

Accepted Manuscript

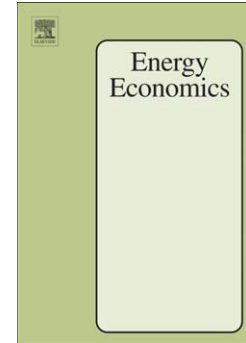
An hour-ahead prediction model for heavy-tailed spot prices

Jae Ho Kim, Warren B. Powell

PII: S0140-9883(11)00129-0
DOI: doi: [10.1016/j.eneco.2011.06.007](https://doi.org/10.1016/j.eneco.2011.06.007)
Reference: ENEECO 2126

To appear in: *Energy Economics*

Received date: 5 November 2010
Revised date: 20 April 2011
Accepted date: 25 June 2011



Please cite this article as: Kim, Jae Ho, Powell, Warren B., An hour-ahead prediction model for heavy-tailed spot prices, *Energy Economics* (2011), doi: [10.1016/j.eneco.2011.06.007](https://doi.org/10.1016/j.eneco.2011.06.007)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

An Hour-Ahead Prediction Model for Heavy-Tailed Spot Prices

Jae Ho Kim and Warren B. Powell

Princeton University

Abstract

We propose an hour-ahead prediction model for electricity prices that captures the heavy-tailed behavior that we observe in the hourly spot market in the Ercot (Texas) and the PJM West hub grids. We present a model according to which we separate the price process into a thin-tailed trailing-median process and a heavy-tailed residual process whose probability distribution can be approximated by a Cauchy distribution. We show empirical evidence that supports our model.

Key words: Heavy-tail, Median-reversion, Mean-reversion

Introduction

For the past decade, many countries have deregulated their electricity markets. Electricity prices in deregulated markets are known to be very volatile and participants in those markets face large risks, accordingly. Daily volatilities of 20-30% are common in electricity markets. For a comparison, stock markets typically have *yearly* volatilities of 20-30% ([20]). There has been a growing interest among electricity market participants about the possibility of profitting through short-term trading by taking advantage of the large volatilities. Unlike stock prices, it is well-known that electricity prices in deregulated markets exhibit reversion to the long-term “average,” allowing us to make directional bets ([6], [15], [25], [31], [35]). Our profit will be determined by how good we are at making such bets - buying electricity when we believe the price will go up and selling electricity when we believe the price will go down.

The strategy for trading electricity must be determined by the behavior of the electricity price process. However, to the best of the authors' knowledge, there are no readily usable price models that describe the characteristics of

electricity price fluctuations in highly volatile markets. In this paper, we analyze four years of hourly electricity price data from Texas, ranging from the beginning of the year 2006 to the end of 2009, as well as three years of data from the PJM West hub for years 2007-2009 to determine an appropriate price model for the data. The method developed in this paper is applicable to any deregulated markets with non-stationary electricity prices with high volatility.

Deregulated electricity markets are not only non-stationary, but also *heavy-tailed* ([1], [3], [5], [17]). A heavy-tailed distribution is defined by its structure of the decline in probabilities for large deviations. In a heavy-tailed environment, the usual statistical tools at our disposal can be tricked into producing erroneous results from observations of data in a finite sample and jumping to wrong conclusions. For heavy-tailed distributions, second and higher-order moments are not finite, i.e., they cannot be computed. They can certainly be measured in finite samples, thus giving the illusion of finiteness, but they typically show a great degree of instability [38].

The central idea behind the price model presented in this paper is as follows. We know that the variation in electricity prices is non-stationary, highly volatile and heavy-tailed. However, we can filter out the heavy-tailed fluctuations via zeroth-order smoothing, i.e., by computing the trailing median of the process, since the quantile function of a set is not affected by rare, extreme events. The trailing median process (TMP), which is a thin-tailed, symmetric, and a slow varying process with very low volatility, reverts towards the mean in the long term. The long-term mean of the TMP, μ , is the long-term “average” of the price process itself. Throughout this paper, we simply refer to it as the long-term mean. On the other hand, the residual process that determines the price fluctuation in the short term is heavy-tailed and difficult to predict, i.e., it is difficult to compute its expected value. The analysis presented in this paper implies that 1) there exists a long-term median for the price process 2) the price process is *median-reverting* towards the TMP in the short term, while 3) we must prepare for an occasional surprise caused by heavy-tailed price spikes.

This paper is organized as follows. In §1, we briefly review the jump-diffusion models for electricity price process and present some motivating observations to show why it is not suitable for real-time trading in the hourly electricity market. We also show that the law of large numbers is not applicable in a heavy-tailed environment and hence empirical averages have little statistical significance. In §2, we define the TMP that forms the thin-tailed portion of the electricity price process and conduct empirical analysis using the spot market data from Texas and PJM. In §3, we define the residual process that forms the heavy-tailed portion of the electricity price process and conduct empirical analyses that lead to one-step prediction model in §4. In §5, we summarize

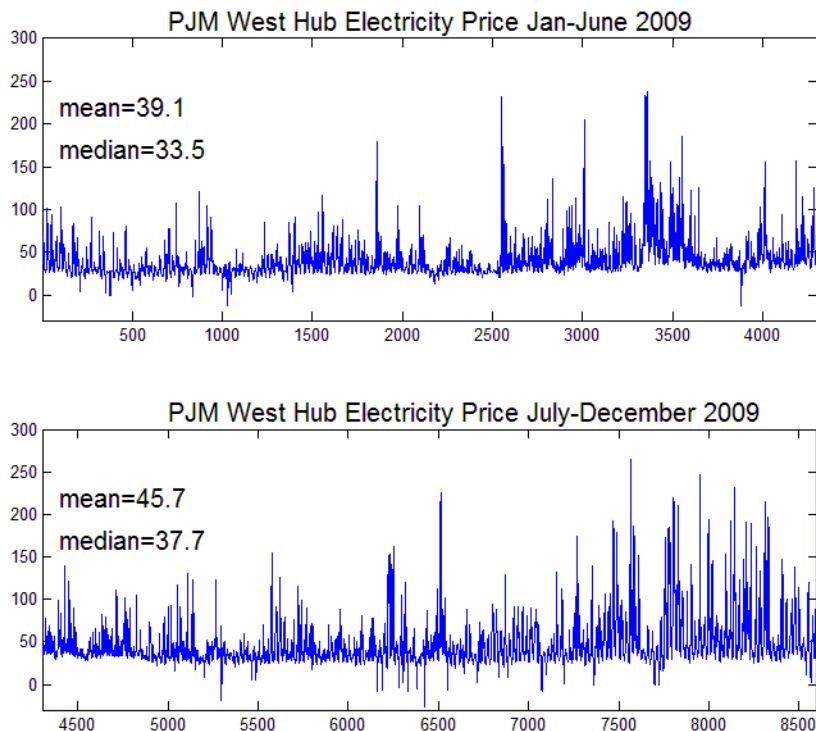


Fig. 1. PJM West Hub electricity price for year 2009.

our conclusions.

1 Motivating Observation

In this section, we present some empirical observations and theoretical discussions that suggest that the popular Gaussian jump-diffusion models may not be applicable to hourly electricity spot prices. The purpose of this section is to show the motivation behind the formal analysis of the electricity price process shown in section §2 and §3 that led to the construction of the heavy-tail model. Throughout this paper, we let $t \in \mathbb{N}_+$ be a discrete time index corresponding to the hourly decision epoch. **Figure 1** shows the plot of the hourly electricity spot market price in PJM West Hub for year 2009. While the median price is around \$35, we see that the price spikes above \$100 occur often enough to skew the empirical mean. The price process seem very volatile, and we cannot dismiss these price spikes as “outliers.” That would make us discard too many data points. However, the electricity price at PJM West Hub is actually relatively stable compared to the price in Ercot Texas shown in **Figure 2**.

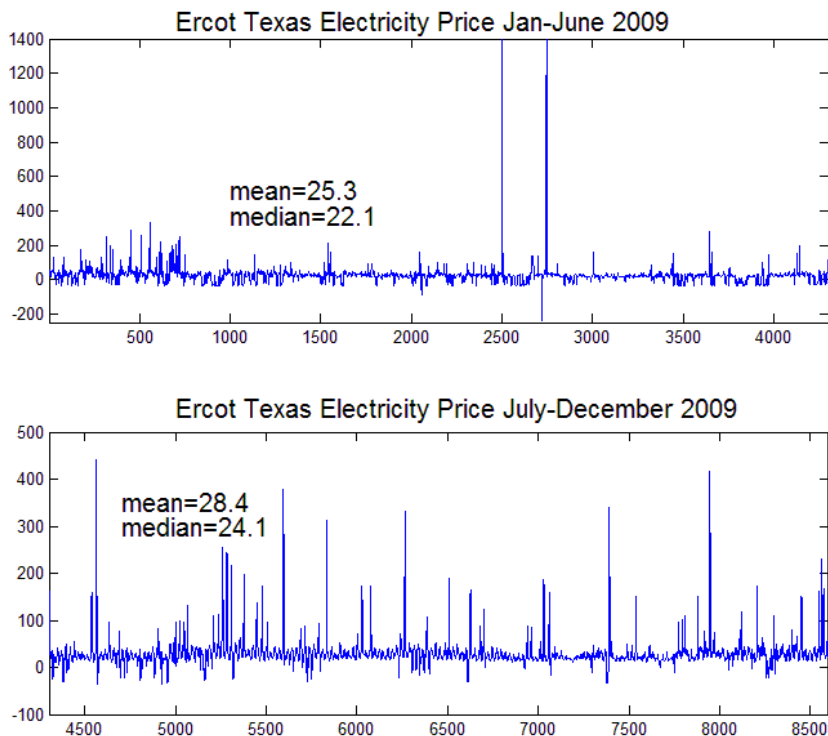


Fig. 2. Ercot Texas electricity price for year 2009.

While the median price is around \$22, the price often spikes above \$100, and occasionally they are above \$200. The maximum price is \$1660. There is even a negative spike where the price goes down to around -\$200. This is possible due to tax subsidies. If we assume that the price process is Gaussian, we can compute the standard deviation by

$$\sigma = \frac{(84.1^{th}\text{-quantile}) - (15.9^{th}\text{-quantile})}{2},$$

which turns out to be 19.4 for the Ercot Texas 2009 data. Then, a price above \$140 is a six-sigma event, which should occur once in a 507 million sample. Out of 8760 samples, we have more than 50 instances when the price is above \$140. A popular model for dealing with this discrepancy is to separate the process into a normal process and a separate jump process with different mean and a larger variance.

