International Conference
"The European Electricity Market EEM-06"
May 24-26, 2006, Warsaw, Poland

Proceedings Volume, pp. 305-312

INTERVAL FORECASTING OF SPOT ELECTRICITY PRICES

Adam MISIOREK

Institute of Power Systems Automation
Wroclaw (Poland)

Rafał WERON

Hugo Steinhaus Center
Wroclaw University of Technology
Wroclaw (Poland)

Abstract

In this paper we assess the short-term forecasting power of different time series models in the electricity spot market. However, instead of evaluating point predictions we concentrate on interval forecasts. The latter are specifically important for risk management purposes where one is more interested in predicting intervals for future price movements than simply point estimates.

We find evidence that non-linear regime-switching models outperform their linear counterparts and that an additional GARCH component significantly improves interval forecasts of linear time series models.

1. INTRODUCTION

Short-term price forecasting (STPF) is 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. Since the day-ahead spot market typically consists of 24 hourly auctions that take place simultaneously one day in advance, forecasting with lead times from a few hours to a few days is of prime importance in day-to-day market operations. It is also the topic of this study. However, instead of evaluating point predictions (for a recent review of point forecasting with time series models consult [20]) we concentrate on interval forecasts. The latter are specifically important for risk management purposes where one is more interested in predicting intervals for future price movements than simply point estimates.

Like in [14] and [20] we limit the range of analyzed models to linear and non-linear time series approaches. For descriptions of model classes used here we refer to these two papers. The main focus is on empirical comparison of the models' short-term interval forecasting performance during normal as well as extremely volatile periods. 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 California power market is chosen as the test ground for two reasons: (i) it offers freely accessible high quality electricity price and load data and (ii) exhibits variable market behavior leading to a market crash in winter 2000/2001. 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. A thorough description of the dataset can be found in the two above mentioned articles. The preprocessed, spreadsheet-ready ASCII format data is available from http://www.im.pwr.wroc.pl/~rweron/exchlink.html.

The paper is structured as follows. In Section 2 we briefly present our models and calibration details. Section 3 provides empirical forecasting results for the studied models. Section 4 concludes and makes suggestions for future work.

2. THE MODELS

For time series modeling, the mean price and the median load were removed to center the data around zero. Removing the mean load resulted in worse forecasts, perhaps, due to the very distinct and regular asymmetric weekly structure with the majority of values lying in the high-load region. 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 [4][10][17].

Short-term 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 (maximum, minimum, mean or median) of all prices on the previous day. The latter created the desired link between bidding and price signals from the entire day.

In our ARMAX-type models we used only one exogenous variable: the hourly values of the system-wide load. At lag 0 the CAISO day-ahead load forecast for a given hour was used, while for larger lags the actual system load was used. Interestingly, the best models turned out to be the ones with only lag 0 dependence. Using the actual load at lag 0, in general, did not improve the forecasts either. This phenomenon can be explained by the fact that the prices are an outcome of the bids, which in turn are placed with the knowledge of load forecasts but not actual future loads.

Furthermore, a large moving average part $\theta(B)\varepsilon_t$ typically decreased the performance, despite the fact that in many cases it was suggested by Akaike's Final Prediction-Error (FPE) criterion. The best results were obtained for pure ARX-type models, i.e. with $\theta(B)\varepsilon_t = \varepsilon_t$. Likewise, a large autoregression part (we tested models with lags up to four weeks) generally led to overfitting and worse out-of-sample forecasts. The optimal autoregressive structure, i.e. yielding the smallest forecast errors for the first week of the test period (April 3-9, 2000), was found to be of the form:

$$\phi(B)p_t = p_t - a_1 p_{t-24} - a_2 p_{t-48} - a_3 p_{t-168} - a_4 mp_t, \quad (1)$$

where mp_t was the minimum of the previous day's 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. Inclusion of 3 dummy variables (for Monday, Saturday and Sunday) helped a lot. The best ARX model structure, in terms of forecasting performance for the first week of the test period, turned out to be (denoted later in the text as **ARX**):

$$\phi(B)p_t = \psi_1 z_t + d_1 D_{Mon} + d_2 D_{Sat} + d_3 D_{Sun} + \varepsilon_t, \quad (2)$$

where $\phi(B)p_t$ is given by (1), ψ_1 is the coefficient of the load forecast z_t and d_1, d_2, d_3 denote the coefficients of the dummies $D_{Mon}, D_{Sat}, D_{Sun}$, respectively. Its simplified version without the exogenous variable (**AR**):

$$\phi(B)p_t = d_1 D_{Mon} + d_2 D_{Sat} + d_3 D_{Sun} + \varepsilon_t, \qquad (3)$$

also performed relatively well.

