International Conference
"The European Electricity Market EEM-05"
May 10-12, 2005, Lodz, Poland

Proceedings Volume, pp. 133-141

FORECASTING SPOT ELECTRICITY PRICES WITH TIME SERIES MODELS

Rafał WERON
Hugo Steinhaus Center
for Stochastic Methods
Wrocław University of Technology
Wrocław (Poland)

Adam MISIOREK[♣]
Institute of Power Systems Automation
Wroclaw (Poland)

Abstract

In this paper we study simple time series models and assess their forecasting performance. In particular we calibrate ARMA and ARMAX (where the exogenous variable is the system load) processes. Models are tested on a time series of California power market system prices and loads from the period proceeding and including the market crash.

1. INTRODUCTION

Short-term load forecasting plays an important role in power system operation and planning. Accurate load prediction saves costs by improving economic load dispatching, unit commitment, etc. At the same time it enhances the function of security control. Hence, it has become one of the major fields of research in electrical engineering [1, 7, 8, 17, 23].

In the last decades, with the introduction of competition and deregulation of the power markets, a new challenge has appeared. It turned out that even accurate load forecasts cannot guarantee profits. The market risk related to trading is considerable due to extreme volatility of electricity prices. This is especially true for spot prices, where the volatility can be as high as 50% on the daily scale, i.e. over ten times higher than for other energy products (natural gas and crude oil). Extreme volatility forces producers and wholesale consumers to hedge not only against volume risk but also against price movements. But to hedge efficiently it is imperative to thoroughly study and, afterwards, accurately model electricity price dynamics. The obtained forecasts will in turn help develop bidding strategies or negotiation skills in order to maximize profits.

However, the prediction of electricity prices with the accuracy currently attainable for loads is a difficult, if not an impossible task. Firstly, because the price process exhibits seasonality – at the daily, weekly and annual timescales, which is not as pronounced as for loads. Secondly, because there are many quantifiable exogenous variables that may be considered, with loads and network constraints being the obvious examples. Thirdly, because there are also psycho- and sociological factors that can cause an unexpected and irrational buyout of certain contracts leading to price spikes.

Despite these obstacles, in the recent years various techniques have been applied to electricity price modelling. They may be classified as:

- parsimonious stochastic models,
- structural or fundamental models,
- non-parametric models.

Many of the approaches considered in the literature are in fact hybrid solutions. In particular, it is difficult to distinguish between the first two classes as many time series or regression models incorporate one or two fundamental factors, like loads or fuel prices. The same holds for the hybrid stochastic models that we study in this paper. Despite their relative simplicity they offer reasonably accurate price forecasts, at least for the calm and moderately volatile periods.

The paper is structured as follows. In Section 2 we review the literature and the modelling approaches. We start with the stochastic models, then turn to hybrid solutions, which incorporate fundamental factors, and conclude with short comments on the two other model classes. In Section 3 we describe the dataset and introduce our models. In the following section we present the calibration results and compare the forecasts with those of other authors.

2. MODELLING APPROACHES

Many stochastic models are inspired by the financial literature and a desire to adapt some of the well known and widely applied in practice approaches. Modifications involve addition of certain electricity price characteristics, like price spikes or mean-reversion. Earliest examples include jump-diffusion [11] and its extension – mean-reverting jump-diffusion of Johnson and Barz [10]. The former is based on Merton's model of discontinuous asset prices [14], while the latter additionally exhibits a Vasicek-type mean-reverting mechanism originally proposed for interest rate dynamics [20]. A serious flaw of the Johnson and Barz model is the slow speed of mean reversion after a jump. Later this has been circumvented by adding downward jumps, allowing time-varying parameters or incorporating non-linearities in the price dynamics, such as regime-switching and stochastic volatility [2, 9, 22].

A common feature of the finance-inspired stochastic models is their main intention to replicate the statistical properties of spot prices with the ultimate objective of derivatives evaluation. Although in this context the models' simplicity and analytical tractability are an advantage, in forecasting the former feature is a serious limitation, while the latter is an excessive luxury.

