Point and interval forecasting of wholesale electricity prices: Evidence from the Nord Pool market

Rafal Weron and Adam Misiorek

Hugo Steinhaus Center, Wroclaw University of Technology

2006

MEPS '06 MODERN ELECTRIC POWER SYSTEMS
Wrocław, September 6-8, 2006
IEEE POLAND SECTION

Point and interval forecasting of wholesale electricity prices: Evidence from the Nord Pool market

Rafał Weron
Hugo Steinhaus Center
Wrocław University of Technology
50-370 Wrocław, Poland
Rafal.Weron@pwr.wroc.pl

Adam Misiorek[1]
Institute of Power Systems Automation
Wystawowa 1
51-618 Wrocław, Poland
misiorek@iase.wroc.pl

Abstract - **In this paper we assess the short-term forecasting power of different time series models in the Nord Pool electricity spot market. We evaluate the accuracy of both point and interval predictions; 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 the interval forecasts of all models are overestimated in the relatively non-volatile periods.**

Keywords: **Wholesale electricity price, Point forecast, Interval forecast, AR model, Threshold AR model.**

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.

This paper is a continuation of our earlier studies on STPF of California electricity prices with time series models [7][8][10]. Here we address the question whether the same techniques yield equally good point and interval forecasts of Nord Pool spot prices. Consequently, we limit the range of analyzed models to linear and non-linear time series approaches that have been found to perform well for pre-crash California power market data (for descriptions of model classes we refer to the above mentioned papers). The list includes autoregression (AR), threshold AR and spike preprocessed AR models. Like in [7][8][10], 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. However, unlike for California, the Nordic area's system wide loads (and their day-ahead forecasts) are not publicly available. The only exogenous information we consider is the hourly air temperature. This lets us expand the range of models to include ARX, TARX and spike preprocessed ARX specifications.

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

2. DATA AND MODELS

In this paper we forecast hourly Nord Pool market clearing prices (MCPs) from the year 2004. Data from the period April 2, 2003 – December 31, 2003 was used only for calibration and from the period January 1 – December 5, 2004 for out-of-sample testing (and step-by-step recalibration). Four five-week periods were selected for model comparison: 26/1–29/2, 26/4–6/6, 26/7–5/9 and 1/11–5/12. Note, that they roughly correspond to the months of February, May, August and November. This lets us evaluate the performance of the models for all seasons of the year. Moreover, this large out-of-sample interval allows for a more thorough analysis of the forecasting results than typically used in the literature single week test samples.

The only exogenous information we consider is the air temperature, since generally it is the most influential (on electricity prices) weather variable [11]. Hourly air temperatures for six Scandinavian cities/locations (Bergen, Helsinki, Malmö, Stockholm, Oslo and Trondheim) were kindly provided by the Swedish Meteorological and Hydrological Institute (SMHI). We used the arithmetic mean of the air temperatures of these six cites as a proxy for the air temperature of the whole Nord Pool region.

The missing and "doubled" price and temperature data values, including those corresponding to the changes to and from the daylight saving time, were treated in the usual way. The former were substituted by the arithmetic average of the two neighbouring values, while the latter by the arithmetic average of the two values for the "doubled" hour. Likewise, the few outliers (but not the spikes; spike preprocessing is addressed later in this Section) were substituted by the arithmetic average of the two neighbouring values. The obtained time series are depicted in Fig. 1. Apparently the seasonal correlation between prices and temperatures during the analyzed period is rather limited. This is confirmed by the correlogram in Fig. 2.

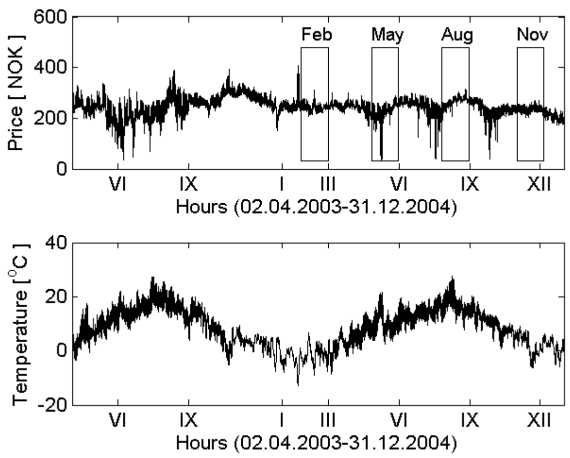

Fig. 1. Hourly Nord Pool market clearing prices (*top panel*) and hourly Scandinavian air temperatures (*bottom panel*) for the period April 2, 2003 – December 31, 2004. The four five-week out-of-sample test periods are denoted by hollow rectangles in the top panel.

