PRACE NAUKOWE AKADEMII EKONOMICZNEJ WE WROCŁAWIU
Nr 1176 — 2007
Inwestycje finansowe i ubezpieczenia
– tendencje światowe a polski rynek

Rafał Weron
Hugo Steinhaus Center, Politechnika Wrocławska
Adam Misiorek
Institute of Power Systems Automation, Wrocław

HEAVY TAILS AND ELECTRICITY PRICES: DO TIME SERIES MODELS WITH NON-GAUSSIAN NOISE FORECAST BETTER THAN THEIR GAUSSIAN COUNTERPARTS?

1. Introduction

In the last decades, with deregulation of power markets and introduction of competition, electricity price forecasts have become a fundamental input to an energy company's decision-making mechanism [10; 14]. Short-term price forecasts (STPF) are of particular interest for participants of auction-type spot electricity markets who are requested to express their bids in terms of prices and quantities. In such markets buy (sell) orders are accepted in order of increasing (decreasing) prices until total demand (supply) is met. Consequently, a generator that is able to forecast spot prices can adjust its own production schedule accordingly and hence maximize its profits.

It has been long known that financial asset returns are not normally distributed. Rather, the empirical observations exhibit excess kurtosis [3; 11; 12]. This heavy-tailed (also called fat-tailed or leptokurtic) character of the distribution of price changes has been repeatedly observed in various financial and commodity markets. The pertinent questions are whether electricity prices are also heavy-tailed, what probability distributions best describe the empirical data and whether models with heavy-tailed innovations perform better in terms of forecasting accuracy than their Gaussian counterparts.

This paper is a continuation of our earlier studies on STPF of California electricity prices with time series models [9; 14; 15]. Here we focus on the above raised questions. In fact, only on the latter as the answer to the first question is pretty straightforward and the second has been already addressed in [2; 8; 13; 14] (and we build on these results). Consequently, we limit the range of analyzed models to autoregressive time series approaches that have been found to perform well for pre-crash California power market data. We expand them by allowing for heavy-tailed innovations in the form of α-stable or generalized hyperbolic noise.

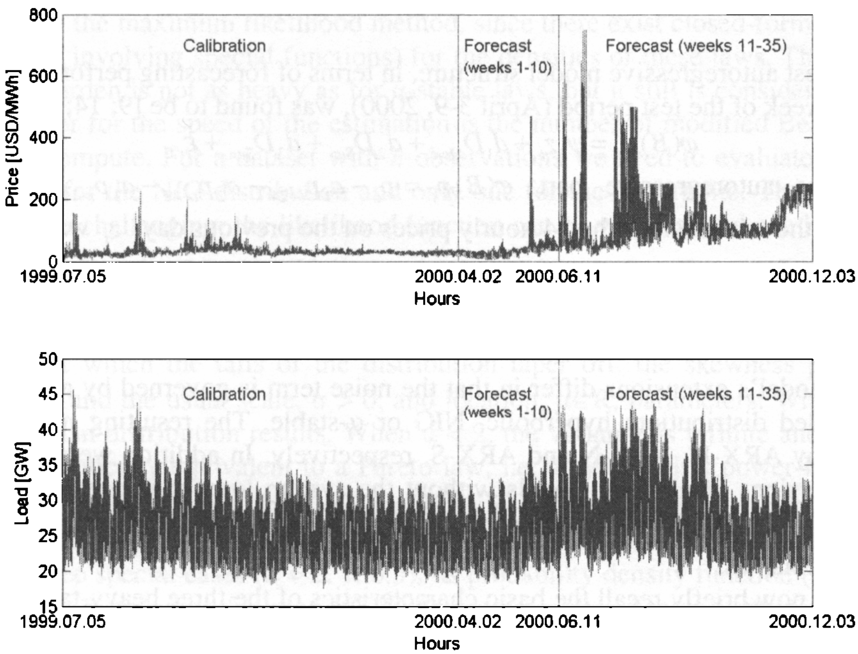

Fig. 1. Hourly system prices (*top panel*) and hourly system loads (*bottom panel*) in California for the period July 5, 1999-December 3, 2000. The changing price cap (750 → 500 → 250 USD/MWh) is clearly visible in the top panel. The day-ahead load forecasts (i.e. the official forecasts of the system operator CAISO) are indistinguishable from the actual loads at this resolution; only the latter have been plotted
Source: own calculatins.