The residuals obtained from the fitted ARX and AR models seemed to exhibit a non-constant variance. Indeed, when tested with the Lagrange multiplier "ARCH" test statistics [9] the heteroscedastic effects were significant at the 5% level. This motivated us to calibrate **ARX-G** and **AR-G** models to the data; "**G**" stands here for GARCH(1,1). They differ from ARX and AR models in that the noise terms in eqns. (2) and (3), respectively, are not just iid(0, σ^2) but are given by:

$$h_t = \varepsilon_t \sigma_t, \text{ with } \sigma_t^2 = \alpha_0 + \alpha_1 h_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \qquad (4)$$

where ε_t is iid with zero mean and finite variance.

Because of the non-linear nature of electricity prices, we also calibrated regime-switching TAR-type models to the spot price time series. They are natural generalizations of the ARX and AR models defined above. Namely, the **TARX** model is given by

$$\begin{cases} \phi_1(B)p_t = \psi_{1,1} z_t + d_{1,1} D_{Mon} + \\ \qquad + d_{1,2} D_{Sat} + d_{1,3} D_{Sun} + \varepsilon_t & \text{when } v_t \geq T, \\ \phi_2(B)p_t = \psi_{2,1} z_t + d_{2,1} D_{Mon} + \\ \qquad + d_{2,2} D_{Sat} + d_{2,3} D_{Sun} + \varepsilon_t & \text{when } v_t < T, \end{cases} \qquad (5)$$

where v_t and T are the threshold variable and the threshold level, respectively. We have tried different threshold variables and threshold levels. The former included combinations of past prices and loads: daily maximum, minimum and mean, value 24 hours ago, latest available value (i.e. value for hour 24 on the previous day), differences between lagged hourly values (for lags of 24 and 168 hours) and differences between lagged daily means (for 1 day and 1 week lags). The threshold levels were either constant or variable (estimated for every hour in a multi-step optimization procedure with ten equally spaced starting points spanning the entire parameter space). The best results – in terms of forecast errors during the first week of the test period – were obtained for v_t equal to the price for hour 24 on the previous day and T estimated for every hour in a multi-step optimization procedure. However, the predictions for later weeks were very disappointing, for details see [20]. Much better results for the whole test period were obtained for v_t equal to the difference in mean prices for yesterday and eight days ago. Since the original optimization process was very slow and did not yield better predictions than a simpler setup where T was set arbitrarily to zero, we have chosen the simpler setup as the best **TARX** model. The **TAR** model was obtained for $\psi_{1,1} = \psi_{2,1} = 0$, i.e. when no exogenous variables were used, and the same threshold variable and threshold level.

3. FORECASTING RESULTS

We investigated the ability of the models to provide interval forecasts. While there is a variety of empirical studies on evaluating point forecasts of electricity spot prices [6][15][16][18][19], literature on interval forecasts is very sparse. To the best of our knowledge, the only account of interval forecasting in electricity markets can be found in [14]. Interval forecasts may be especially relevant for risk management purposes where one is more interested in predicting intervals for future price movements than simply point estimates.

For all considered models interval forecasts were determined analytically; for details on calculation of conditional prediction error variance and interval forecasts we refer to [1][3][11][12][13]. Afterwards, following [1] or [5], 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 (CI) and determined the actual percentage of exceedances of the 50%, 90% and 99% two sided day-ahead CI of the models by the actual market clearing price (MCP). If the model implied interval forecasts were accurate then the percentage of exceedances should be approximately 50%, 10% and 1%, respectively. Note that for each week, 168 hourly values were determined and compared to the actual MCP.