Interestingly, the continuous-time stochastic models have familiar counterparts in discrete time. For instance, Vasicek-type mean reversion is equivalent to an autoregressive AR(1) process. It does not take much to make the next step and utilize the classical engineering approach of modelling a random phenomenon (here: electricity price P_t) via time series, like ARMA:

$$A(p)P_t \equiv P_t - a_1 P_{t-1} - \ldots - a_p P_{t-p} = \\ = \varepsilon_t + b_1 \varepsilon_{t-1} + \ldots + b_q \varepsilon_{t-q} \equiv B(q)\varepsilon_t,$$

or the more general ARIMA and SARIMA. Application of the latter allowed to obtain mean daily forecast errors of 5-10% in the non-spiky periods for data from mainland Spain and Californian markets [4].

The introduction of exogenous variables in these models – in the form of fundamental factors, like loads or plant data – is straightforward. Such a combined approach allowed Nogales et al. [16] to reduce the mean daily forecast errors to 3-7% for the same data.

In the case of ARMA processes, the inclusion of exogenous variables leads to the so called ARMAX models [13]:

$$A(p)P_t = C(r,k)Z_t + B(q)\varepsilon_t,$$

where

$$C(r,k)Z_t \equiv Z_{t-k} + c_1 Z_{t-k-1} + \ldots + c_r Z_{t-k-r},$$

Z_t is the value of the exogenous variable (e.g. system load) at time t, and ε_t – like above – is independent and identically distributed noise with finite variance. In general, there can be several exogenous variables. Additionally, seasonal constraints can be imposed or, like for the stochastic models, a regime-switching mechanism can be introduced [5].

Since spot electricity prices display excessive volatility, which is time-varying with evidence of heteroscedasticity both in unconditional and conditional variance [6], models that incorporate GARCH effects have been also applied. Karakatsani and Bunn [12] tested four approaches (including regression-GARCH, and Time-Varying Parameter regression with exogenous variables) to explain the stochastic dynamics of spot volatility and understand agent reactions to shocks. Limitations of GARCH models due to extreme values were resolved when a regression model with the assumptions of an implicit jump component for prices and a leptokurtic distribution for innovations was used. A similar approach was taken by Mugele et al. [15] who applied GARCH time series with α-stable innovations for modelling the asymmetric and heavy-tailed nature of electricity spot price returns.

Although the hybrid approaches involve some fundamental factors, the “classical” structural models intend to uncover a richer structure for electricity prices in order to better understand the complex market performance. For example, based on stochastic climate factors (temperature and precipitation), Vahviläinen and Pyykkönen [19] modelled hydrological inflow and snow-pack development that affect hydro power generation, the major source of electricity in Scandinavia. Utilizing 31 deterministic and stochastic formulas their model was able to capture the observed fundamentally motivated market price movements in the medium-term horizon.

Taking into account that forecasting is perceived by the pragmatic market players as more important than derivatives valuation and/or risk management it is no wonder that, due to their universality, artificial intelligence methods have been also adopted for price prediction. For example, Wang and Ramsay [21] proposed a hybrid approach based on neural networks and fuzzy logic, with examples from the England-Wales market and daily mean errors around 10%. While Szkuta et al. [18] proposed a three-layered neural network with backpropagation, showing results from the Victorian electricity market, with daily mean errors around 15%. However, like in the case of load forecasting, artificial intelligence techniques are not intuitive and no simple physical interpretation may be attached to their components. Hence, they do not allow engineers and system operators to better understand the power market’s behaviour.

3. DATA AND MODELS

Like in [4] and [16], we forecast CalPX market clearing prices from the period proceeding and including the Californian market crash. This lets us evaluate the performance of the models during normal (calm) periods, as well as during highly volatile periods.

The time series of system prices, system-wide loads, and day-ahead load forecasts was constructed using data obtained from the UCEI institute (www.ucei.berkeley.edu) and the California’s independent system operator CAISO (oasis.caiso.com). The missing and “doubled” data values corresponding to the changes to and from the daylight saving time (summer time) were treated in the usual way. The former were substituted by the arithmetic average of the two neighboring values, while the latter by the arithmetic average of the two values for the “doubled” hour. Likewise, the few outliers (but not the spikes) were substituted by the arithmetic average of the two neighboring values. The obtained time series are depicted in Figs. 1 and 2. The day-ahead load forecasts (i.e. the official forecasts of the system operator CAISO) are indistinguishable from the loads at this resolution.