Like in the previous papers, an assumption is made that only publicly available information is used to predict spot prices, i.e. generation constraints, line capacity limits or other fundamental variables are not considered. The available dataset[1] includes hourly system prices, system-wide loads, and day-ahead load forecasts for the California market. The time series used in this study are depicted in Figure 1.

[1] The dataset *CA_hourly.dat* is part of the MFE Toolbox, see [14].

The data from the period July 5, 1999-April 2, 2000 was used for calibration and from the period April 3-December 3, 2000 for out-of-sample testing. Since in practice the market-clearing price forecasts for a given day are required on the day before, we used the following testing scheme. To compute price forecasts for hour 1 to 24 of a given day, data available to all procedures included price and demand historical data up to hour 24 of the previous day plus day-ahead load predictions for the 24 hours of that day.

2. The base model and its extensions

The best autoregressive model structure, in terms of forecasting performance for the first week of the test period (April 3-9, 2000), was found to be [9; 14; 15]:

$$\varphi(B)p_t = \psi_1 z_t + d_1 D_{Mon} + d_2 D_{Sat} + d_3 D_{Sun} + \varepsilon_t,$$

where the autoregressive part $\varphi(B)p_t = p_t - a_1 p_{t-24} - a_2 p_{t-48} - a_3 p_{t-168} - a_4 mp_t$, mp_t was the minimum of the 24 hourly prices on the previous day, z_t was the load forecast and $D_{Mon}, D_{Sat}, D_{Sun}$ were the dummy variables (for Monday, Saturday and Sunday). In this base model, denoted in the text as ARX, the noise term ε_t is i.i.d. Gaussian.

The model's extensions differ in that the noise term is governed by a different, heavy-tailed distribution: hyperbolic, NIG or α-stable. The resulting models are denoted by ARX-H, ARX-N and ARX-S, respectively. In addition we study simplified versions of all four models without the system load component, i.e. with $\psi_1 = 0$. The letter 'X', which stands for 'eXogenous variable', is dropped from the respective names.

Let us now briefly recall the basic characteristics of the three heavy-tailed families (for a more thorough treatment see [12]). The generalized hyperbolic distribution is defined as a normal variance-mean mixture where the mixing distribution is the generalized inverse Gaussian law with parameter λ, i.e. it is conditionally Gaussian. The hyperbolic and NIG (normal inverse Gaussian) laws are special cases obtained for λ = 1 and λ = –0.5, respectively. The PDF of the hyperbolic H(α, β, δ, μ) law can be written as:

$$f_H(x) = \frac{\sqrt{\alpha^2 - \beta^2}}{2\alpha\delta \mathbf{K}_1\left(\delta\sqrt{\alpha^2 - \beta^2}\right)} \mathrm{e}^{-\alpha\sqrt{\delta^2 + (x-\mu)^2} + \beta(x-\mu)},$$

and of the NIG(α, β, δ, μ) distribution as:

$$f_{NIG}(x) = \frac{\alpha\delta \mathbf{K}_1\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right)}{\pi\sqrt{\delta^2 + (x-\mu)^2}} \mathrm{e}^{\delta\sqrt{\alpha^2 - \beta^2} + \beta(x-\mu)},$$

where $\delta > 0$ and $\mu \in R$ are the usual scale and location parameters, while α and β determine the shape, with α being responsible for the steepness and β, $|\beta| < \alpha$, for the skewness. The normalizing constant $K_1(t)$ is the modified Bessel function of the third kind with index 1. The tail behavior is often classified as 'semi-heavy', i.e. the tails are lighter than those of non-Gaussian stable laws, but much heavier than Gaussian. It is characterized by the following asymptotic relation: $f(x) \sim |x|^{\lambda-1} \exp((\mp\alpha + \beta)x)$. In particular, the hyperbolic log-density forms a hyperbola – hence the name of the distribution [1].