Figures 1-3 show the deviations of the point forecasts and 50%, 90% and 99% two-sided CI from the actual market clearing price (MCP). Results for six models (AR, ARX, AR-G, ARX-G, TAR and TARX) and for two weeks (April 3-9, 2000 and August 21-27, 2000) of the test period are displayed. Obviously, the width of the interval varies much more for the AR/ARX-G and TAR/TARX than for the linear time series. This is the case both for variations of interval lengths within certain hours or days and differences between calm and volatile periods.

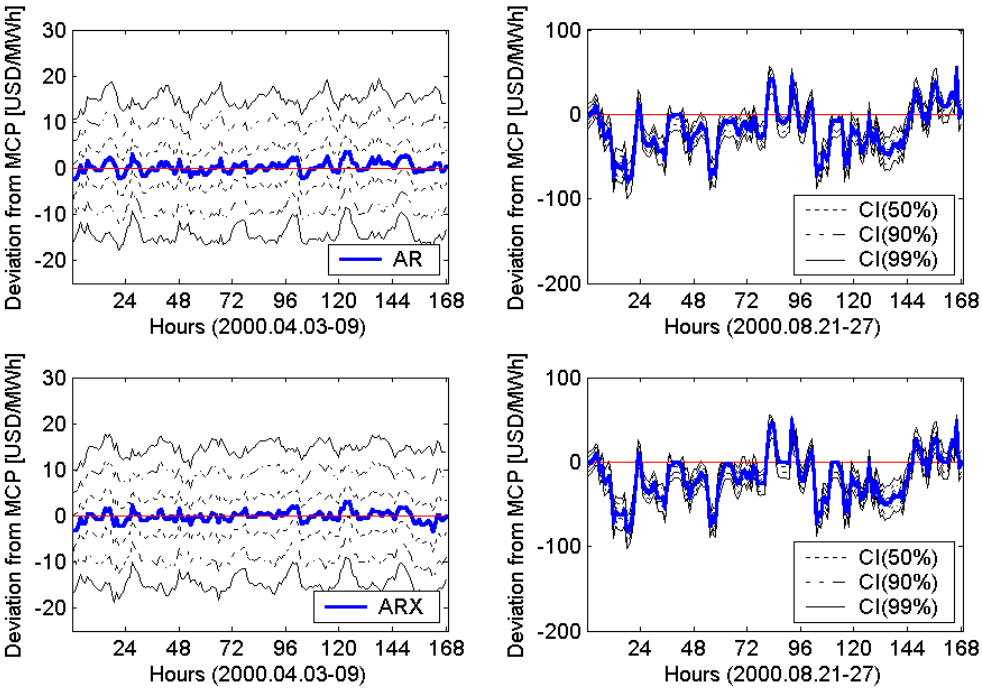

FIGURE 1 Deviation of the day-ahead point forecasts and their respective 50\%, 90\% and 99\% two-sided confidence intervals (CI) from the actual market clearing price (MCP) for two models: AR (*top panels*) and ARX (*bottom panels*), and for two weeks of the test period: the first week (April 3-9, 2000; *left panels*) and the 21st week (August 21-27, 2000; *right panels*).