We used the data from the period July 5, 1999 – April 2, 2000 solely for the purpose of calibration. Such a relatively long period of data was needed to achieve high accuracy. For example, limiting the calibration period to data coming only from the year 2000 (like in [4, 16]) led to a decrease in forecasting performance by up to 70%.

Consequently, the period April 3 – December 3, 2000 was used 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.

The models considered in this study comprised simple time series specifications with and without exogenous variables, namely

ARMAX and ARMA processes. The calibration was performed in Matlab (prediction error estimate) and SAS (maximum likelihood and conditional least squares estimates) computing environments. The logarithmic transformation was applied to price, $p_t = \log(P_t)$, and load, $z_t = \log(Z_t)$, data to attain a more stable variance. The mean (or the median for loads) was also removed to center the data around 0. Furthermore, since each hour displays a rather distinct price profile reflecting the daily variation of demand, costs, and operational constraints the modeling was implemented separately across the hours, leading to 24 sets of parameters. This approach was also inspired by the extensive research on demand forecasting, which has generally favored the multi-model specification for short-term predictions [1, 7].

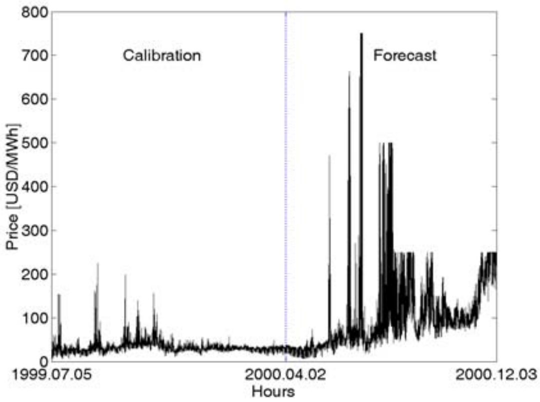

FIGURE 1. Hourly system prices in California for the period July 5, 1999 – December 3, 2000. The changing price cap (750 → 500 → 250 USD/MWh) is clearly visible.

Seasonal market conditions were captured by the autoregressive structure of the models: the log price p_t was made dependent on the log prices for the same hour on the previous days, and the previous weeks, as well as a certain function (e.g. mean, minimum) of all prices on the previous day. The latter created the desired link between bidding and price signals from the entire day.

Since the system load partly explains the price behavior (especially on the daily scale) it was used as the fundamental variable, see Fig. 3. In the calm period (till mid-May) the dependence between the log price and the log system load is almost linear with a slight downward bend for small values of the load. Later the prices tend to jump during high load hours, leading to an S-shaped curvilinear dependence. This observation suggests using non-linear regression for the spiky periods. Testing the effectiveness of such an approach is left for future research.

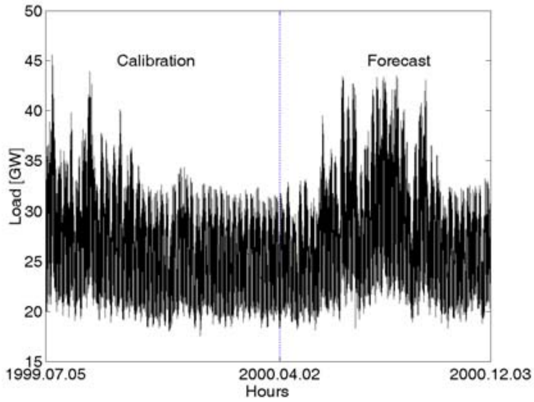

FIGURE 2. Hourly system loads in California for the period July 5, 1999 – December 3, 2000.

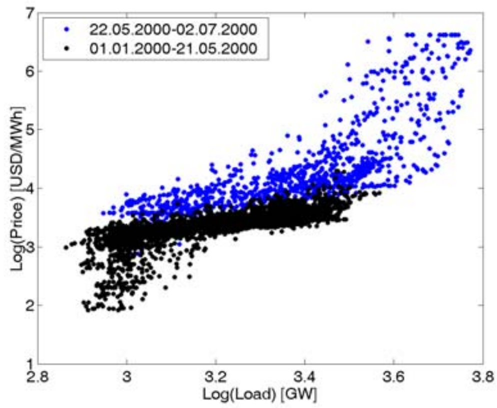

FIGURE 3. The dependence between the hourly log prices and hourly log system loads in California for the period January 1 – July 2, 2000.