The parameter estimation of generalized hyperbolic distributions can be performed by the maximum likelihood method, since there exist closed-form formulas (although, involving special functions) for the densities of these laws. The computational burden is not as heavy as for α-stable laws, but it still is considerable. The main factor for the speed of the estimation is the number of modified Bessel functions to compute. For a dataset with n observations we need to evaluate n Bessel functions for the NIG distribution and only one for the hyperbolic. The optimization is also challenging: the likelihood function can be very flat and can have local minima.

Stable laws – also called α-stable, stable Paretian or Lévy stable – require four parameters for complete description: the tail exponent $\alpha \in (0,2]$, which determines the rate at which the tails of the distribution taper off, the skewness parameter $\beta \in [-1, 1]$ and the usual scale, $\sigma > 0$, and location, $\mu \in R$, parameters. When $\alpha = 2$, the Gaussian distribution results. When $\alpha < 2$, the variance is infinite and the tails are asymptotically equivalent to a Pareto law, i.e. they exhibit a power-law decay of order $x^{-\alpha}$. In contrast, for $\alpha = 2$ the decay is exponential. From a practitioner's point of view the crucial drawback of the stable distribution is that, with the exception of three special cases ($\alpha = 2, 1, 0.5$), its probability density function (PDF) and cumulative distribution function (CDF) do not have closed form expressions. They have to be evaluated numerically, see [12] for details, either by approximating complicated integral formulas or by taking the Fourier transform of the characteristic function $\varphi(t)$:

$$\log\varphi(t) = \begin{cases} -\sigma^\alpha |t|^\alpha \left\{1 + i\beta \operatorname{sign}(t) \tan\frac{\pi\alpha}{2}\left[\left(\sigma|t|\right)^{1-\alpha} - 1\right]\right\} + i\mu t, & \alpha \neq 1, \\ -\sigma|t|\left\{1 + i\beta \operatorname{sign}(t)\frac{2}{\pi}\log\left(\sigma|t|\right)\right\} + i\mu t, & \alpha = 1. \end{cases}$$

The estimation of stable law parameters is in general severely hampered by the lack of known closed–form PDFs. Numerical approximation or direct numerical integration are nontrivial and burdensome from a computational point of view. As a consequence, the maximum likelihood (ML) estimation algorithm based on such approximations is difficult to implement and time consuming for samples encountered in practice. Yet, the ML estimates are almost always the most accurate, followed by regression-type estimates and quantile methods.

3. Empirical results

To assess the prediction performance of the models, different statistical measures can be utilized. The most widely used measures are those based on absolute errors, i.e. absolute values of differences between the actual, P_h, and predicted, $\hat{P}_h$, prices for a given hour, h. The Mean Absolute Percentage Error (MAPE) is a typical example. However, when applied to electricity prices, MAPE values could be misleading. In particular, when electricity prices drop to zero, MAPE values become very large regardless of the actual absolute differences $|P_h - \hat{P}_h|$. The reason for this is the normalization by the current (close to zero, and hence very small) price P_h. Alternative normalizations have been proposed in the literature. For instance, the absolute error $|P_h - \hat{P}_h|$ can be normalized by the average price attained during the day: $\bar{P}_{24} = \frac{1}{24}\sum_{h=1}^{24} P_h$. The resulting measure, also known as the Mean Daily Error [5; 14], is given by:

$$\text{MDE} = \frac{1}{24}\sum_{h=1}^{24}\frac{|P_h - \hat{P}_h|}{\bar{P}_{24}}.$$