1.1 Infinite Variance Tail Process

One such popular model for the electricity price is the jump-diffusion model that superimposes a mean-reverting Wiener (Gaussian) process with an inde-

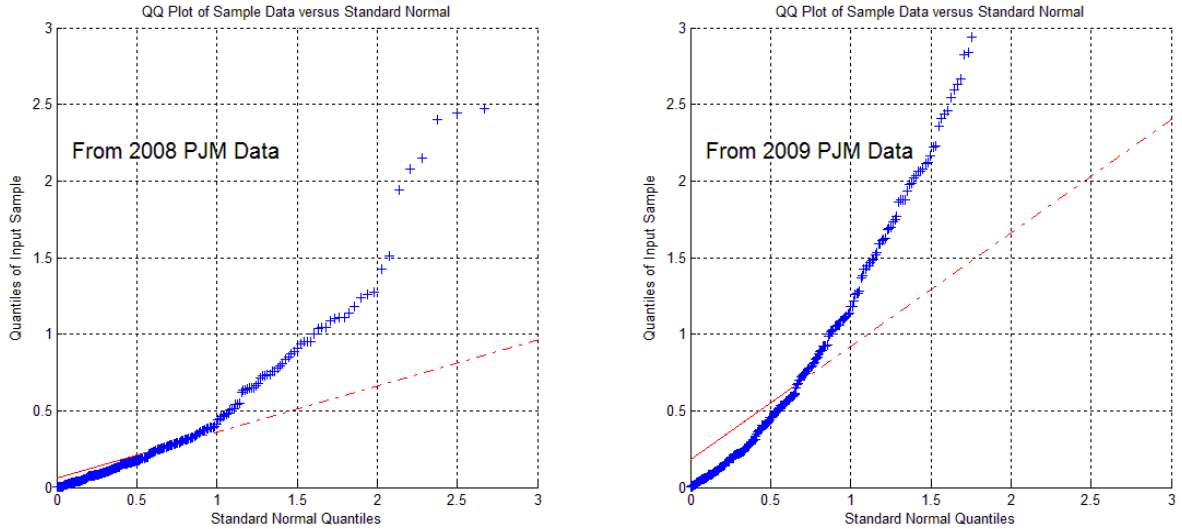


Fig. 3. Q-Q plots of the set $(\nu_t)_t$ versus standard normal computed from PJM West Hub data

pendent jump component ([15], [31]):

$$p_{t+1} = \mu + \kappa(p_t - \mu) + \hat{\varepsilon}_{t+1} + \hat{\nu}_{t+1},$$

where p is the price (or logarithm of the price), μ is the mean, κ is the mean-reversion parameter, $\hat{\varepsilon}$ is the i.i.d Gaussian noise and $\hat{\nu}$ is an independent jump process. After we compute the mean μ and the autoregressive parameter κ through the standard least-squares method, we can compute the combined residual $\hat{\varepsilon}_{t+1} + \hat{\nu}_{t+1}$. Suppose we categorize samples that are three-sigma events as jump processes. That is, the set $(\hat{\nu}_t)_t$ is constructed by taking the sample from the set $(\hat{\varepsilon}_t + \hat{\nu}_t)_t$ that deviates more than three standard deviation from the median. In other words, $(\hat{\nu}_t)_t$ is the tail process. **Figure 3** shows the positive tail of the Gaussian Q-Q plot of the normalized set $(\hat{\nu}_t)_t$ constructed from the PJM West Hub data for year 2008 and 2009. The Q-Q plots show that the tail deviates significantly from Gaussian. The result is similar if we plot the sample against exponential distribution. One might argue that we must break down $(\hat{\nu}_t)_t$ itself further into a core process and a jump process. However, the result will be the same. We can break down the residual process into infinitely many subsets, but we will not be able to fit a Gaussian or any other thin-tailed process. This is because of the tail structure of the residual process. The tail follows a power law, instead of an exponential law. In other words, for large y ,

$$\mathbb{P}[\hat{\nu}_t > y] \approx Cy^{-\beta},$$

for some C and $\beta > 0$.

Using the hourly electricity price in Texas Ercot, a simple figure can demonstrate that the variance is not finite. Given the set $(\hat{\nu}_t)_t$ constructed from the

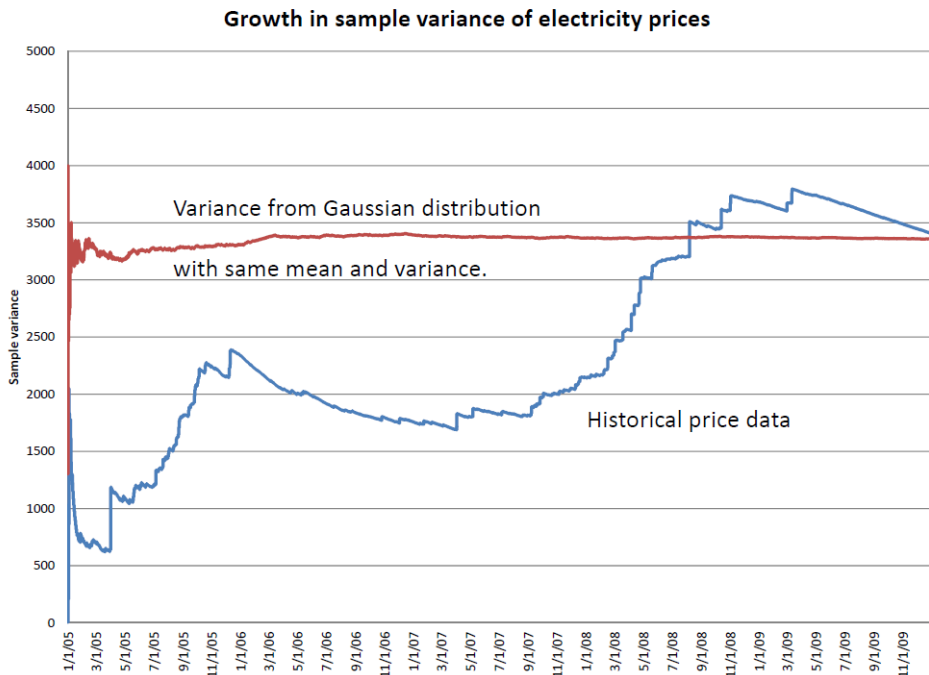


Fig. 4. Divergence of Empirical Variance

electricity price process of Ercot, define

$$\mu_T := \frac{1}{T} \sum_{t=1}^T \hat{\nu}_t.$$

Then, the empirical variance of the process $(\hat{\nu}_t)_t$ computed using data known up to time T is defined as

$$\sigma_T^2 := \frac{1}{T-1} \sum_{t=1}^T (\hat{\nu}_t - \mu_T)^2.$$

If the price process $(\hat{\nu}_t)_t$ has a second moment with dependencies decreasing corresponding to the time-lag, $\lim_{T \rightarrow \infty} \sigma_T^2$ must exist. **Figure 4** shows that the empirical variance computed from historical data does not converge even with tens of thousands of data point; it continues to grow without a bound, suggesting that the variance can be infinite. Of course, if we try to compute the variance, we will get some finite number because we only have finite amount of data.

Besides the empirical evidence against the jump-diffusion model as shown above, there are two additional reasons why jump-diffusion models are not widely adopted by practitioners participating in the electricity market. First, as shown in [6], [11], and [33], calibrating and verifying a continuous jump-diffusion model is notoriously difficult in practice where we do not have continuous data. When fitting jump-diffusion parameters to real data, one has to

separate the jump process from the diffusion process [33], which requires an arbitrary distinction on what counts merely as “big noise” and what counts as a “jump.” Even if we can achieve an agreement on how to distinguish between big noise and a jump, empirical studies show that the jump diffusion model shown above do not fit the real data well ([5], [11], [15], [16], [20] and [23]).

Moreover, empirical analysis of various commodity, electricity, and even equity price data from [1], [3], [5], and [17] also shows that jump-diffusion models do not hold up to reality when we compare the kurtosis implied by the jump-diffusion models with those actually computed from real market data. A typical jump-diffusion model with exponential jumps not only implies that all of the moments are finite, but also implies that we know on average how often the jumps happen and on average how big a jump is when it does happen. A jump-diffusion model allows us to compute various high-order statistics, giving us the illusion of certainty and a false sense of confidence ([37], [38]).

Next, consider the multifactor model ([15], [31])

$$p_{t+1} = \mu_{t+1} + \kappa(p_t - \mu_t) + \hat{\varepsilon}_{t+1}(p_t, t) + \hat{\nu}_{t+1}(p_t, t),$$

where μ_t is the deterministic mean incorporating seasonality while $\hat{\varepsilon}$ and $\hat{\nu}$ are no longer i.i.d - their volatilities and/or other characteristics are changing as the price changes. For the multi-factor model shown above with a separate seasonality component and changing volatilities, it is difficult to come up with a standardized model-fitting procedure that can be broadly accepted by the community.

Since there are infinitely many ways one can construct the volatility surfaces $\hat{\varepsilon}_{t+1}(\cdot)$ and $\hat{\nu}_{t+1}(\cdot)$, with hindsight, we can always find the appropriate volatility surfaces that can reasonably describe the past data. However, in practice, having a lot of freedom is a weakness, not a strength, due to the measurability constraints. When trading in real time, we must use past data to fit a model that is supposed to describe what is likely to happen in the future. If we have a lot of freedom in constructing the volatility surfaces, we are likely *over-fitting to that one particular sample path of past data*, and will not do a good job in describing the electricity price process going forward due to the highly non-stationary and volatile nature of the electricity price. In fact, [3], [13] and [14] have shown that complex jump specifications and changing volatilities add little explanatory power.

1.2 Price Process Has No First Moment

We have shown that the residual tail process $(\hat{\nu}_t)_t$ does not have a finite second moment. However, we know that the interdependence within the set $(\hat{\nu}_t)_t$ is

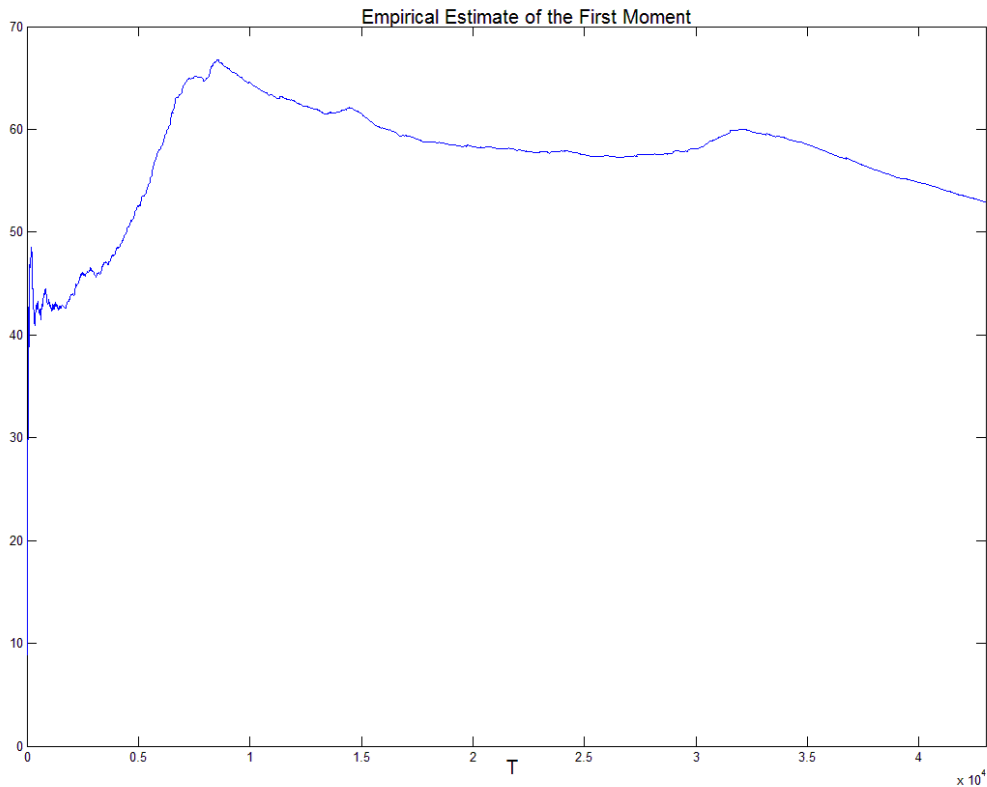


Fig. 5. Empirical Estimate of the First Moment

smaller than the interdependence within the original prices process $(p_t)_t$. The process $(p_t)_t$ is “heavier” than the set $(\hat{v}_t)_t$ due to self-similarity and long range dependence. Let

$$m_T := \frac{1}{T} \sum_{t=1}^T |p_t|$$

be an empirical estimate of the first moment computed using data known up to time T . If the dependence are decreasing corresponding to the time-lag, $\lim_{T \rightarrow \infty} m_T$ must exist. **Figure 5** shows that the empirical estimate of the first moment computed from the Texas Ercot data does not converge even with tens of thousands of data points. The implications of not having a well-defined first moment is profound; the strong law of large numbers is not applicable to the sequence. We know that it is a common practice to estimate the long-term mean of the price using the empirical average,

$$\tilde{m}_T := \frac{1}{T} \sum_{t=1}^T p_t.$$

If the strong law of large numbers holds, the empirical average is a good estimate for the real expectation, because the average using *any* sample path must converge to the real expectation as $T \rightarrow \infty$. Since this is not the case, the empirical average computed from a finite amount of data is just a number

and has no statistical significance. That is, the empirical average computed using the past data does not help us ascertain anything about the expectation of the price in the future.