4. FORECASTING PERFORMANCE

Collecting all the facts together, for each hour our model has the following general form:

$$A(p)p_t = C(r,0)z_t + B(q)\varepsilon_t,$$

where $C(r,0)z_t$ is such that at lag 0 the CAISO day-ahead load forecast for a given hour is used, while for larger lags the actual system load is used. Interestingly, using the actual load at lag 0, in general, does not improve the forecast. This phenomenon might be explained by the fact that the prices are an outcome of the bids, which in turn are placed without the knowledge of the future actual system load.

Furthermore, a large moving average part $B(q)\varepsilon_t$ typically decreases the performance, despite the fact that in many cases it is suggested by Akaike's Final Prediction-Error (FPE) criterion [13]. The best results were obtained for pure ARX models, i.e. with $B(q)\varepsilon_t = \varepsilon_t$. Likewise, a large autoregression part generally led to overfitting and worse out-of-sample forecasts. The optimal structure was found to be of the form:

$$A(p)p_t \equiv p_t - a_1 p_{t-24} - a_2 p_{t-48} - \\ - a_3 p_{t-168} - a_4 mp_t,$$

where mp_t is a function of all prices on the previous day. The best results were obtained for the minimum of the 24 hourly prices.

This very simple structure was unable to cope with the weekly seasonality, the results for Mondays, Saturdays, and Sundays were significantly worse than for the other days. Separate modeling of each hour of the week (leading to 168 ARX models) was not satisfactory, probably due to a much smaller calibration set. Incorporation of 7 dummy variables (one for each day of the week) did not improve the results significantly. However, inclusion of 3 dummy variables (for Monday, Saturday, and Sunday) helped a lot.

The best model structure, in terms of forecasting performance for the first week of the test period (April 3-9, 2000), turned out to be:

$$p_t - a_1 p_{t-24} - a_2 p_{t-48} - a_3 p_{t-168} - a_4 mp_t = \\ = c_1 z_t + d_{Mon} + d_{Sat} + d_{Sun} + \varepsilon_t,$$

where $d_{Mon}, d_{Sat}, d_{Sun}$ denote the coefficients of the dummies. Note, that all models were estimated using an adaptive scheme, i.e. instead of using a single model for the whole sample, for every day (and hour) in the test period we calibrated the model (given its structure) to the previous values of prices and loads and obtained a forecasted value for that day (and hour). Originally, at each time step also the model structure was optimized by minimizing the FPE criterion [13] for a given set of model structures. However, this procedure, apart from being time consuming, did not produce satisfactory results. The models were apparently overfitted. Hence, we decided to use only one model structure for all hours and all days.

To assess the prediction performance of the models, different statistical measures were utilized. They were chosen to comply with those used in [3, 4, 16] for the results to be comparable. The forecast accuracy was checked afterwards, once the true market prices were available. For all the weeks under study, two types of average prediction errors were computed: one corresponding to the 24 hours of each day and one to the 168 hours of each week. The Mean Daily Error (i.e. a variant of the Mean Absolute Percentage Error) was computed as:

$$MDE = \frac{1}{24}\sum_{h=1}^{24} \frac{|p_h - p_h^{est}|}{\overline{p}_{24}},$$

where $\overline{p}_{24}$ is the mean price for a given day (to avoid the adverse effect of prices close to zero) and p_h^{est} is the predicted price for a given hour. Analogously to MDE, the Mean Weekly Error was computed as:

$$MWE = \frac{1}{168}\sum_{h=1}^{168} \frac{|p_h - p_h^{est}|}{\overline{p}_{168}},$$

where $\overline{p}_{168}$ is the mean price for a given week.