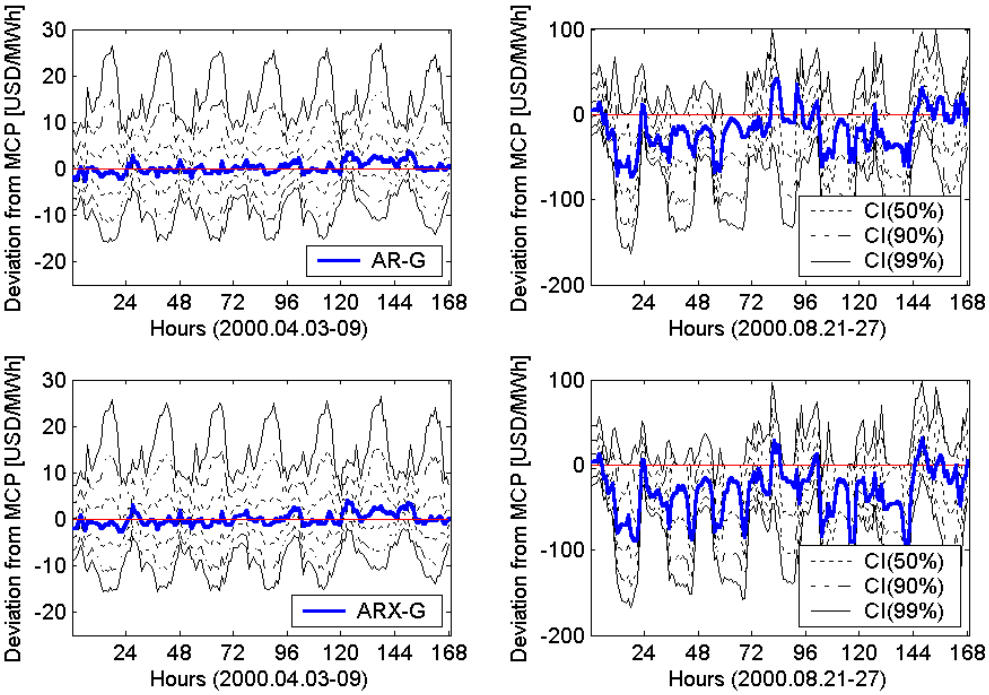

FIGURE 2 Deviation of the day-ahead point forecasts and their respective 50\%, 90\% and 99\% two-sided confidence intervals (CI) from the actual market clearing price (MCP) for two models: AR-G (*top panels*) and ARX-G (*bottom panels*), and for two weeks of the test period: the first week (April 3-9, 2000; *left panels*) and the 21st week (August 21-27, 2000; *right panels*).

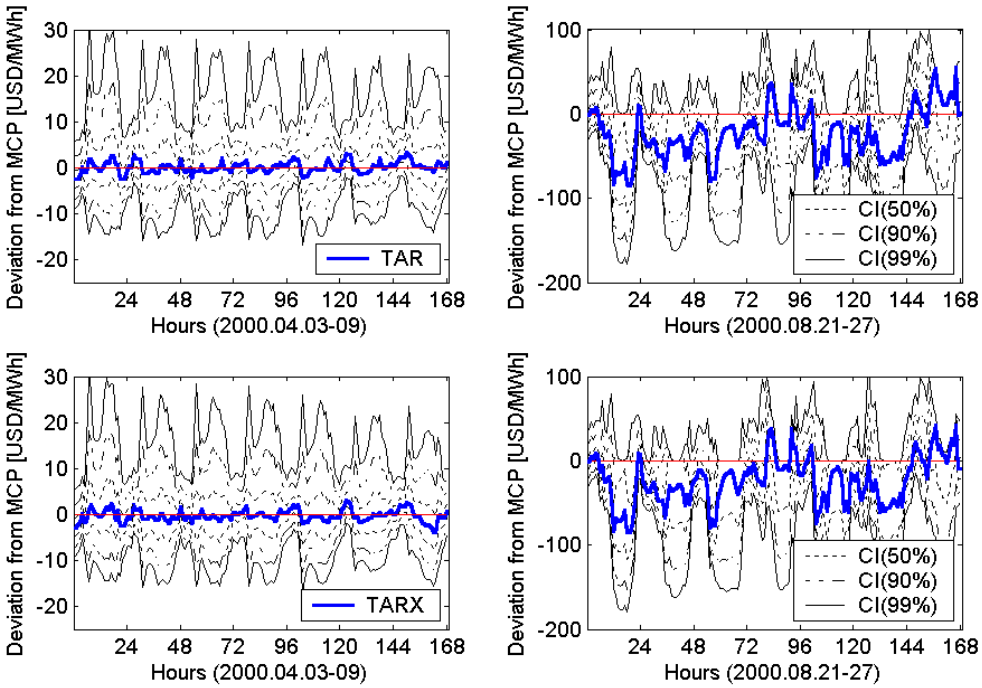

FIGURE 3 Deviation of the day-ahead point forecasts and their respective 50\%, 90\% and 99\% two-sided confidence intervals (CI) from the actual market clearing price (MCP) for two models: TAR (*top panels*) and TARX (*bottom panels*), and for two weeks of the test period: the first week (April 3-9, 2000; *left panels*) and the 21st week (August 21-27, 2000; *right panels*).