1.3 Empirical Averages and Seasonality

When working with traditional models, which are linear and additive, it is common practice to break down the mean into several components. One often assumes that the true mean within the mean-reverting model is a superposition of the long-term mean, the seasonal (monthly) mean, and the mean corresponding to what hour of the day or the day of the week it is ([15], [20]). Such an approach is possible in the first place because the assumed model is linear and it is often assumed that we can readily compute the expectation. The expectation is a linear operator and hence the expectation of the sum is the sum of the expectations.

If we compute the empirical average of all the past prices at 4pm, that number is likely to be greater than the empirical average computed using all the past prices at 2am, for example. Thus, it is often assumed that there is a cycle depending on the hour of the day and day of the week. However, if the strong law of large numbers does not hold, there is not much statistical information we can get from empirical averages. While it is true that electricity prices are generally higher in the busy afternoon than earlier in the day, that kind of deterministic variation depending on the hour of the day and day of the week cannot be reliably quantified. We can demonstrate this point through the following experiment.

Define

$$d_m := \text{difference between the 4pm price and the 2am price on day } m.$$

Figure 6 shows the plot of

$$\kappa_T := \frac{1}{T} \sum_{m=1}^T |d_m|$$

for $1 \leq T \leq 1500$, for Texas Ercot data. Even after 1500 days, difference between the 4pm price and the 2am price is not converging. This implies that the law of large numbers is not applicable and we cannot quantify the expected difference between 4pm price and 2am price using empirical averages.

While empirical averages are unreliable, one way to observe the dependence on the hour of the day is to compute the quantiles. **Table 1** shows that while there is a difference between 2am and 4pm in the median, the difference is

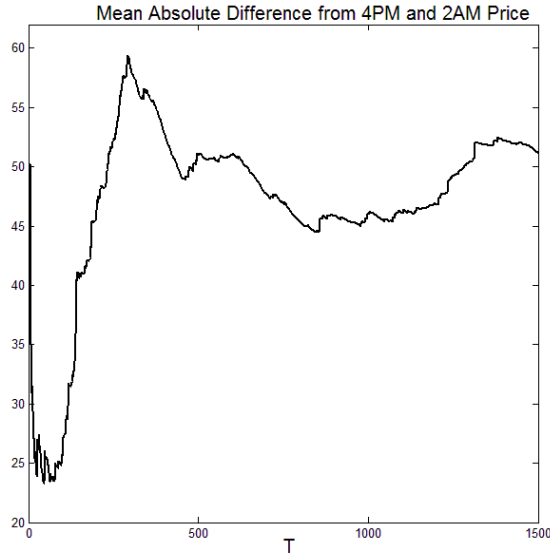


Fig. 6. Difference between 4pm price and 2am price

	5 th	25 th	50 th	75 th	95 th
4PM	14.1	37.1	55.9	80.3	183.5
2AM	-18.3	20.7	38.2	52.0	122.8
4PM - 2AM	-32.4	3.5	22.8	43.4	136.9
6PM	17.3	41.6	59.7	81.5	139.7
12PM	-21.0	29.2	45.3	58.7	90.5
6PM - 12PM	-24.3	3.3	17.3	33.9	114.8

Table 1

Quantiles of price data at specific times of day, and the quantiles of the changes in prices, showing that the differences in the distribution of prices is quite different from the distribution of the differences.

small in the overall spread between .9-quantiles and .1-quantiles. Moreover, one can see that the quantile of the difference d_m is fairly different from the quantile of 4pm prices minus the quantile of 2am prices. Thus, unlike the usual linear additive Gaussian models, the 2am price and the “4pm price minus 2am price” are not independent. About 20% of the time, the 2am price is actually greater than the 4pm price. Given the 2am price, we cannot quantify what our prediction for the 4pm price should be, neither in the expectation nor in the median.

While we believe linear methods are unreliable, we want to see what kind of results we get if we employ the usual standard methods, anyway. Define

$$\begin{aligned}\beta_0 &:= p_0, \\ \beta_t &:= (1 - \alpha)\beta_{t-1} + \alpha p_t,\end{aligned}$$

to be the exponentially smoothed baseline for some $0 < \alpha \leq 1$. We compute α such that mean-absolute error for the one-step prediction

$$g_T := \frac{1}{T} \sum_{t=0}^T |p_t - \beta_{t-1}|$$

is minimized. The value of α will depend on the specific data set. Again, we do not believe it is meaningful to use such criteria for optimization in a heavy-tailed environment since we cannot reliably compute an approximation to the first moment. However, similar approaches are widely used in the industry, hence it is important to investigate how well such method can do. Next, define

$$\delta_t := p_t - \beta_{t-1}$$

to be the deviation from the baseline. Note that $24 \times 7 = 168$ and define

$$\varphi_t := \frac{1}{6} (\delta_{t+1-168} + \delta_{t+1-168 \times 2} + \dots + \delta_{t+1-168 \times 5} + \delta_{t+1-168 \times 6})$$

to be the average deviation from the baseline that depends on both the hour of the day and the day of the week. In this context, we have determined that using six weeks worth of data in a rolling horizon basis is reasonable because it does not go back so far enough so that we are affected by long-term seasonal variation (spring versus summer, for example) while having six data points somewhat smooths-out the “noise.” Note that $24 \times 7 \times 6 = 1008$. Define

$$\lambda_t := \frac{1}{42} (\delta_{t+1-24} + \delta_{t+1-24 \times 2} + \dots + \delta_{t+1-24 \times 41} - \delta_{t+1-24 \times 42})$$

to be the average deviation from the baseline that depends only on the hour of the day. We compare how well the three estimators β_t , $\beta_t + \varphi_t$, and $\beta_t + \lambda_t$ does in predicting p_{t+1} . We want to check if the dependence on the hour of the day and the day of the week can be captured by empirical averages so that either $\beta_t + \varphi_t$ or $\beta_t + \lambda_t$ does at least better than β_t . Define

$$\begin{aligned}A_T &:= \frac{1}{T - 1007} \sum_{t=1008}^T |p_{t+1} - (\beta_t + \lambda_t)|, \\ B_T &:= \frac{1}{T - 1007} \sum_{t=1008}^T |p_{t+1} - (\beta_t + \varphi_t)|, \\ C_T &:= \frac{1}{T - 1007} \sum_{t=1008}^T |p_{t+1} - \beta_t|,\end{aligned}$$

to be the mean-absolute error terms for the three estimators. We use mean absolute deviations since the mean squared deviations tend to be dominated

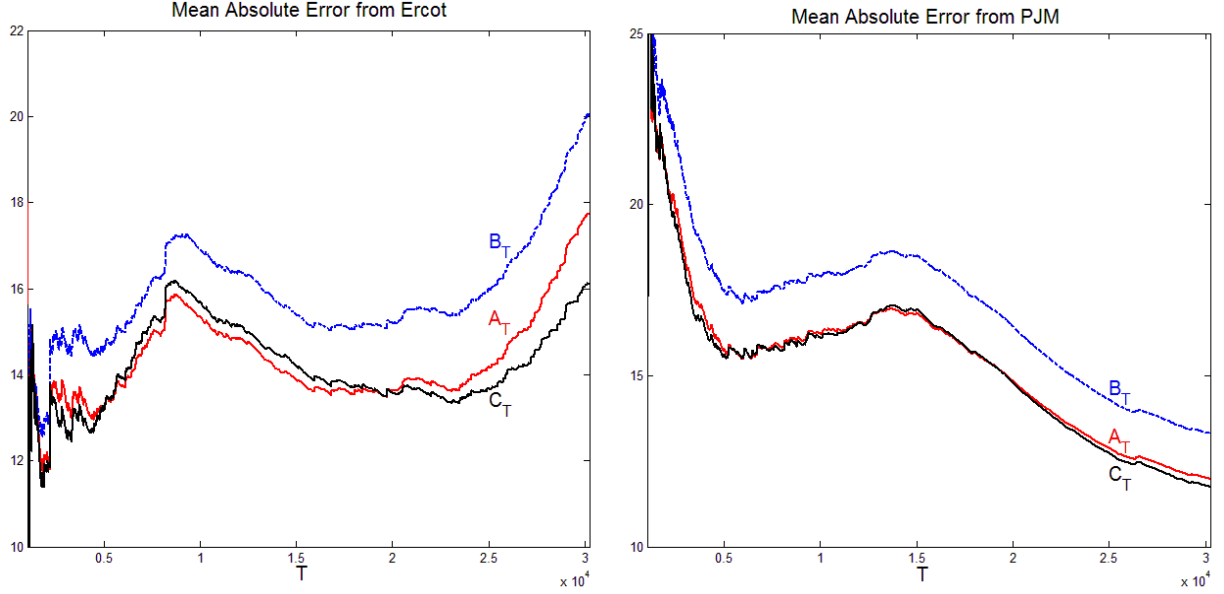


Fig. 7. Hour-Ahead Mean Absolute Error Estimate

by a small number of large spikes. **Figure 7** shows that for both Ercot and PJM data, β_t is the best predictor for p_{t+1} while $\beta_t + \lambda_t$ comes in second and $\beta_t + \varphi_t$ is the worst predictor for p_{t+1} . Factoring in the hour of the day and the day of the week computed from the empirical averages seems to have no statistical value.