The forecast accuracy was checked afterwards, once the true market prices were available. The error statistics for the whole test period (April 3-December 3, 2000) and all models – separately for models with and without the exogenous variable – are given in Table 1. Furthermore, to distinguish the rather calm first 10 weeks of the test period from the more volatile weeks 11-35 (see Fig. 1), the summary statistics are displayed separately for the two periods. These statistics are based on the 245 Mean Daily Errors. In particular, the number of days a given model was best in terms of MDE, the number of times a given heavy-tailed model was better than its Gaussian counterpart in terms of MDE, the mean and standard deviation of MDEs, and the mean deviation from the best model. The latter statistics is defined as $\frac{1}{T}\sum_{t=1}^{T}\left(\text{MDE}_{i,t} - \text{MDE}_{\text{Best model},\,t}\right)$, where i ranges over all evaluated models (i.e. i = 4) and T is the number of days (70, 175) in the sample.

All computations were performed in Matlab 7.0. The AR(X) models were calibrated using the *armax.m* function, which minimizes the Final Prediction Error criterion [7]. The heavy-tailed models were calibrated by numerically maximizing the likelihood function with the AR(X) models' parameters as starting points of the unconstrained simplex search routine (*fminsearch.m* function). Obviously this approach requires large computational times, as the PDFs have to be evaluated many times.

The obtained results are somewhat surprising. In both periods and both categories (with/without the exogenous variable) most often the Gaussian model yielded the best point forecasts. And this picture is not blurred by the 'large' number of its heavy-tailed competitors – generally they performed inferior rather than superior compared to AR(X). The only exceptions are the AR-H model in the volatile period (88 out of 175 days better than AR) and the ARX-H model in the calm period (38 out of 70 days better than ARX). The picture is a bit more favorable to the heavy-tailed models if we look at the other statistics. In the calm period, the heavy-tailed models not only yielded lower on average and less dispersed MDEs, but also gave lower mean deviation from the best model for a given week. In other words, the heavy-tailed models were closer to the 'optimal model' composed of the best performing model in each week. In particular, the AR-N, ARX-N and ARX-S specifications performed particularly well. However, in the volatile period the AR(X) models were again the best (except for the standard deviation of MDEs).

Table 1. Error measures for the considered models.
Best results in each category are emphasized in bold

	AR	**AR-H**	**AR-N**	**AR-S**
	Weeks 1-10 (relatively calm period)			
Times best	**35**	11	10	14
Times better than AR (max. 70)	–	27	21	**29**
Mean MDE	12.57	12.58	**12.30**	12.41
Standard deviation of MDE	13.18	12.35	**11.59**	11.79
Mean deviation from the best	1.09	1.10	**0.82**	0.93
	Weeks 11-35 (volatile, atypical period)			
Times best	**68**	31	35	41
Times better than AR (max. 175)	–	**88**	72	81
Mean MDE	**18.24**	18.29	18.54	18.47
Standard deviation of MDE	23.55	21.92	**19.12**	19.71
Mean deviation from the best	**1.39**	1.44	1.69	1.63
	ARX	**ARX-H**	**ARX-N**	**ARX-S**
	Weeks 1-10 (relatively calm period)			
Times best	**27**	13	20	10
Times better than AR (max. 70)	–	**38**	31	32
Mean MDE	11.98	11.94	11.64	**11.63**
Standard deviation of MDE	12.71	12.06	**11.61**	11.66
Mean deviation from the best	1.20	1.16	0.86	**0.85**
	Weeks 11-35 (volatile, atypical period)			
Times best	**71**	26	41	37
Times better than AR (max. 175)	–	**82**	77	77
Mean MDE	**17.71**	18.01	18.10	18.15
Standard deviation of MDE	21.22	21.24	**17.13**	17.59
Mean deviation from the best	**1.80**	2.11	2.20	2.25

Source: own calculations.

