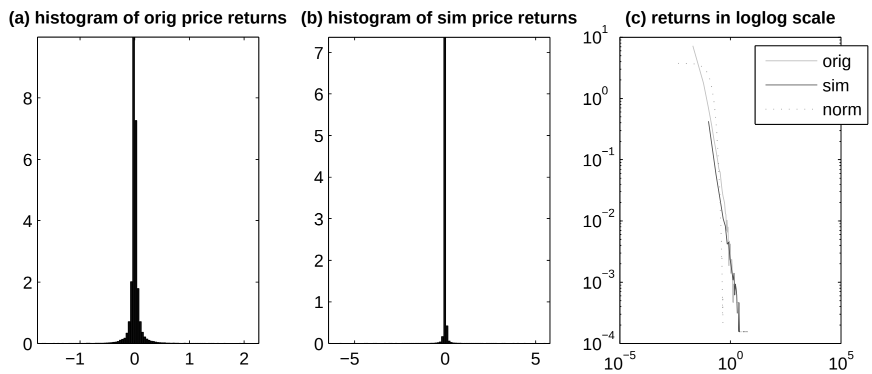

Figure 10: Histogram of Finland price log-returns: (a) original and (b) simulated, and (c) in the log-log scale.

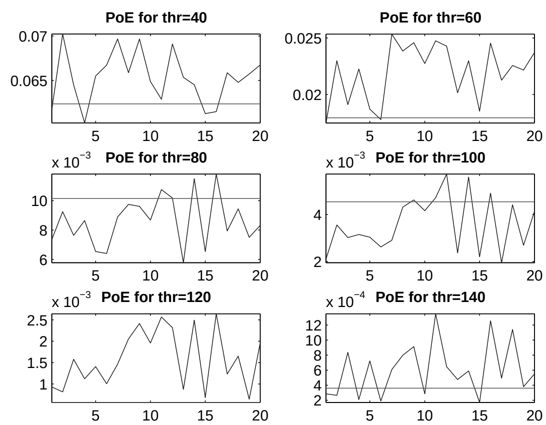

Figure 11: Probability of exceedance.

4.2. Out-of-sample simulation

In this section we present final verification of the model performance by an out-of-sample simulation. The parameters of the model described in previous sections are estimated once again, not for a smaller data set – learning set of 2000 days. Then, based on the estimates we run the simulation and compare the outcome with the original prices as presented in Figure **??**.

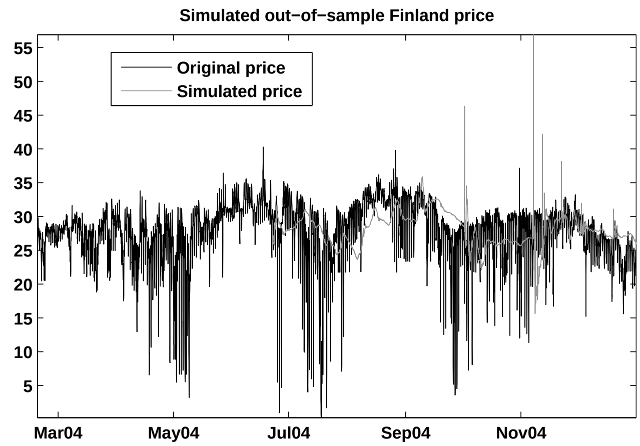

Figure 12: Out of sample multiple mean reversion simulation.

We can see that the underlying mean level follows the original trend. Though we do notice some weaknesses of the simulation. Firstly, our model does not include any component driving behaviour of prices in particular hours. That is, we do consider the dependence of prices within specific trading periods of consecutive days in case of spiky regime, but do not employ any methodology that would make the prices always higher in peak hours and lower otherwise. Therefore, the original data shows a lot more variance due to not only that type of seasonality, but the weekly periodicity as well.

Also, our simulation generated some upwards outstanding observations, whereas such were lacking in the original price series. The cause of it is that the past five years used for model calibration were more spiky at this particular price level, while our out-of-sample simulation falls into economically and hydro-storage stable period in the Nordic electricity market.

5. Conclusion and suggestions for future work

In this work, we have studied the electricity hourly spot prices and made an attempt to provide a simulation that reconstructs real spot price behavior, mostly in terms of spike characteristics. First, we investigated basic statistical features of the available data. We studied in detail the distribution of spikes as well as their inter- and intra-day structure, motivating in this way the use of three different reversion rates in the model proposed later on.

For studying spike distribution over time we have investigated spike density with respect to given price levels, claiming that higher prices are more likely to experience jumps than lower ones. Therefore, we have sliced the total price interval and computed frequencies of spikes within the specified price intervals. This gave a basis for spikes generated as Poisson-distributed events. Spike sizes were suggested to be sampled from the true data distribution.

Next, we estimated the respective mean reversion rates for price behavior within non-spiky regime and, also, 24 and 48 hours after a spike. As to the mean reversion level, we have proposed a common but not constant series projected from a 60-day moving regression of average daily temperatures in Finland. It proved to follow well the true price values, without overfitting.

Moreover, we proposed the model as a differential equation with two superimposed processes (regular mean prices and spikes) together with a Matlab pseudo-code for final price simulation. The simulation was done with step size 1 in a fashion of looking back all the time and verifying spike occurrence in the past, so that they could be properly mean-reverted with a time interval of 24 hours.

Finally, we presented results which by general graphical as well as more detailed statistical comparison proved to resemble well the true data behavior. Both price and spike parameters did not differ significantly from these from counterparts. The small differences observed could be possibly still decreased by putting more emphasis on within-day price behavior, *i.e.* enriching the simulation with the probability structure of particular times of day that may be more spiky than others. Also the regular price path has a strong 24-hour periodic structure, which was not fully captured here.

The study provided a look on the prices as a process switching between three regimes. But distinctively from the recent approaches, we account for the fact that probability of switch cannot be a purely random process, as prices are so much dependent on themselves and external variables. Thus we propose regime switched independent of each other, but related to the given price level. Our contribution also stands out by the fact that the model still remains as one continuous process, opposite to approaches like **?** or **?** where prices were split in 24 hourly vectors or cross-sections. The model we proposed allows an advanced approach for simulation accounting for the hour-to-hour as well as 24-hourly dependencies.

Electricity markets are very complex and one may suppose that along with the deterministic factors, there comes a lot of psychological background as well. The traders' behavior could be compared to animal population spatial dynamics. They do observe one another and thus create the general price path, which could be also understood as the global (in *macroscale*) population formation. However, there is a limit for overcrowding (in *microscale*) which in power trading could be interpreted as physical impossibility of two market participants to buy the same dose of electricity. Therefore, our suggestion for future work is to employ models proposed by **?** in mathematical biology. There, each individual price path simulated from the model proposed in this paper would represent a single trader, and the multiple simulation would provide coupling between the participants (in *mesoscale*) via suitable interaction potentials.