Additionally, the weekly mean square errors were calculated as the square roots of the average of 168 square differences between the predicted and the actual prices:

Note, that the normalization by 1/168 was not used later in the text to comply with the results of [4] and [16]

$$WMSE = \sqrt{\cancel{\frac{1}{168}} \sum_{h=1}^{168} \left(p_h - p_h^{est} \right)^2} \, .$$

Mean daily errors for the first week of the test period (April 3-9, 2000) are given in Table 1, see also Fig. 4. ARX denotes our best model and AR denotes this model without the exogenous variable, i.e. for $c_1 \equiv 0$. Surprisingly the performance is not much worse, in fact the forecasts for Monday and Sunday are even better. For comparison we include the results for other models studied in the literature. DR denotes the dynamic regression model (equivalent to ARX) and TF denotes the transfer function model (equivalent to ARMAX); both studied in [16]. Furthermore, ARIMA stands for a seasonal ARIMA model and ARIMA-E for the same model but with an explanatory variable (system load); both tested in [4].

TABLE 1. Mean daily errors (MDE) in percent for the first week of the test period (April 3-9, 2000). Best results are emphasized in bold. Results not passing the naïve test are italicized.

	ARX	AR	Naïve	DR	TF	ARIMA
Mo	3.91	3.73	5.68	**2.6**	2.8	4.35
Tu	**2.33**	2.99	3.77	3.3	3.3	*6.17*
We	**2.06**	*2.28*	2.19	*2.7*	2.9	*2.60*
Th	**1.58**	1.94	2.97	1.9	2.2	2.53
Fr	*2.92*	*3.62*	2.89	2.5	**2.3**	*3.57*
Sa	3.98	5.39	8.72	3.7	**3.6**	8.46
Su	4.87	3.96	10.11	4.0	**3.7**	7.44

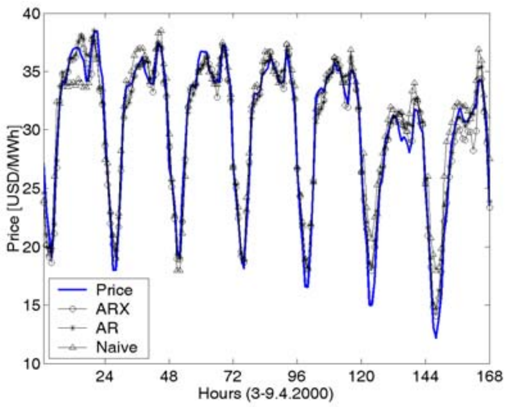

FIGURE 4. Prediction results for the first week of the test period (April 3-9, 2000).

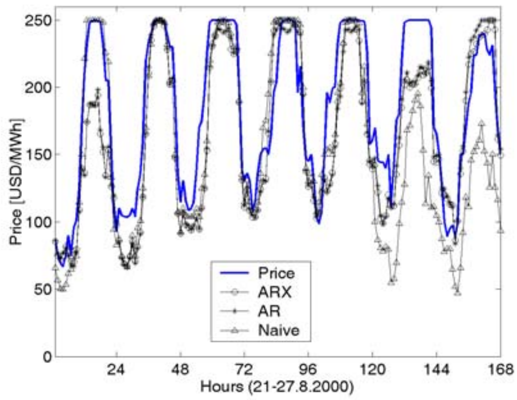

FIGURE 5. Prediction results for the 21st week of the test period (August 21-27, 2000).

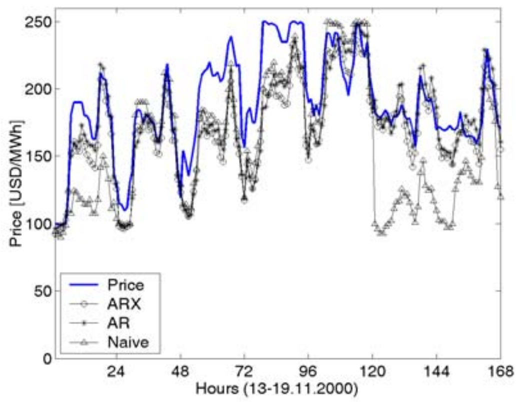

FIGURE 6. Prediction results for the 33rd week of the test period (November 13-19, 2000).