While the goal of this paper is to model the price process and the decision making process on an hour-ahead basis, it is also interesting see how well these predictors do on a day-ahead basis. We want to compare how β_t , $\beta_t + \varphi_{t+23}$, and $\beta_t + \lambda_{t+23}$ do as a predictor for p_{t+24} . Define

$$A_T := \frac{1}{T - 1007} \sum_{t=1008}^T |p_{t+24} - (\beta_t + \lambda_{t+23})|$$

$$B_T := \frac{1}{T - 1007} \sum_{t=1008}^T |p_{t+24} - (\beta_t + \varphi_{t+23})|$$

$$C_T := \frac{1}{T - 1007} \sum_{t=1008}^T |p_{t+24} - \beta_t|$$

to be the mean-absolute error terms for the three day-ahead estimators. **Figure 8** shows that the results are similar for day-ahead predictions. We conclude that for the hourly spot market price, empirical averages are not useful in capturing the seasonality. In a heavy-tailed environment, higher order statistics such as the mean and the variance are unreliable. The only robust and verifiable statement we can make is that of the zeroth order statistics. For example, we can say something like the following: the probability that the electricity

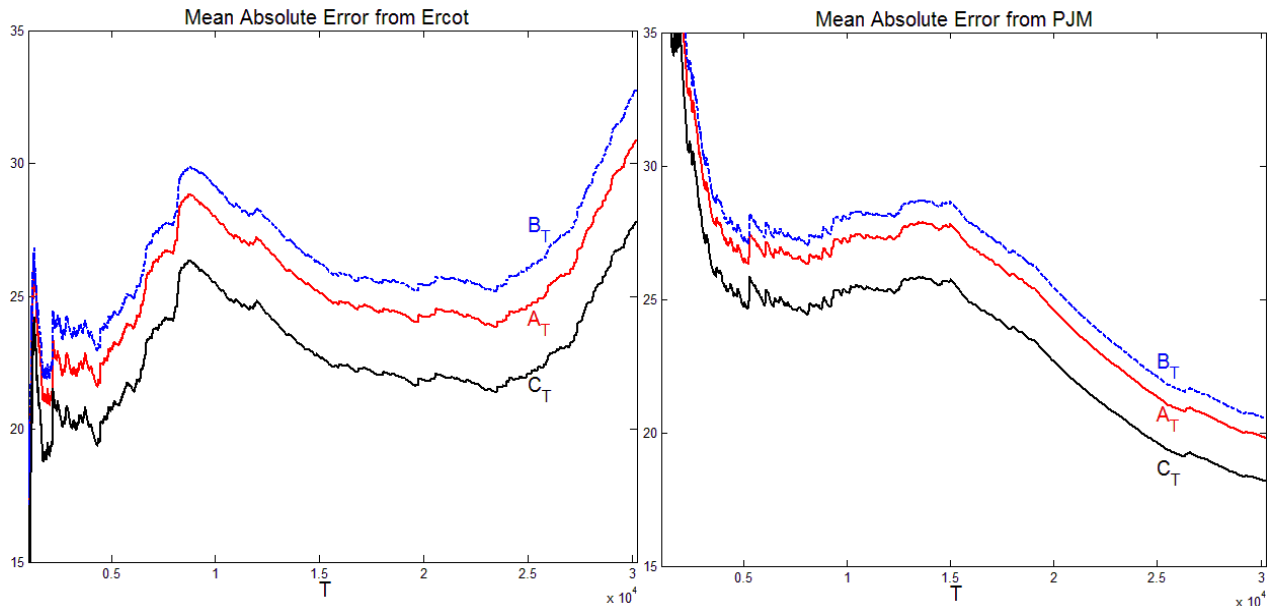


Fig. 8. Day-Ahead Mean Absolute Error Estimate

price in 4pm is greater than the electricity price at 2am is 80%. How much greater, though, is difficult to quantify both in the expectation or in the median. Suppose the empirical average of the past 4pm prices is \$50 while the empirical average of the past 2am prices is \$20. This does not imply that there will be on average a \$30 difference between the 4pm prices and 2am prices in the future. Empirical averages are just some numbers when the strong law of large numbers is not applicable.

We have also studied the affect of correcting for the hour-of-the-day in the traditional Gaussian jump-diffusion process, whose parameters were fitted in the standard way - minimizing the mean-squared error. We compared the mean-squared errors between a Gaussian jump diffusion process with and without the hour-of-the-day correction using the PJM West hub data. The square root of the mean square prediction error was 27.02 and the 95% confidence interval was (26.38, 27.66) when we did not account for the hour-of-the-day affect. The square root of the mean squared prediction error was 26.72 and the 95% confidence interval is (26.04, 27.40) when we did correct for the hour-of-the-day affect by subtracting the empirical means for every hour. Thus, we can not say using the empirical mean to correct for the hour-of-the-day affect has made a statistical difference. And of course, we believe that those numbers do not really have statistical meaning anyway since the law of large numbers do not hold.

When we are buying and selling electricity on an hourly basis, the deterministic dependence on hour of the day or the day should not be our focus. We may buy electricity from 2am to 10am and sell electricity from 4pm to 10pm. Such strategy is certainly profitable. However, we do not want to sit back and

ignore opportunities that arise from “random” variations that can happen in any hour. Even if it is 2am, if the price is high, we would rather sell than buy. Much larger profits can be made by focusing on the big movement - the surprising heavy-tailed spikes - that may occur at any hour.

2 Analysis of the Trailing Median Process

When we analyze heavy-tailed data, the only information we can get that is robust and reliable is that of the zeroth order statistics - specifically, the probability distribution and the quantile function that gives us the median, among other things. This is because computing a quantile function only requires us to sort and count and is thus not affected by the occasional heavy-tail events with large impacts. Higher-order statistics are unstable and unreliable because they can be affected by a few heavy-tail samples. In this paper, we develop an electricity price model based on the evolution of the median of the price.

2.1 Definition and Observation

First, we assume that price changes due to seasonality are negligible within a one month period and hence the median of the price during that period is approximately constant. Given t , suppose $(p_{t'})_{t-719 \leq t' \leq t}$ is some hourly price data over the past one month period, i.e., 720 hours. For each t , define

$$\tilde{\mu}_t := \text{median of the set } (p_{t'})_{t-719 \leq t' \leq t}.$$

We refer to $\dots, \tilde{\mu}_{t-1}, \tilde{\mu}_t, \tilde{\mu}_{t+1}, \dots$ as the *monthly TMP*. By construction, the monthly TMP is a thin-tailed process. Define

$$\mu := \frac{1}{T} \sum_{t=1}^T \tilde{\mu}_t$$

to be the sample mean of the monthly TMP, where T is the length of the available data. μ is the long-term average of the price process. **Figure 9** shows plots of the monthly TMP constructed from the Texas electricity price data for year 2006 through 2009. The sample mean is fairly consistent from year to year except for year 2009, when the market displayed great instabilities with a lot of negative price spikes (the electricity price can be negative due to tax subsidies). However, one can see that towards the end of 2009, the price is climbing back up towards the long-term mean. From **Figure 9**, we can observe that the monthly TMP reverts to the mean in the long term, but the

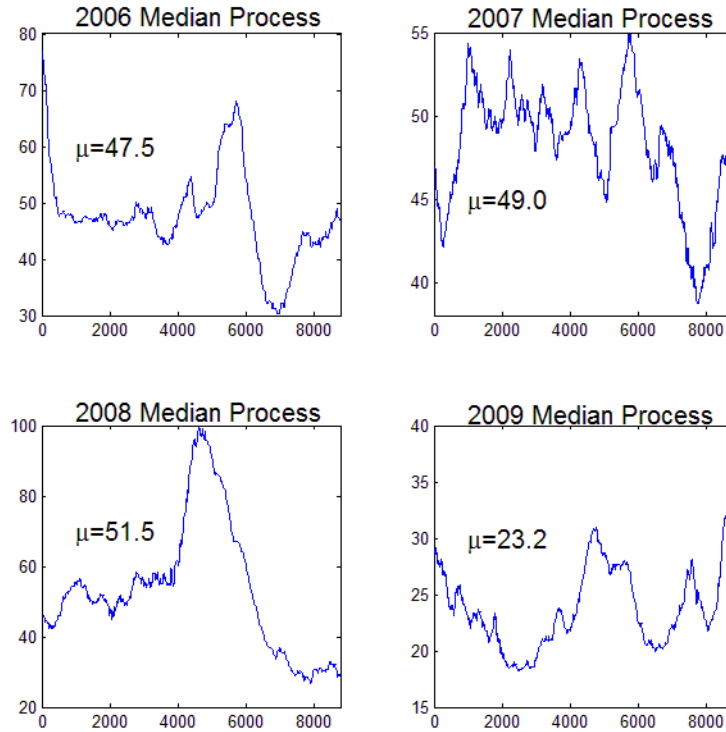


Fig. 9. Median process constructed using the hourly electricity price data from Texas

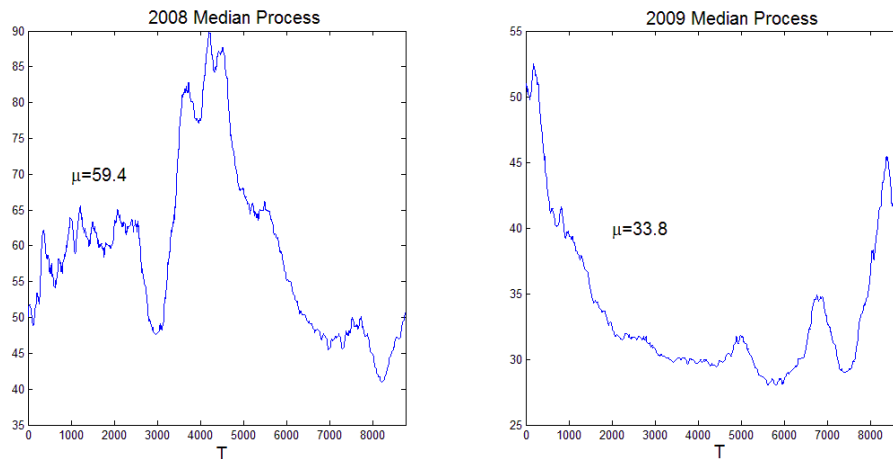


Fig. 10. Median process constructed using the hourly electricity price data from PJM West hub

force of mean-reversion is very weak such that if we look at the plot for a one month period, for example, we would not see any evidence of mean-reversion.

For the purpose of one-step prediction, suppose we assume that the monthly TMP is a martingale. Then, at time t , our best guess for $\tilde{\mu}_{t+1}$ is $\tilde{\mu}_t$, and the

mean absolute error in our prediction would have been

$$\frac{1}{T-1} \sum_{t=1}^{T-1} \left| \frac{\tilde{\mu}_{t+1} - \tilde{\mu}_t}{\tilde{\mu}_{t+1}} \right| = 6.40 \times 10^{-4}$$

for the four years spanning 2006 through 2009. That is a prediction error of 0.064%. **Figure 10** shows the TMP for PJM West data for year 2008 and 2009. For years 2007 through 2009, the prediction error from assuming the TMP is a martingale is 3.93×10^{-6} . However, **Figures 9** and **10** shows that the monthly TMP is mean-reverting. Can we do a better job at one-step prediction using a mean-reverting model instead of assuming that the monthly TMP is a martingale? The answer is: *theoretically yes, but practically no*. Suppose

$$\tilde{\mu}_{t+1} = \mu + \beta (\tilde{\mu}_{t+1} - \mu) + \hat{\nu}_{t+1}, \quad \forall t,$$

where $(\hat{\nu}_t)_{t \geq 1}$ is i.i.d with $\mathbb{E}[\hat{\nu}_t] = 0$. Then, $(\tilde{\mu}_t)_{t \geq 1}$ is a martingale if $\beta = 1$ and mean-reverting if $\beta < 1$. If we try to find β using the standard least-squares method on the four years of data from Texas, we get $\beta = 1.0021$. If we break down the data set for each year or a season, we get β ranging from .98 to 1.03. The numbers are similar for the PJM West data. However, if we take a bottom-up approach and try to simulate sample paths that look similar to the plots shown in **Figures 9** and **10**, we realize that the true value of β should be less than 1, but very close to 1. That will ensure the process to be mean-reverting, but with a very weak force of mean-reversion so that it can take several thousand hours for the TMP to approach μ . Therefore, for small $n \in \mathbb{N}_+$, $\beta^n \approx 1$, and we can assume that the TMP is locally a martingale.