Examining the deviations of the CI from the actual MCP for the first week of the test period (left panels in Figures 1-3), we find that for all models almost all confidence intervals include the actual MCP. This is especially true for the 90% and 99% intervals, but even for the 50% confidence level deviations from the actual MCP are rarely high enough to exclude the price from the interval.

Note also that for the AR/ARX models there is only a small intra-day and intra-week variation of the intervals. For the AR/ARX-G and TAR/TARX models the width of the CI varies much more depending on the considered day and hour.

Examining results for the 21st week of the test period (right panels in Figures 1-3), we find that the estimated intervals of the linear models are clearly too narrow for the volatile period. Deviations of the point and interval forecasts from the actual MCP are quite high and thus, the estimated CI fail to provide adequate estimates for the range of future spot prices. We find that more than 60% (70%) of the 99% (90%) CI fail to comprise the actual MCP during the period August 21-27, 2000.

Quite different results are obtained for the AR/ARX-G and TAR/TARX models. Due to the higher volatility in this period, the AR/ARX-G models give higher estimates for the conditional volatility while the TAR/TARX models give higher probabilities for staying in the spiky regime. Thus, CI forecasts for these models are much wider in comparison to the first week of the test period. For certain hours forecasted 99% CI are wider than 150 USD/MWh giving an extensive range of possible one-day ahead spot prices. As a result, only about 2% (15%) of the 99% (90%) CI fail to comprise the actual MCP for the AR-G model. Results for the non-linear TARX model are similar, yielding approximately 2% (10%) exceedances of the 99% (90%) CI.

Figure 4 displays the actual percentage of exceedances for the 35 weeks in the out-of-sample period for the six models. For the AR/ARX models (upper panels), obviously the number of exceedances is systematically too high due to the too narrow interval forecasts. Especially for the period starting from week 11 the empirically observed number of exceedances is higher than the theoretical confidence level would suggest. This is true for all confidence levels of 50%, 90% and 99%. Most demonstrative, we observe that during volatile periods for several weeks more than 40% of the 99% CI are exceeded. We conclude that despite the rather good results for point forecasts of the ARX model [20], it fails in interval forecasting. Determined CI are clearly too narrow and do not distinguish adequately between phases of lower or higher volatility. The residuals seem to have much heavier tails than suggested by these models. Thus, there is a big difference in model performance between coverage of the actual MCP by the intervals during calm and volatile periods. As indicated by Table 1, the mean percentage of exceedances rises drastically for the AR/ARX models during the volatile time period.

TABLE 1 Mean percent of exceedances of the 50%, 90% and 99% two-sided day-ahead confidence intervals (CI) by the actual market clearing price (MCP) for the six considered models.

	AR			**ARX**		
Weeks	50%	90%	99%	50%	90%	99%
1-10	35.42	10.18	5.60	34.52	10.71	5.36
11-35	76.38	51.64	35.05	75.00	51.57	35.31
	AR-G			**ARX-G**		
	50%	90%	99%	50%	90%	99%
1-10	41.85	12.08	5.12	41.85	13.15	4.88
11-35	42.86	13.45	5.60	47.93	15.64	6.81
	TAR			**TARX**		
	50%	90%	99%	50%	90%	99%
1-10	38.63	11.55	5.18	36.55	11.07	4.70
11-35	37.14	9.60	3.76	37.71	9.24	3.31