Following [3] a naïve but challenging test was used as a benchmark for all forecasting procedures. The forecasts were compared to the 24 prices of a day similar to the one to be forecast. A similar day is characterized as follows. A Monday is similar to the Monday of the previous week and the same rule applies for Saturdays and Sundays; analogously, a Tuesday is similar to the previous Monday, and the same rule applies for Wednesdays, Thursdays, and Fridays. The naïve test is passed if hourly errors for the estimates are smaller than for the prices of the similar day. More often than expected, forecasting procedures do not pass this test.

Mean weekly errors and weekly mean square errors for four selected weeks (1, 2, 21, and 33) of the test period are given in

Table 2, see also Figs. 5 and 6. Surprisingly, as was already noted in [4], the inclusion of the fundamental variable (system load) is not always optimal. While for the first 28 weeks of the test period ARX was better than or roughly the same as AR, the situation changed in favor of the latter in late 2000 when the minimum daily price increased above 70 USD/MWh. For the relatively calm periods a ca. 10% decrease in MWE was observed, however, during the spiky weeks the improvement was negligible. An even worse effect was obtained for the ARIMA-E model in [4]. It is also worth noting that our AR model is overall significantly better than ARIMA. The major difference between them is that for each day the former utilizes 24 hourly 7-parateter structures, whereas the latter one common 48-parameter specification for all hours.

TABLE 2. Mean weekly errors (MWE) in percent and weekly mean square errors (WMSE) for four selected weeks (1, 2, 21, and 33) of the test period. Best results are emphasized in bold. DR and TF values are known only for the first week (MWE: 2.95 and 2.97, WMSE: 13.61 and 13.47, respectively) [16].

	ARX	AR	Naïve	ARIMA	ARIMA-E
			MWE		
1	**3.04**	3.36	5.00	5.01	5.21
2	**4.71**	5.28	8.62	-	-
21	**13.9**	14.1	18.22	15.65	21.03
33	11.1	**10.6**	18.57	13.60	13.68
			WMSE		
1	**15.2**	16.6	26.8	21.19	21.82
2	**20.7**	22.7	38.0	-	-
21	424	**422**	585	470	675
33	348	**331**	546	393	397

It is very surprising that DR and TF – which also use one common multi-parameter specification for all hours – are significantly better than ARIMA and especially ARIMA-E. After all, TF and ARIMA-E are more or less equivalent in terms of variables used. Possibly this is related to the way the load data is included in both methods. In ARIMA-E it is just an explanatory variable, but in TF it is bundled with the autoregressive part of the model [24]. DR and TF are also slightly better than ARX, at least for the first week of the test period, see Table 2. If this relation is sustained for other periods (and other data sets) has yet to be tested.

5. REFERENCES

1. Bunn, D.: Forecasting loads and prices in competitive power markets, Proc. IEEE 88 (2) 163-169, 2000.
2. Bunn, D., Karakatsani, N.: Forecasting Electricity Prices, London Business School Working Paper, 2003.
3. Conejo, A.J., Contreras, J., Espínola, R., Plazas, M.A.: Forecasting electricity prices for a day-ahead pool-based electric energy market, International Journal of Forecasting, *in print*, 2005.
4. Contreras, J., Espínola, R., Nogales, F.J., Conejo, A.J.: ARIMA Models to Predict Next-Day Electricity Prices, IEEE Trans. Power Systems 18(3), 1014-1020, 2003.
5. Davison M., Anderson, L., Marcus, B., Anderson, K.: Development of a Hybrid Model for Electrical Power Spot Prices, IEEE Trans. Power Systems 17(2), 257-264, 2002.
6. Escribano, A., Pena J.I., Villaplana, P.: Modelling electricity prices: International evidence, Working Paper, Universidad Carlos III de Madrid, 2002.
7. Feinberg, E.A., Genethliou, D.: Load forecasting. In: Applied Mathematics for Restructured Electric Power Systems: Optimization, Control, and Computational Intelligence, Chow, J.H., Wu, F.F., Momoh, J.J. (ed.), Springer, 2005.
8. Hippert, H.S., Pedreira, C.E., Souza, R.C.: Neural networks for short-term load forecasting: a review and evaluation, IEEE Trans. Power Systems 16(1), 44–55, 2001.
9. Huisman, R., Mahieu, R.: Regime jumps in electricity prices, Energy Economics 25, 425-434, 2003.
10. Johnson, B., Barz, G.: Selecting stochastic processes for modelling electricity prices. In: Energy Modelling and the Management of Uncertainty, Risk Books, London, 1999.
11. Kaminski, V.: The challenge of pricing and risk managing electricity derivatives. In: The US Power Market, Barber, P. (ed.), Risk Books, London, 1997.
12. Karakatsani, N., Bunn, D.: Modelling stochastic volatility in high-frequency spot electricity