If we had an infinite amount of data, we could theoretically compute the exact β and forecast TMP for the entire future, but it is not possible in practice. As long as we only have a finite amount of data, there will be an error in the estimate, and even a small error will be compounded if we try to forecast even for just a few weeks into the future. Thus, we do not care exactly what β is. We just need to know that the TMP is approximately a martingale in the short term while it approaches μ in the long term. The fact that $(\tilde{\mu}_t)_{t \geq 1}$ is approximately a martingale follows naturally from the way we defined $(\tilde{\mu}_t)_{t \geq 1}$; a median of a set is a very stable measure. The monthly TMP can be seen as the seasonal component of the electricity price.

2.2 Dependence on the Hour of the Day

As mentioned above, when working with the traditional models, which are linear and additive, it is a common practice to break down the mean into several components. However, the premise of our paper is that when the data is heavy-tailed, the mean either does not exist or is very difficult to compute

	10 th	25 th	50 th	75 th	90 th
3 AM	42.60	46.21	49.13	53.09	63.98
7 AM	42.61	46.23	49.10	53.04	63.90
11 AM	42.60	46.21	49.06	53.04	63.90
3 PM	42.64	46.17	49.06	53.04	63.98
7 PM	42.60	46.17	49.10	53.04	63.98
11 PM	42.57	46.21	49.14	53.00	63.98

Table 2

Dependence of $\tilde{\mu}_t$ on hour of the day for Texas Ercot

with a finite amount of data. In this case, the empirical mean is just a number that does not represent an estimate of a parameter in the model such as the true mean. Therefore, we work with the median instead, which by definition is robust and always exists. Our median, the monthly TMP, is computed from a rolling-horizon of 720 hours of data, so it roughly accounts for the seasonal (monthly) fluctuation. Suppose we want to further break down the median in our median-reverting model to multiple components: the long-term median, the seasonal median and the median corresponding to what hour of the day it is. There is no obvious way of decomposing and superimposing the different components. The sum of the medians is not the median of the sum and the median of a conditional median is not the median. However, we can show that the process $(\tilde{\mu}_t)_{t \geq 1}$ does not depend on what hour of the day or day of the week it is. We separate the set $(\tilde{\mu}_t)_{t \geq 1}$ into 24 subsets corresponding to the hour of the day and analyze the subsets to see if there is a meaningful difference among them. For each of the 24 subsets, we computed the 10th, 25th, 50th, 75th and the 90th-percentile. **Table 2** shows the result for some selected hour of the day from the Texas Ercot data. The table shows that the probability distribution of $\tilde{\mu}_t$ does not depend on the hour of the day. As shown in (1) above, the difference between 4pm prices and 2am prices are substantial only in quantiles somewhat away from the median, which does not affect $(\tilde{\mu}_t)_{t \geq 1}$. We have purposely defined $(\tilde{\mu}_t)_{t \geq 1}$ so that it is not affected by large deviations in the tail-end of the quantiles.

The results are similar if we separate the set $(\tilde{\mu}_t)_{t \geq 1}$ into 7 subsets corresponding to the day of the week and analyze their statistical properties. Thus, there is no reason to further break down the set $(\tilde{\mu}_t)_{t \geq 1}$ as we normally do when working with the mean in a linear additive model.

Of course, this does not mean that the price itself, $(p_t)_{t \geq 1}$, does not depend on the hour of the day. **Figure 11** shows the number of times within the five year span 2005-2009 that the electricity price exceeded above \$100 depending on the hour of the day. For 5pm, the price exceeds \$100 about 16% of the time while for 5am, the price exceeds \$100 about 2% of the time. However, the

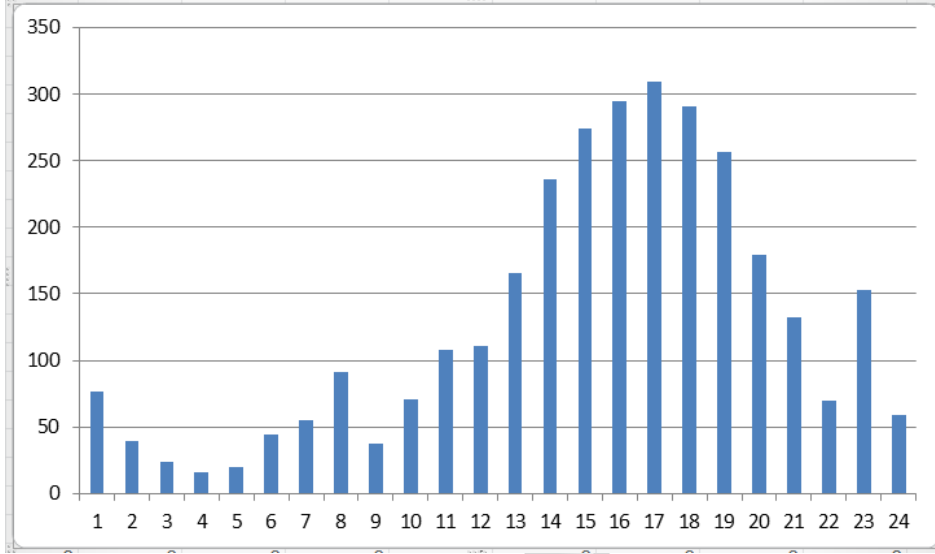


Fig. 11. Number of times price exceeded above \$100

data shown in **Figure 11** does not give us practical information that can be used in real-time trading because the hourly electricity prices are not actually i.i.d. If it is currently 4pm, then the probability that the 5pm price exceeds \$100 will depend heavily on what the current price level at 4pm is. If it is currently 4am and the electricity price is above \$200, the probability that the 5am price is above \$100 will be much larger than 2%. If it is currently 4pm and the electricity price is \$5, the probability that the 5pm price is greater than \$100 will be much lower than 16%. We are trying to build an hour-ahead prediction model, and we must construct our time series carefully. This will be done in the following two sections.

3 Analysis of the Residual Process

By the definition of $\tilde{\mu}_{t+1}$, which is computed from $(p_{t'})_{t-718 \leq t' \leq t+1}$, we know that $\tilde{\mu}_{t+1}$ will be close to $\tilde{\mu}_t$, which is computed from $(p_{t'})_{t-719 \leq t' \leq t}$, regardless of what p_{t-719} or p_{t+1} are. Thus, given $(p_{t'})_{t-719 \leq t' \leq t}$ at time t , $\tilde{\mu}_t$ is our best guess for $\tilde{\mu}_{t+1}$. However, at time t , what we really want to know is the probability distribution of p_{t+1} . Define

$$Y_t := \frac{p_t - p_{t-1}}{\tilde{\mu}_{t-1} - p_{t-1}}, \quad \forall t, \quad (1)$$

as the ratio between the change in price and the distance between the price level and the TMP. We assume that the time series $\dots, Y_{t-1}, Y_t, Y_{t+1}, \dots$ is stationary ergodic with the strong Markov property and constant median.

In this section, we show that the time series $\dots, Y_{t-1}, Y_t, Y_{t+1}, \dots$ is approxi-

mately “independent” and hence it can be categorized as the noise that derives the random evolution of the price process. Then, we compute its tail-index to show that the set is heavy-tailed.

3.1 Serial Dependence Test for Heavy-Tailed Data

Define

$$\tilde{\mu}_Y := \text{med}(Y_t) = \text{median of the set } \{\dots, Y_{t-1}, Y_t, Y_{t+1}, \dots\}, \forall t.$$

Then,

$$\mathbb{P}[Y_t - \tilde{\mu}_Y > 0] = \mathbb{P}[Y_t - \tilde{\mu}_Y < 0] = \frac{1}{2}, \forall t.$$

We want to verify that $\dots, Y_{t-1}, Y_t, Y_{t+1}, \dots$ are the correct residuals that capture the random evolution of the price process $(p_t)_{t \geq 0}$ by showing that the serial dependence within the time series $\dots, Y_{t-1}, Y_t, Y_{t+1}, \dots$ is negligible. Specifically, we want to show that the best median-predictor of p_{t+1} at time t is

$$\begin{aligned} \text{med}(p_{t+1}) &= p_t + \text{med}(Y_{t+1}) (\tilde{\mu}_t - p_t) \\ &= p_t + \tilde{\mu}_Y (\tilde{\mu}_t - p_t). \end{aligned}$$

We can achieve the above if we can show that $\dots, Y_{t-1}, Y_t, Y_{t+1}, \dots$ are mutually independent. However, in practice, it is not possible to test the independence hypothesis against all possible alternatives. Instead, we resort to the sign test shown in [39], which tests the null hypothesis that $\dots, Y_{t-1}, Y_t, Y_{t+1}, \dots$ are mutually independent against the alternative that these variables are positively or negatively serially dependent (we explain what we mean by this shortly).

Define

$$Z_{t,k} := (Y_t - \tilde{\mu}_Y)(Y_{t+k} - \tilde{\mu}_Y).$$

If $\dots, Y_{t-1}, Y_t, Y_{t+1}, \dots$ are mutually independent, then we can show that

$$\begin{aligned}
\mathbb{P}[Z_{t,k} > 0] &= \mathbb{P}[Y_t - \tilde{\mu}_Y > 0, Y_{t+k} - \tilde{\mu}_Y > 0] \\
&\quad + \mathbb{P}[Y_t - \tilde{\mu}_Y < 0, Y_{t+k} - \tilde{\mu}_Y < 0] \\
&= \mathbb{P}[Y_{t+k} - \tilde{\mu}_Y > 0 \mid Y_t - \tilde{\mu}_Y > 0] \mathbb{P}[Y_t - \tilde{\mu}_Y > 0] \\
&\quad + \mathbb{P}[Y_{t+k} - \tilde{\mu}_Y < 0 \mid Y_t - \tilde{\mu}_Y < 0] \mathbb{P}[Y_t - \tilde{\mu}_Y < 0] \\
&= \mathbb{P}[Y_{t+k} - \tilde{\mu}_Y > 0] \mathbb{P}[Y_t - \tilde{\mu}_Y > 0] \\
&\quad + \mathbb{P}[Y_{t+k} - \tilde{\mu}_Y < 0] \mathbb{P}[Y_t - \tilde{\mu}_Y < 0] \\
&= \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \\
&= \frac{1}{2}, \quad \forall k \geq 1.
\end{aligned}$$

We define our null hypothesis as

$$\mathcal{H}_0 : \text{med}(Z_{t,k}) = 0, \quad \forall k \geq 1 \text{ and } \forall t.$$

We say that the sequence $\dots, Y_{t-1}, Y_t, Y_{t+1}, \dots$ has positive serial dependence if

$$\text{med}(Z_{t,k}) > 0, \quad \forall k \geq 1 \text{ and } \forall t,$$

and negative serial dependence if

$$\text{med}(Z_{t,k}) < 0, \quad \forall k \geq 1 \text{ and } \forall t.$$

Instead of computing the autocovariance, which is a second-order statistic, we are computing what we may call the ‘autoco-median,’ which is a zeroth-order statistic. Suppose we index the residuals from 1 to n : Y_1, Y_2, \dots, Y_n . Define

$$q_k := \frac{1}{n-k} \sum_{t=1}^{n-k} \mathbf{1}\{Z_{t,k} > 0\}.$$

where $\mathbf{1}\{\cdot\}$ is the indicator function. Note that $q_k \rightarrow \mathbb{P}[Z_{t,k} > 0]$ for large n . Then, under \mathcal{H}_0 , $q_k \rightarrow \frac{1}{2}$, for large n .