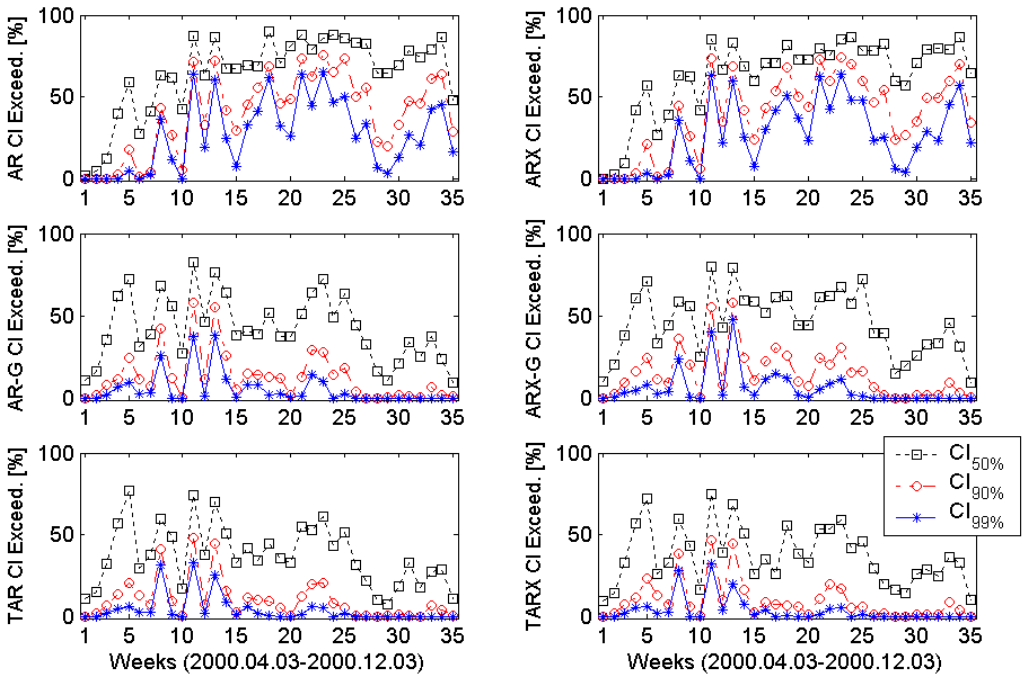

FIGURE 4 Percent of exceedances of the 50%, 90% and 99% two-sided day-ahead confidence intervals (CI) by the actual market clearing price (MCP) for the six considered models during all 35 weeks of the test period.

The results are substantially better for the AR/ARX-G and TAR/TARX models (middle and lower left panels, respectively). However, we find that due to heavy tails in the residuals, for a confidence level of 99% the intervals are still too narrow. Here, the average number of exceedances of the intervals is higher than it is expected theoretically, but is substantially lower than for the linear models. Especially the TARX model gives very good results for the interval forecasts for the 90% confidence level where the number of exceedances is approximately 11% during the calm and 9% during the spiky period. Still, as it is indicated by Figure 4, the number of exceedances shows large variations through time: while we observe periods where the actual percentage of exceedances is significantly too high (e.g. weeks 8, 11, 13, 16-18), there are also several weeks where the percentage of exceedances is clearly too low. However, in terms of the mean percent of exceedances (see Table 1) there is only a slight difference between the calm (weeks 1-10) and volatile (weeks 11-25) period, indicating the overall adequacy of the models for both calm and volatile periods. Hence, we conclude that AR/ARX-G and TAR/TARX models are clearly superior to AR/ARX models when it comes to interval forecasts.

In view of the very good performance of the TAR/TARX models in point forecasts [14] and their superior performance for interval predictions, the TARX model can be regarded as the overall winner.

4. CONCLUSIONS

In this paper we investigated the forecasting power of various time series models for electricity spot prices. The models included different specifications of linear autoregressive time series with heteroscedastic noise and/or additional fundamental variables. Further, non-linear threshold regime-switching models (TAR and TARX) were considered. The models were tested on a time series of hourly system prices and loads from the California power market.

Following [1] and [5] we evaluated the quality of the predictions by comparing the nominal coverage of the models to the true coverage. We found that the TAR/TARX models, as well as the AR/ARX-G specifications gave the best results. The simpler AR/ARX models predicted too narrow confidence intervals and thus, the number of exceedances of the intervals was systematically too high. Still, even for AR/ARX-G and TAR/TARX models, due to heavy-tailed residuals, the intervals for the 99% confidence level were too narrow.