prices, London Business School Working Paper, 2004.
13. Ljung, L.: System Identification – Theory for the User, 2 ed., Prentice Hall, Upper Saddle River, 1999.
14. Merton R.C.: Option pricing when underlying stock returns are discontinuous, Journal of Financial Economics 3, 125-144, 1976.
15. Mugele, Ch., Rachev, S.T., Trück, S.: Stable modeling of different European power markets, Investment Management & Financial Innovations, *in print*, 2005.
16. Nogales, F.J., Contreras, J., Conejo, A.J., Espínola, R.: Forecasting next-day electricity prices by time series models, IEEE Trans. Power Systems 17, 342-348, 2002.
17. Nowicka-Zagrajek, J., Weron, R.: Modeling electricity loads in California: ARMA models with hyperbolic noise, Signal Processing 82(12), 1903-1915, 2002.
18. Szkuta, B.R., Sanabria, L.A., Dillon, T.S.: Electricity price short-term forecasting using artificial neural networks, IEEE Trans. Power Syst. 14, 851-857, 1999.
19. Vahviläinen, I., Pyykkönen, T.: Stochastic factor model for electricity spot price – the case of the Nordic market, Energy Economics, forthcoming, 2005.
20. Vasicek, O.: An equilibrium characterization of the term structure, Journal of Financial Economics 5, 177-188, 1977.
21. Wang, A.J., Ramsay, B.: A neural network based estimator for electricity spot-pricing with particular reference to weekend and public holidays, Neurocomputing 23, 47-57, 1998.
22. Weron, R., Bierbrauer, M., Trück, S.: Modeling electricity prices: jump diffusion and regime switching, Physica A 336, 39-48, 2004.
23. Weron, R., Misiorek, A.: Modeling and forecasting electricity loads: A comparison, International Conference EEM-04 proceedings, 135-142, 2004.
24. Private communication with Javier Contreras.

Dr. Rafał Weron
received his M.Sc. (1995) and Ph.D. (1999) degrees in applied mathematics from the Wroclaw University of Technology (WUT), Poland, where he currently is posting an Assistant Professor position. His research focuses on risk management in the power markets and computational statistics as applied to finance and insurance. He is the (co-)author of 3 books and over 60 research articles, book chapters, and conference papers. His professional experience includes design of the risk management module for Opole Power Plant (Elektrownia Opole S.A.), development of insurance strategies for Polish Power Grid Co. (PSE S.A.) and Hydro-storage Power Plants Co. (ESP S.A.), as well as implementation of yield curve calibration and option pricing software for LUKAS Bank S.A. (Credit Agricole Group). He has also been a consultant or executive teacher to a large number of banks and corporations.

Mailing address:
Rafał Weron
Hugo Steinhaus Center
Wroclaw University of Technology
Wyb. Wyspiańskiego 27, 50-370 Wrocław
POLAND
Phone: (+48) (71) 320-3530.
Fax: (+48) (71) 320-2654.
e-mail: rweron@im.pwr.wroc.pl

Adam Misiorek, M.Sc.
was born in 1975 in Kielce, Poland. He received the M.Sc. degree in applied mathematics from the Wroclaw University of Technology (WUT). Currently he is statistical analyst/programmer at Institute of Power Systems Automation (IASE) and a PhD student at the Institute of Mathematics at WUT. His areas of interest include forecasting and mathematical methods of risk management.

Mailing address:
Adam Misiorek
Institute of Power Systems Automation
1 Wystawowa Str., 51-618 Wrocław
POLAND
Phone: (+48) (71) 348-4221.
Fax: (+48) (71) 348-2183.
e-mail: misiorek@iase.wroc.pl

♣ Research partially supported by KBN grant no. 4 T10B 030 25.