In practice, however, it is usually sufficient to check just the statistics corresponding to the first lag $k = 1$. If q_1 is larger than $\frac{1}{2}$, that implies $(Y_t)_{t \geq 1}$ has positive serial dependence. If q_1 is significantly smaller than $\frac{1}{2}$, that implies $(Y_t)_{t \geq 1}$ has negative serial dependence. Using the hourly electricity price data from Texas, q_1 turns out to be .5233, .5311, .5375, and .5311, for the years 2006, 2007, 2008, and 2009, respectively. q_2 turns out to be .5078, .5057, .5042, and .5021, respectively. For $k \geq 3$, q_k ’s are all .5+/-0.001. From q_1 , we can see that the residual $(Y_t)_{t \geq 1}$ has a small remnant of positive dependence. However, for the purpose of constructing an electricity price model used for trading, we believe that the dependence is small enough so that we can assume the time series $(Y_t)_{t \geq 1}$ is “purely random.”

	year	2006	2007	2008	2009
$(Y_t)_{t \geq 1}$,	$k = 1$.5233	.5311	.5375	.5311
	$k = 2$.5078	.5057	.5042	.5021
	$k = 3$.5000	.5000	.5000	.5000
$(p_t)_{t \geq 1}$,	$k = 1$.8442	.8492	.8680	.8655
	$k = 2$.7729	.7718	.8020	.7896
	$k = 3$.7081	.7070	.7539	.7254
$(p_t/p_{t-1})_{t \geq 1}$,	$k = 1$.5622	.5934	.5778	.5967
	$k = 2$.5494	.5559	.5490	.5432
	$k = 3$.5108	.5159	.5086	.5212
$(p_t - p_{t-1})_{t \geq 1}$,	$k = 1$.5617	.5928	.5745	.6007
	$k = 2$.5507	.5543	.5371	.5406
	$k = 3$.5095	.5118	.4934	.5154

Table 3

Serial dependence test for various time series constructed from the Texas price data

Table 3 shows the serial dependence in our choice of residual $(Y_t)_{t \geq 1}$, as well as the serial dependence on the original set $(p_t)_{t \geq 1}$ and the simple difference sets $(p_t - p_{t-1})_{t \geq 1}$ and $(p_t/p_{t-1})_{t \geq 1}$, for the Texas region. Of course, there are many ways one can construct a residual set, and it is not possible to ever prove that our choice of residual $(Y_t)_{t \geq 1}$ is the best against all possible alternatives. Among the aforementioned sets, however, the set $(Y_t)_{t \geq 1}$ turns out to have the least amount of serial dependence among them. Moreover, it turns out that the distribution of the set $(Y_t)_{t \geq 1}$ can be approximated by a well-known probability distribution, as will be shown in §4. **Table 4** shows the serial dependence for PJM West hub.

3.2 Tail Index Estimation

Next, we estimate the tail index of the time series $(Y_t)_{t \geq 1}$ to see how heavy its tail is. The most widely used tail index estimator is Hill's estimator ([10], [19]). An extensive study of Hill's estimator have shown that the estimator is asymptotically optimal as long as the dependence between the random variables weaken as the time separation becomes larger ([21], [34]). Since our data indicate that the dependence is essentially non-existent for $k > 2$, the Hill's estimator is applicable to our data. According to the Hill's estimator, the tail index is found to be close to 1.

	year	2007	2008	2009
$(Y_t)_{t \geq 1}$,	$k = 1$.5119	.5368	.5218
	$k = 2$.5076	.5052	.5012
	$k = 3$.5000	.5000	.5000
$(p_t)_{t \geq 1}$,	$k = 1$.8091	.8257	.8764
	$k = 2$.7381	.7523	.8114
	$k = 3$.6935	.7038	.7655
$(p_t/p_{t-1})_{t \geq 1}$,	$k = 1$.6043	.5811	.5897
	$k = 2$.5459	.5533	.5471
	$k = 3$.5151	.5168	.5233
$(p_t - p_{t-1})_{t \geq 1}$,	$k = 1$.4736	.4750	.5534
	$k = 2$.4891	.4921	.5190
	$k = 3$.5008	.5079	.5140

Table 4

Serial dependence test for various time series constructed from the PJM West price data

Assume $(Y_t)_{t \geq 1}$ have the same marginal distributions and there exists an $\alpha > 0$ such that

$$\lim_{y \rightarrow \infty} \frac{\mathbb{P}[Y_t > \lambda y]}{\mathbb{P}[Y_t > y]} = \frac{1}{\lambda^\alpha}, \text{ for all } \lambda > 0.$$

This implies that $\mathbb{P}[Y_t > y] \approx Cy^{-\alpha}$ for some $C > 0$ and large y . Then, α is called the tail index and $\text{var}(Y_t) = \infty$ if $0 < \alpha < 2$.

Let $Y_{(i)}$ denote the order statistics, i.e., the results of ordering the original sample in the increasing order:

$$Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(n)}.$$

Let $1 \leq r_1 < r_2 \leq n$ be the index such that

$$\ln \mathbb{P}[Y_t > Y_{(i)}] \approx \ln C - \alpha \ln Y_{(i)},$$

for $r_1 \leq i \leq r_2$. That is, $Y_{(i)}$ for $r_1 \leq i \leq r_2$ are the data points in the tail part whose distribution follows the power law as described above. The procedure for identifying these data points, otherwise known as extreme order statistics, can be found in [8], for example. Then, using the approximation

$$\mathbb{P}[Y_t > Y_{(n-i+1)}] \approx \frac{i}{n},$$

we know that the slope that fits the data points $(\ln \frac{i}{n}, \ln Y_{(n-i+1)})$ for $r_1 \leq i \leq$

α -quantile	.1	.2	.3	.5	.7	.8	.9
2am	-20.9	-8.4	-3.4	-.22	4.7	11.2	26.0
12pm	-24.5	-10.1	-5.1	-.09	5.6	11.9	24.2
4pm	-20.3	-7.1	-2.7	0.3	4.0	9.3	19.1
8pm	-21.2	-7.3	-2.1	0.9	6.5	12.3	20.8

Table 5

Quantile function of $(Y_t)_t$ depending on the hour of the day constructed from Ercot data

r_2 approximates $-1/\alpha$. Then, the Hill's estimator for the tail index is

$$\hat{\alpha} = \left[\frac{1}{r_2 - r_1 + 1} \sum_{i=r_1}^{r_2} (\ln Y_i - \ln Y_{r_1}) \right]^{-1},$$

which is the conditional maximum likelihood estimator for α . Using the hourly electricity price data from Texas, $\hat{\alpha}$ turns out to be .94, 1.02, .90, and .95, for the years 2006, 2007, 2008, and 2009, respectively. Using the hourly electricity price data from PJM West, $\hat{\alpha}$ turns out to be 1.12, 1.08, 1.13, for the years 2007, 2008, and 2009, respectively. It is common to use the student-t distribution to represent a heavy-tailed random variable whose support lies on the real line. One special case of the student-t distribution is the Cauchy distribution, which corresponds to the case where the tail index of the student-t distribution is $\alpha = 1$. Thus, if we assume $(Y_t)_{t \geq 1}$ has a student-t distribution, we know that its distribution can be approximated by the Cauchy distribution. In the next section, we show empirical evidence that indeed the Cauchy distribution fits the distribution of $(Y_t)_{t \geq 1}$ well.

3.3 Dependence on the Hour of the Day

Next, we want to know if $(Y_t)_t$ depends noticeably on the hour of the day. First, even if we separate the set $(Y_t)_t$ into 24 subsets, each of the subsets has tail index ranging from .9 to 1.2. **Table 5** shows the quantile function of $(Y_t)_t$ computed from Texas Ercot data. We can see that there is no significant difference based on the hour of the day and it is not obvious how we would make corrections based on the hour of the day. The results are similar for PJM West Hub data. It is important to remember that $(Y_t)_t$ determines the *hourly evolution* of the electricity spot price. Suppose right now, at time t , it is 3pm on April 12. If we want to forecast what the electricity price will be at 2am and 4pm on July 26, there is nothing much we can say other than there is an 80% probability that the 4pm price is greater than the 2am price, for example. On the other hand, if we want to forecast what the electricity price will be at time $t + 1$, 4pm on April 12, our forecast will depend on the probability

distribution of Y_{t+1} , which does not depend on the time of the day. In other words, p_{t+1} depends a lot more on what p_t and $\tilde{\mu}_t - p_t$ is at time t than the fact that $t + 1$ corresponds to 4pm.

4 One-Step Prediction Model for Electricity Spot Price

Finally, we present our mathematical model describing the electricity price process. The purpose of our model is to describe the statistical characteristic of electricity price at time $t + 1$, given the information we have up to time t . First, recall that $\tilde{\mu}_Y = \text{med}(Y_t)$, and let

$$\hat{\varepsilon}_t = \frac{Y_t}{\tilde{\mu}_Y}, \quad \forall t.$$

Given \mathcal{F}_t , we assume our price process follows

$$p_{t+1} = p_t + (1 - \kappa) (\tilde{\mu}_t - p_t) \hat{\varepsilon}_{t+1}, \quad (2)$$

where $\kappa := 1 - \tilde{\mu}_Y$ determines the force of median-reversion. Then,

$$\text{median}(p_{t+1} - \tilde{\mu}_t) = \kappa (p_t - \tilde{\mu}_t),$$

given \mathcal{F}_t . When we compute κ from the market data, we get $0 < \kappa < 1$, implying that the electricity price process is median-reverting. Given \mathcal{F}_t , we know that $\text{median}(\hat{\varepsilon}_{t+1}) = 1$ by definition, and we can assume $\mathbb{E}[\hat{\varepsilon}_{t+1}^2] = \infty$ based on the tail-index that we have computed in the previous section. From (2), if $p_t = \tilde{\mu}_t$, the process will stall and just stay at $\tilde{\mu}_t$. However, we can assume $\mathbb{P}[p_t = \tilde{\mu}_t] = 0$ for all t since p_t is a continuous random variable.

Recall that $\tilde{\mu}_t$ is a slowly moving median computed from the past one month of hourly data and it is locally a martingale. If we assume $\tilde{\mu}_t \approx \tilde{\mu}_{t+1}$,

$$\begin{aligned} p_{t+2} &= p_{t+1} + (1 - \kappa) (\tilde{\mu}_{t+1} - p_{t+1}) \hat{\varepsilon}_{t+2} \\ &= p_t + (1 - \kappa) (\tilde{\mu}_t - p_t) \hat{\varepsilon}_{t+1} \\ &\quad + (1 - \kappa) (\tilde{\mu}_{t+1} - p_t - (1 - \kappa) (\tilde{\mu}_t - p_t) \hat{\varepsilon}_{t+1}) \hat{\varepsilon}_{t+2} \\ &= p_t + (1 - \kappa) (\tilde{\mu}_t - p_t) \hat{\varepsilon}_{t+1} \\ &\quad + (1 - \kappa) (\tilde{\mu}_{t+1} - p_t) \hat{\varepsilon}_{t+2} - (1 - \kappa)^2 (\tilde{\mu}_t - p_t) \hat{\varepsilon}_{t+1} \hat{\varepsilon}_{t+2} \\ &\approx p_t + (1 - \kappa) (\tilde{\mu}_t - p_t) (\hat{\varepsilon}_{t+1} + \hat{\varepsilon}_{t+2}) - (1 - \kappa)^2 (\tilde{\mu}_t - p_t) \hat{\varepsilon}_{t+1} \hat{\varepsilon}_{t+2}. \end{aligned}$$

Next, even though the market data shows that there is a slight positive dependence between $\hat{\varepsilon}_{t+1}$ and $\hat{\varepsilon}_{t+2}$, suppose they are i.i.d for a moment to study its implications. Then, given \mathcal{F}_t ,

$$\begin{aligned}
\text{median}(p_{t+2}) &\approx p_t + 2(1 - \kappa)(\tilde{\mu}_t - p_t) - (1 - \kappa)^2(\tilde{\mu}_t - p_t) \\
&= p_t + (1 + \kappa)(1 - \kappa)(\tilde{\mu}_t - p_t) \\
&= p_t + (1 - \kappa^2)(\tilde{\mu}_t - p_t).
\end{aligned}$$

We can generalize the result and show that if $\tilde{\mu}_t \approx \tilde{\mu}_{t+1} \approx \dots \approx \tilde{\mu}_{t+n}$ and $\hat{\varepsilon}_{t+1}, \hat{\varepsilon}_{t+2}, \dots, \hat{\varepsilon}_{t+n}$ are i.i.d, then

$$\text{median}(p_{t+n}) = p_t + (1 - \kappa^n)(\tilde{\mu}_t - p_t),$$

given \mathcal{F}_t . Thus, at time t , we know that p_{t+n} is reverting towards the local median $\tilde{\mu}_t$ for some small n , while it will be reverting towards the long-term median μ for large n .