We conclude that in contrast to other studies on the failure of non-linear models in forecasting [2][7][8], in our case the TARX model gave the best overall results. Both for point (see 14) and interval forecasting the model outperformed most of its competitors and was the best in several of the considered criteria. Consequently, we recommend the threshold AR/ARX models for forecasting of highly volatile electricity spot prices. The striking underperformance of the AR/ARX models in interval predictions, in our opinion, makes them an inferior candidate for forecasting of electricity spot prices in general. For purposes like determining risk figures based on confidence intervals, the non-linear approaches (TAR, TARX) and models with an additional GARCH component are clearly more favorable.

Finally, we note that the differences between pure price models and specifications with fundamental/exogenous variables are generally negligible. TARX is slightly better than TAR, AR and ARX are roughly the same, only ARX-G underperforms in the volatile period compared to AR-G.

5. REFERENCES

1. Baillie R., Bollerslev T.: *Prediction in dynamic models with time-dependent conditional variances*, Journal of Econometrics 52, 1992, 91-113.
2. Bessec M., Bouabdallah O.: *What causes the forecasting failure of Markov-switching models? A Monte Carlo study*, Studies in Nonlinear Dynamics and Econometrics 9(2), 2005, Article 6.
3. Bollerslev T.: *Generalized autoregressive conditional heteroscedasticity*, Journal of Econometrics 31, 1986, 307-327.
4. Bunn D.W.: *Forecasting loads and prices in competitive power markets*, Proc. IEEE 88(2), 2000, 163-169.
5. Christoffersen P., Diebold F. X.: *How relevant is volatility forecasting for financial risk management*, Review of Economics and Statistics 82, 2000, 12-22.
6. Contreras J., Espínola R., Nogales F.J., Conejo A.J.: *ARIMA models to predict next-day electricity prices*, IEEE Trans. Power Systems 18(3), 2003, 1014-1020.
7. Crawford G., Fratantoni M.: *Assessing the forecasting performance of regime-switching, ARIMA and GARCH models of house prices*, Real Estate Economics 31(2), 2003, 223-243.
8. Dacco R., Satchell C.: *Why do regime-switching models forecast so badly?*, Journal of Forecasting 18, 1999, 1-16.
9. Engle R.F.: *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation*, Econometrica 50, 1982, 987-1007.
10. 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.
11. Hamilton J.: *Time Series Analysis*, Princeton University Press, 1994.
12. Hansen B.E.: *Inference in TAR models*, Studies in Nonlinear Dynamics and Econometrics 2(1), 1997, 1-14.
13. Ljung L.: *System Identification – Theory for the User*, 2nd ed., Prentice Hall, Upper Saddle River, 1999.
14. Misiorek A., Trück S., Weron R.: *Point and Interval Forecasting of Spot Electricity Prices: Linear vs. Non-Linear Time Series Models*, Studies in Nonlinear Dynamics and Econometrics, 2006, *forthcoming*.
15. Nogales F., Contreras J., Conejo A., Espínola R.: *Forecasting next-day*

electricity prices by time series models, IEEE Trans. Power Systems 17, 2002, 342-348.
16. Rodriguez C.P., Anders G.J.: *Energy price forecasting in the Ontario competitive power system market*, IEEE Transactions on Power Systems 19(1), 2004, 336-374.
17. Shahidehpour M., Yamin H., Li Z.: *Market Operations* in Electric Power Systems: Forecasting, Scheduling, and Risk Management, Wiley, 2002.
18. Szkuta B., Sanabria L., Dillon T.: *Electricity price short-term forecasting using artificial neural networks*, IEEE Trans. Power Syst. 14, 1999, 851-857.
19. Weron R., Misiorek A.: *Forecasting spot electricity prices with time series models*, Proceedings of the European Electricity Market EEM-05 Conference, Łódź, 2005, 133-141.
20. Weron R., Misiorek A.: *Short-term electricity price forecasting with time series models: A review and evaluation*. In: Complex Electricity Markets, Mielczarski, W. (ed.), Łódź, 2006.

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

Dr. Rafał Weron
received his M.Sc. (1995) and Ph.D. (1999) degrees in applied mathematics from the Wroclaw University of Technology (WUT), Poland. He currently holds a position of Assistant Professor at WUT. 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 4 books and over 70 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: rafal.weron@pwr.wroc.pl