Table 2. Mean percent of exceedances of the 50, 90 and 99% two-sided day-ahead confidence intervals (CI) by the actual system price for the considered models

	AR			AR-H			AR-N			AR-S		
Weeks	50%	90%	99%	50%	90%	99%	50%	90%	99%	50%	90%	99%
1-10	42.62	**14.05**	6.01	**55.60**	14.46	4.17	62.62	15.06	3.27	61.43	16.31	**0.60**
11-35	**43.90**	**13.52**	5.74	56.38	15.88	4.45	64.71	17.60	1.95	62.83	18.45	**0.55**
	ARX			**ARX-H**			**ARX-N**			**ARX-S**		
Weeks	50%	90%	99%	50%	90%	99%	50%	90%	99%	50%	90%	99%
1-10	41.96	**13.93**	5.60	**53.69**	14.52	4.35	60.95	14.70	3.39	58.57	15.48	**0.77**
11-35	**46.10**	**13.60**	5.52	58.86	16.74	4.26	65.02	18.40	2.05	65.38	19.31	**0.60**

Source: own calculations.

Apart from point forecasts, we investigated the ability of the models to provide interval forecasts. For all considered models interval forecasts were determined analytically; for details on calculation of conditional prediction error variance and interval forecasts we refer to [6; 14]. Afterwards, following [4] and [9], we evaluated the quality of the interval forecasts by comparing the nominal coverage of the models to the true coverage. Thus, for each of the models we calculated confidence intervals (CIs) and determined the actual percentage of exceedances of the 50%, 90% and 99% two sided day-ahead CIs of the models by the actual system price, see Table 2. If the model implied interval forecasts were accurate then the percentage of exceedances should be approximately 50%, 10% and 1%, respectively. Note that in the calm period (first 10 weeks) 1680 hourly values were determined and compared to the system price for each of the models, while in the volatile period (weeks 11-35) – 4200 hourly values.

Examining the exceedances of the 50% interval we note that while the Gaussian models yield too wide CIs, all of the heavy-tailed alternatives behave quite the opposite. In this respect they exhibit a performance similar to the Markov regime--switching model analyzed in [9]. Also, the AR(X)-H model is better than the NIG and α-stable competitors, and comparable to AR(X). Looking at the exceedances of the 90% interval we see all models performing alike and yielding too narrow CIs. Yet, the AR(X) CIs are slightly better (wider) than those of the other models. Finally, the exceedances of the 99% interval present a different picture. The α-stable innovations lead to the widest (even a bit too wide) and closest to the optimal CIs. Next in line are the NIG, hyperbolic and Gaussian models, all of which yield too narrow CIs. In this category, ARX-H and ARX-N models behave comparably to the nonlinear Threshold ARX (TARX) model analyzed in [9].

4. Conclusions

In this paper we investigated the forecasting power of time series models for electricity spot prices. Motivated by the good fit of various heavy-tailed distributions to electricity price returns we focused on comparing linear autoregressive models with Gaussian and heavy-tailed innovations (hyperbolic, NIG and α-stable). The models were tested on a time series of hourly system prices and loads from California. We evaluated the quality of the predictions both in terms of the Mean Daily Error (for point forecasts) and in terms of the nominal coverage of the models to the true coverage (for interval predictions).

The is no unanimous winner of the presented competition. During relatively calm weeks the AR-N, ARX-N and ARX-S models led to the best 'on average' point forecasts, but could not beat the AR(X) models in the number of best forecasts. Surprisingly, in the volatile period the AR(X) models yielded the best point forecasting performance. Regarding interval forecasts the evidence is also mixed. Gaussian models behave well for the 50% and 90% intervals, but are worse for the 99% CI than the rest. Overall the NIG models seem to be reasonable heavy-tailed alternatives to AR(X), but the performance does not fully justify the computational burden involved.

References

[1] Barndorff-Nielsen O.E., *Exponentially Decreasing Distributions for the Logarithm of Particle Size*, Proc. Royal Soc. London A 1977 (353), 401-419.