However, while this model is simple and describes the statistical characteristic of hour-ahead price very well as will be shown in section §4-2, we caution the reader against relying on the model for long-term projections. The trailing median process is approximately a martingale, but it is not a perfect martingale, and while the serial dependence on the residuals are low as shown in section §3-1, they are still not perfectly zero, and of course they are not independent. Thus, if we project far out into the future, the errors arising from assuming the residuals are i.i.d Cauchy and assuming the TMP is a martingale, while individually small, will be compounded. The model is developed in order to help understand the short-term fluctuations and help us make hourly decisions to buy and sell electricity. For example, if it is 2pm on January 15, we may rely on the above model to gauge the statistical characteristic of electricity spot price on 5pm on January 15. However, it will be not be very useful in projecting what the electricity spot price will be on March 15, 4am, for example.

4.1 Qualitative Analysis

If $p_t \leq \tilde{\mu}_t$, then

$$\begin{cases} p_{t+1} \leq p_t, & \text{if } \hat{\varepsilon}_{t+1} \leq 0, \\ p_t < p_{t+1} \leq \tilde{\mu}_t, & \text{if } 0 < \hat{\varepsilon}_{t+1} \leq \frac{1}{1-\kappa}, \\ \tilde{\mu}_t < p_{t+1}, & \text{if } \frac{1}{1-\kappa} < \hat{\varepsilon}_{t+1}. \end{cases}$$

Similarly, if $p_t > \tilde{\mu}_t$, then

$$\begin{cases} p_{t+1} \geq p_t, & \text{if } \hat{\varepsilon}_{t+1} \leq 0, \\ p_t > p_{t+1} \geq \tilde{\mu}_t, & \text{if } 0 < \hat{\varepsilon}_{t+1} \leq \frac{1}{1-\kappa}, \\ \tilde{\mu}_t > p_{t+1}, & \text{if } \frac{1}{1-\kappa} < \hat{\varepsilon}_{t+1}. \end{cases}$$

Therefore, $\{\hat{\varepsilon}_{t+1} \leq 0\}$ corresponds to the worst event in which the price deviates further away from the median, and $\{0 < \hat{\varepsilon}_{t+1} \leq \frac{1}{1-\kappa}\}$ corresponds to the best event in which the price is reverting back toward the median. $\{\frac{1}{1-\kappa} < \hat{\varepsilon}_{t+1}\}$ corresponds to the over-reversion in which the price overshoots and moves past the median while reverting. As will be shown in the next section, the probability distribution of $\hat{\varepsilon}_{t+1}$ is smooth and regular for the case where $\{\hat{\varepsilon}_{t+1} > 0\}$. It is also heavy-tailed. It is smooth and regular because reversion towards the median is a natural, predictive phenomenon in the market. It is heavy-tailed because no matter how far the price has deviated from the median, the market can abruptly correct itself and return to normalcy in a very short period of time. It also implies that the market is prone to over-reversion. The probability distribution of $\hat{\varepsilon}_{t+1}$ for the case where $\{\hat{\varepsilon}_{t+1} \leq 0\}$ is somewhat difficult to characterize even when $\hat{\varepsilon}_{t+1}$ is relatively close to 0. It is not only the big deviations that causes the price spikes that are irregular; the small deviations are not smooth, as well. Overall, there is a fundamental asymmetry and skewness in the hourly electricity price data. Reversion toward the median is smooth and strong all the way up to the heavy-tailed end corresponding to the over-reversion. On the other hand, movement in the opposite direction - moving further away from the median - is unnatural and unpredictable even for relatively small deviations.

4.2 Cauchy Approximation

We found that the distribution of $\hat{\varepsilon}_t$ is well-described by the Cauchy distribution, especially for the positive outcomes. If a random variable Y has a *Cauchy* $(1, \gamma)$ distribution, then

$$\alpha = F_Y(y) = \mathbb{P}[Y \leq y] = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} \left(\frac{y-1}{\gamma} \right), \quad (3)$$

and

$$y = F_Y^{-1}(\alpha) = 1 + \gamma \tan \left[\pi \left(\alpha - \frac{1}{2} \right) \right].$$

The Cauchy distribution ensures that the median-reversion of the price process is smooth and heavy-tailed. **Figure 12** shows the Q-Q plots constructed from Texas electricity price including all the data from the .01-quantile to the .99-quantile of the set $(\hat{\varepsilon}_t)_{t \geq 1}$ versus a set constructed by randomly generated i.i.d. *Cauchy* $(1, \gamma)$ samples, where 1 is the median and γ is the scale parameter computed from the spread between the 90th percentile and the 50th percentile of the data. That is,

$$\gamma := \frac{\{90^{th}\text{-quantile of the set } (\hat{\varepsilon}_t)_{t \geq 1}\} - 1}{\tan \left[\pi \left(.9 - \frac{1}{2} \right) \right]}.$$

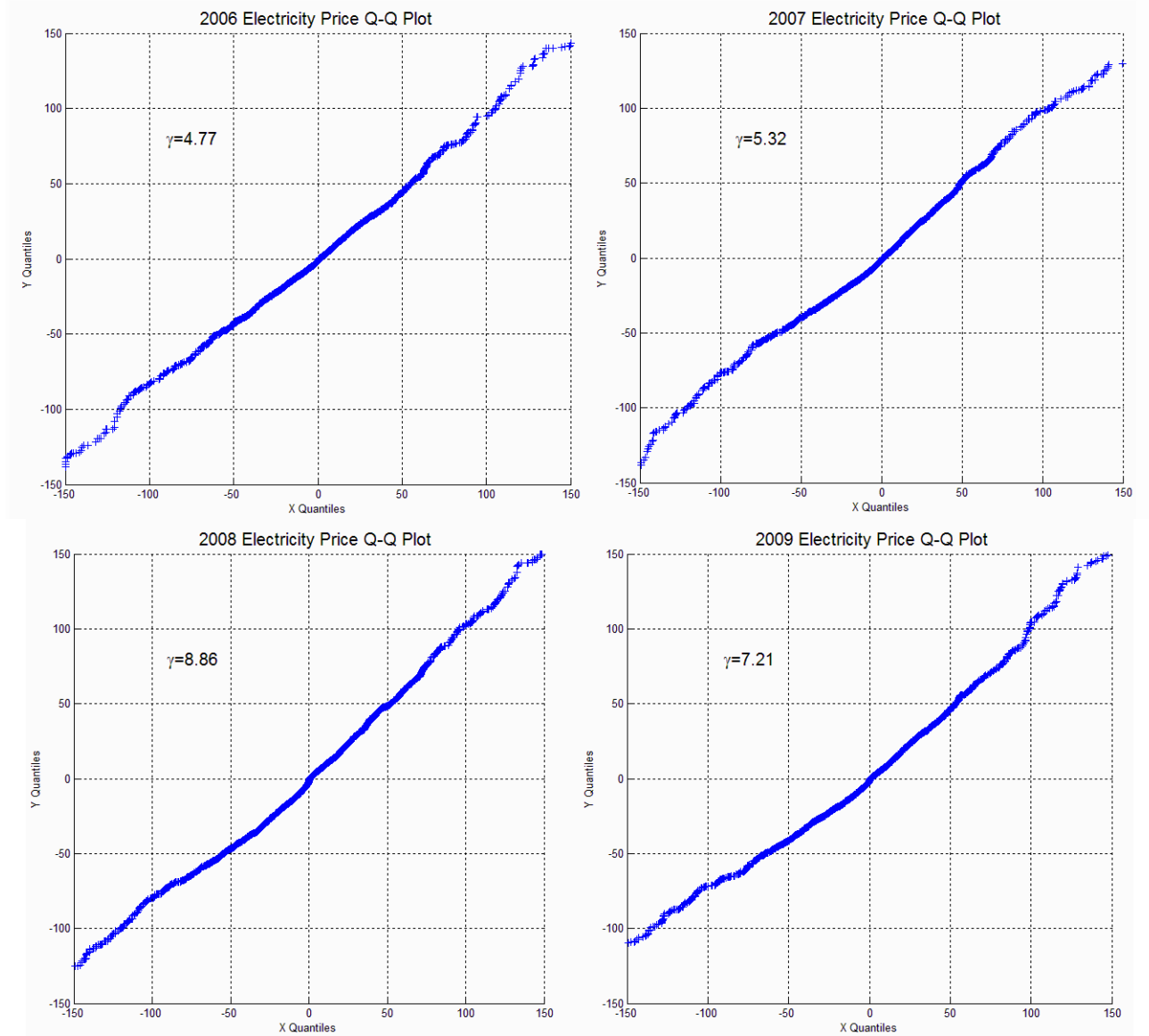


Fig. 12. QQ plot using Texas electricity spot market price from 2006 to 2009.

The axis on the figures range from -150 to 150. That is because the .01-quantile is around -150 and the .99-quantile is around 150. .05 and .95 quantiles are around -35 and 35, respectively. .1 and .9-quantiles are around -15 and 15, respectively. There are subtle deviations in the tail ends of the Q-Q plot, and one can notice that the slope of the plot in the negative side deviates slightly from one, showing that the real data is slightly skewed compared to the Cauchy distribution. Nonetheless, the overall straight lines imply that the Cauchy distribution is a very good fit, especially on the right hand side. If $(\hat{\epsilon}_{t'})_{t-(T-2) \leq t' \leq t}$ were truly Cauchy, the mean would be indeterminate, but we are assuming that the mean is 0. From the Q-Q plots, we can argue that $(\hat{\epsilon}_{t'})_{t-(T-2) \leq t' \leq t}$ have a slightly skewed and truncated Cauchy distribution.

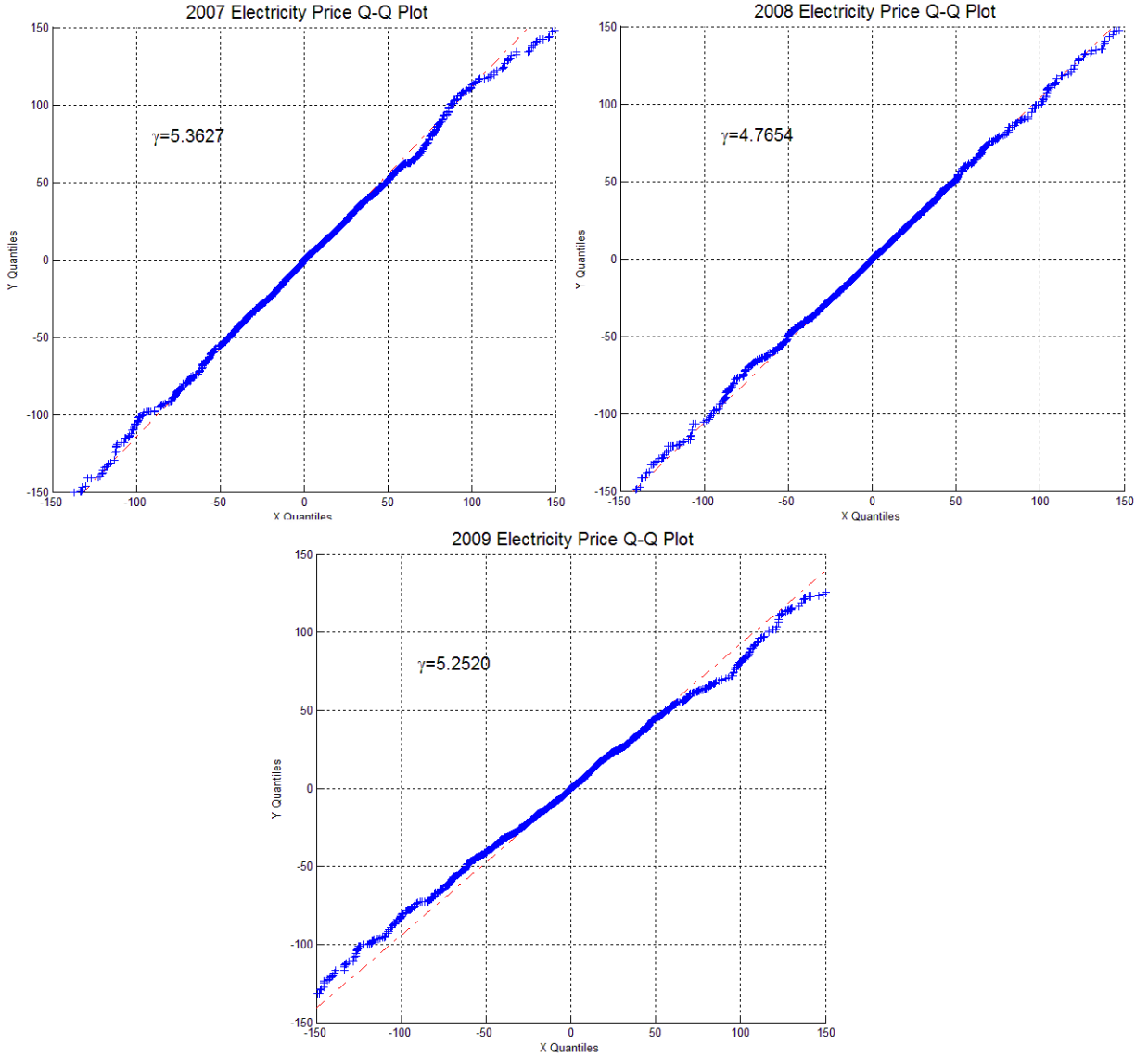


Fig. 13. QQ plot using PJM West electricity spot market price from 2007 to 2009.

Figure 13 shows the Q-Q plots constructed from the PJM West hub electricity price data. **Table 6** shows that the quantiles of the Texas data $(\hat{\varepsilon}_{t'})_{t-(T-2)\leq t'\leq t}$ are close to that of (3) on the right hand side. The .5-quantile and the .9-quantiles are matched by construction. We compare the real data and the data generated from i.i.d Cauchy random variables from the .45-quantile to the .95-quantile (the right hand side). The results are similar for the PJM West hub data. Therefore, at time t given \mathcal{F}_t , our prediction for p_{t+1} is that it is approximately Cauchy distributed with median $\tilde{\mu}_t + \kappa(p_t - \tilde{\mu}_t)$ and scale factor γ .

α -quantile	.45	.5	.55	.6	.65	.7	.75	.8	.85	.9	.95
2006 Data	.358	1	1.767	2.61	3.65	4.86	6.31	8.11	10.42	15.7	34.2
Cauchy(1, 4.77)	.244	1	1.756	2.56	3.43	4.56	5.87	7.66	10.36	15.7	32.1
2007 Data	.243	1	1.86	2.85	3.86	5.12	6.80	8.96	11.53	17.4	35.6
Cauchy(1, 5.32)	.156	1	1.84	2.73	3.71	4.87	6.32	8.33	11.46	17.4	34.6
2008 Data	.073	1	2.136	3.51	5.37	7.56	10.23	14.2	19.5	28.3	57.9
Cauchy(1, 8.86)	-.404	1	2.404	3.88	5.51	7.43	9.86	13.2	18.4	28.3	57.0
2009 Data	.189	1	2.17	3.47	4.85	6.68	9.06	12.2	16.1	23.2	48.7
Cauchy(1, 7.21)	-.143	1	2.14	3.34	4.67	6.24	8.21	10.9	15.2	23.2	46.5

Table 6

Empirical evidence for the Cauchy distribution

5 Conclusion

We have analyzed the behavior of electricity prices in Ercot Texas and PJM West hub and showed that it is median-reverting and heavy-tailed. We presented empirical evidence to show that it is difficult to quantify seasonality using higher-order statistics in an heavy-tailed environment. While we can assert that the probability that the 4pm price is greater than the 2am price is 80%, we cannot assess on average how much greater the 4pm price will be. The results shown in this paper imply that a good policy for managing and trading electricity should depend on the empirical CDF and the quantile function rather than the empirical mean and the variance used in traditional linear models.

6 Acknowledgements

We would like to acknowledge the careful comments by Curt Lefebvre.

References

- [1] Bouchaud, J.P. 2001. Power laws in economics and finance: some ideas from physics. *Quantitative Finance*. **1**(1) 105-112.
- [2] Barton, J.P., Infield, D.G. 2004. Energy storage and its use with intermittent renewable energy. *IEEE Trans. Energy Conversion*. **19**(2). 441-448.

- [3] Breymann, W. Dias, A., Embrechts, P. 2003. Dependence structure for multivariate high-frequency data in finance. *Quantitative Finance*. **3**(1) 1-14.
- [4] Brunetto, C., Tina, G. 2007. Optimal hydrogen storage sizing for wind power plants in day ahead electricity market. *Renewable Power Generation*. **1**(4) 220–226.
- [5] Bystrom, H.N.E. 2002. Extreme value theory and extremely large electricity price changes. *International Review of Economics & Finance*. **14**(1) 41-55.
- [6] Cartea, A., Figueroa, M.G. 2005. Pricing in electricity markets: a mean-reverting jump-diffusion model with seasonality. *Applied Mathematical Finance*. **12**(4) 313-335.
- [7] Cont, R. 2001. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*. **1**(2) 223-236.
- [8] Danielsson, J., de Haan, L., Peng, L., de Vries, C.G. 2001. Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis*. **76**(1) 226-248.
- [9] Delgado, M.A. 1996. Testing serial dependence using the sample distribution function. *Journal of Time Series Analysis*. **17**(3) 271-285.
- [10] Dress, H., de Haan, L., Resnick, S. 2000. How to make a Hill plot. *Annals of Statistics*. **28**(1) 254-274.
- [11] Drost, F.C., Nijman, T.E., Werker, B.J.M. 1998. Estimation and testing in models containing both jumps and conditional heteroscedasticity. *Journal of Business and Economic Statistics*. **16**(2) 237-243.
- [12] Dufour, J.-M. 1981. Rank tests for serial dependence. *Journal of Time Series Analysis*. **2**(3) 117-128.
- [13] Engle, R.F., Patton, A.J. 2001. What good is a volatility model?. *Quantitative Finance*. **1**(2) 237-245.
- [14] Eraker, B. 2004. Do stock prices and volatility jump? reconciling evidence from spot and option prices. *The Journal of Finance*. **59**(3) 1367-1403.
- [15] Eydeland, A., Wolyniec, K. 2003. *Energy and Power Risk Management*. John Wiley & Sons, New Jersey.
- [16] Farmer, J.D., Gillemot, L., Lillio, F., Mike, S., Sen, A. 2004. What really causes large price changes?. *Quantitative Finance*. **4**(4) 383-397.
- [17] Farmer, J.D., Lillio, F. 2004. On the origin of power law tails in price fluctuations. *Quantitative Finance*. **4**(1) C7-C11.
- [18] Graves, F. and Jenkin, T. and Murphy, D. 1999. Opportunities for electricity storage in deregulating markets. *The Electricity Journal*. **12**(8) 46-56.
- [19] Hill, B.M. 1975. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*. **3**(5) 1163-1174.

- [20] Huisman, R., Maiheu, R. 2003. Regime jumps in electricity prices. *Energy Economics*. **25**(5) 425-434.
- [21] Hsing, T. 1991. On tail index estimation using dependent data. *Annals of Statistics*. **19**(3) 1547-1569.
- [22] Ibrahim, H., Ilinca, A., Perron, J. 2008. Energy storage systems-characteristics and comparisons. *Renewable and Sustainable Energy Reviews*. **12**(5) 1221-1250.
- [23] Knittel, C.R., Roberts, M.R. 2005. An empirical examination of restructured electricity prices. *Energy Economics*. **27**(5) 791-817.
- [24] LeBaron, B. 2001. Stochastic volatility as a simple generator of apparent financial power laws and long memory. *Quantitative Finance*. **1**(6) 621-631.
- [25] Lucia, J.J., Schwartz, E.S. 2002. Electricity prices and power derivatives: Evidence from the Nordic power exchange. *Review of Derivatives Research*. **5**(1) 5-20.
- [26] MacKerron, G., Pearson, P. 2000. *International Energy Experience: Markets, Regulation and the Environment*. Imperial College Press, London.
- [27] Nogales, F. J., Contreras, J., Conejo, A. J., Espinola, R. 2002. Forecasting next day electricity prices by times series models. *IEEE Trans. Power Systems*. **17**(2) 342-348.
- [28] Peng, L. 2001. Estimating the mean of a heavy-tailed distribution. *Statistics and Probability Letters*. **52**(3) 255-264.
- [29] Peng, L. 2004. Empirical-likelihood based confidence interval for the mean with a heavy-tailed distribution. *Annals of Statistics*. **32**(3) 1192-1214.
- [30] Peng, L., Qi, Y. 2006. Confidence regions for high quantiles of a heavy-tailed distribution. *Annals of Statistics*. **34**(4) 1964-1986.
- [31] Pilipovic, D. 2007. *Energy Risk: Valuing and Managing Energy Derivatives*. McGraw-Hill, New York.
- [32] Puterman, M. L. 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley and Sons, New York.
- [33] Ait-Sahalia, Y. 2004. Disentangling diffusion from jumps. *Journal of Financial Economics*. **74**(3) 487-528.
- [34] Resnick, S., Starica, C. 1998. Tail index estimation for dependent data. *Annals of Applied Probability*. **8**(4) 1156-1183.
- [35] Schwartz, E.S. 1997. Stochastic behavior of commodity prices: Implications for valuation and hedging. *The Journal of Finance*. **52**(3) 923-974.
- [36] Sioshansi, R., Denholm, P., Jenkin, T., Weiss, J. 2009. Estimating the value of electricity storage in PJM: Arbitrage and some welfare effects. *Energy Economics*. **31**(2) 269-277.

- [37] Taleb, N.N. 2007. Black swans and the domains of statistics. *The American Statistician*. **61**(3) 198-200.
- [38] Taleb, N. N. 2009. Errors, robustness, and the fourth quadrant. *International Journal of Forecasting*. **25**(4) 744-759.
- [39] Wilcoxon, F., Katti, S.K., Wilcox, R.A. 1973. Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed-rank test. *Selected Tables in Mathematical Statistics I*. 171-259.