[2] Bottazzi G., Sapio S., Secchi A., *Some Statistical Investigations on the Nature and Dynamics of Electricity Prices*, „Physica A" 2005, 355, 54-61.

[3] Carr P., Geman H., Madan D.B., Yor M., *The Fine Structure of Asset Returns: An Empirical Investigation*, „Journal of Business" 2002, 75, 305-332.

[4] Christoffersen P., Diebold F.X., *How Relevant is Volatility Forecasting for Financial Risk Management*, „Review of Economics and Statistics" 2000, 82, 12-22.

[5] Conejo A.J., Contreras J., Espinola R., Plazas M.A., *Forecasting Electricity Prices for a Day-Ahead Pool-Based Electric Energy Market*, „International Journal of Forecasting" 2005, 21(3), 435-462.

[6] Hamilton J., *Time Series Analysis*, Princeton University Press, 1994.

[7] Ljung L., *System Identification – Theory for the User*, Prentice Hall, 1999.

[8] Misiorek A., *Ciężkie ogony cen, czyli dlaczego trudno mierzyć ryzyko na rynku energii*, Materiały konferencyjne REE'06, Kazimierz Dolny, 24-27 kwietnia, 2006.

[9] Misiorek A., Trück S., Weron R., *Point and Interval Forecasting of Spot Electricity Prices: Linear vs. Non-Linear Time Series Models*, Stud. Nonlinear Dynamics & Econometrics, 2006, 10(3), Art. 2.

[10] *Modelling Prices in Competitive Electricity Markets*, red. D.W. Bunn, Wiley 2004.

[11] Rachev S., Mittnik S., *Stable Paretian Models in Finance*, Wiley 2000.

[12] Weron R., *Computationally Intensive Value at Risk Calculations*, [w:] *Handbook of Computational Statistics: Concepts and Methods*, red. J.E. Gentle, W. Härdle, Y. Mori, Springer 2004, 911-950, http://www.quantlet.com/mdstat/ scripts/stf/html/stfhtml.html.
[13] Weron R., *Heavy Tails and Electricity Prices*, Invited paper presented at the Deutsche Bundesbank's 2005, Annual Fall Conference, Eltville, 10-12 November 2005.
[14] Weron R., *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*, Wiley, 2006, http://www.im.pwr.wroc.pl/~rweron/MFE.html.
[15] Weron R., Misiorek A., *Short-Term Electricity Price Forecasting with Time Series Models: A Review and Evaluation*, [w:] *Complex Electricity Markets*, red. W. Mielczarski, Chapter 10, 2006, 231-254.

CIĘŻKIE OGONY A CENY ENERGII ELEKTRYCZNEJ: CZY MODELE SZEREGÓW CZASOWYCH Z SZUMEM NIEGAUSSOWSKIM PROWADZĄ DO LEPSZYCH PROGNOZ NIŻ MODELE GAUSSOWSKIE?

Streszczenie

Residua modeli szeregów czasowych wykorzystywanych do prognoz procesów energetycznych, m.in. cen na giełdach energii elektrycznej, nie mają rozkładu gaussowskiego, lecz charakteryzują się znacznie cięższymi ogonami. Jednak, w literaturze naukowej wykorzystywano dotąd metody zakładające właśnie gaussowski rozkład innowacji. Niniejsza praca ma na celu odpowiedzieć na pytanie, jaki wpływ na dopasowanie modeli oraz na jakość prognoz ma zastosowanie modeli z szumem ciężkoogonowym (hiperbolicznym, NIG bądź α-stabilnym).

Wyniki analiz przeprowadzonych na danych kalifornijskich nie są jednoznaczne. Okazuje się, że modele z szumem NIG oraz α-stabilnym prowadzą do średnio dokładniejszych prognoz, ale modele gaussowskie częściej zwracają najlepsze wyniki.