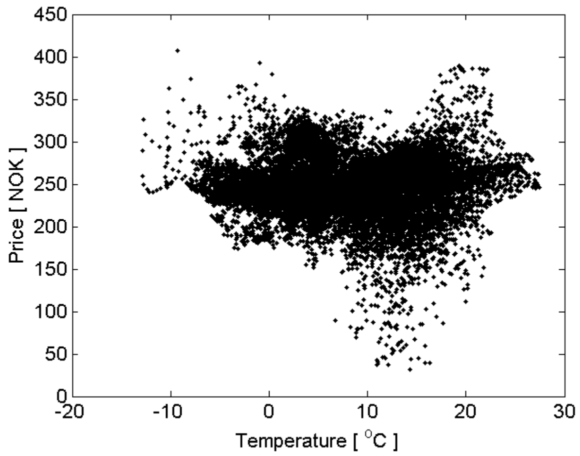

Fig. 2. The correlogram of the hourly spot price vs. hourly air temperature during the period April 2, 2003 – December 31, 2004 indicates that the seasonal correlation between these two factors is rather limited.

The models considered in this study comprised simple time series specifications with and without exogenous variables, namely ARX and AR processes, spike preprocessed ARX and AR models and regime-switching threshold autoregression (TARX/TAR) models. The calibration was performed in Matlab using the prediction error estimate method.

The logarithmic transformation $p_t = \log(P_t)$ was applied to price data to attain a more stable variance. Furthermore, the mean was removed to center the data around 0. Since each hour displays a rather distinct price profile reflecting the daily variation of demand, costs and operational constraints the modelling 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][4][9][11].

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.

Furthermore, we have found that a large moving average part (of ARMA-type specifications) 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/AR-type models. Likewise, a large autoregression part (we tested models with lags up to four weeks) generally led to overfitting and worse out-of-sample forecasts.

Analogous results were obtained earlier for California power market data [7][8][10]. Consequently, we have chosen to use here the same AR structure that was found to be optimal for California:

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

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 resulting **AR** model can be written as:

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

where $\phi(B)p_t$ is given by (1) and d_1, d_2, d_3 denote the coefficients of the dummies $D_{Mon}, D_{Sat}, D_{Sun}$, respectively. The corresponding **ARX** model structure is given by:

$$\phi(B)p_t = \psi_1 z_t + d_1 D_{Mon} + d_2 D_{Sat} + d_3 D_{Sun} + \varepsilon_t, \qquad (3)$$

where ψ_1 is the coefficient of the actual air temperature z_t observed at the forecasted hour.

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

$$\begin{cases} \phi_1(B)p_t = \psi_{1,1} z_t + d_{1,1} D_{Mon} \\ \qquad + d_{1,2} D_{Sat} + d_{1,3} D_{Sun} + \varepsilon_t & \text{when } v_t \geq T, \\ \phi_2(B)p_t = \psi_{2,1} z_t + d_{2,1} D_{Mon} \\ \qquad + d_{2,2} D_{Sat} + d_{2,3} D_{Sun} + \varepsilon_t & \text{when } v_t < T, \end{cases} \qquad (4)$$

where v_t and T are the threshold variable and the threshold level, respectively. Based on our earlier experience [7], we have decided to use v_t equal to the difference in mean prices for yesterday and eight days ago and set $T = 0$. The simpler **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.

Price spikes pose a serious problem for linear time series models, which assume stationarity of the signal. Possible solutions involve excluding or limiting price spikes [9][11]. In the first case we treat the abnormal prices as outliers and substitute them with the average of the neighboring observations or with “similar-day” prices. However, price spikes are inherent in electricity prices, so we do not want to delete them completely from the calibration process. Instead of excluding them, we can limit their severity or damp all observations above a certain threshold.

Based on the results of [11], we have decided to use spike damping. Namely, we set an upper limit T, equal to the mean plus three standard deviations of the calibration sample prices, on the price (not the log-price) and if the price P_t is higher than T, it is set to $T + T\log_{10}(P_t/T)$. This scheme allows to differentiate between “regular” and “extreme” spikes. Spike preprocessing was used only in combination with **ARX** and **AR** models. The resulting models (calibrated to spike-damped data) are denoted later in the text by **p-ARX** and **p-AR**, respectively.

Finally, 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 temperatures) and obtained a forecasted value for that day (and hour). Note, that the model structure itself was not optimized at each time step as this procedure did not lead to better results.

3. FORECASTING RESULTS

The prediction accuracy was checked afterwards, once the true market prices were available. To assess the point forecasting performance of the models, different statistical measures can be utilized. Here we use the Mean Weekly Error [3][7][9][11]:

$$MWE = \frac{1}{168}\sum_{h=1}^{168}\frac{\left|P_h - \hat{P}_h\right|}{\overline{P}_{168}}, \qquad (5)$$

where P_h is the actual price for hour h, $\hat{P}_h$ is the predicted price and $\overline{P}_{168}$ is the mean price for a given week.

Forecasting results for the whole test period are summarized in Tab. 1. The MWEs are reported only for pure price models: it turned out that the models with the exogenous variable (air temperature) yielded worse predictions than their simpler counterparts. This result is not that surprising if we recall Fig. 2, which indicates that the seasonal correlation between electricity prices and temperatures is negligible.

Tab. 1. Mean Weekly Errors (MWE) of day-ahead point forecasts for all weeks of the test period (given in percent). Best results in each week are emphasized in bold. Results not passing the naïve test are underlined.

Month/Week	AR	TAR	p-AR	Naïve
February/1	1.78	2.69	1.77	**1.52**
February/2	3.07	3.61	**3.07**	4.13
February/3	**3.16**	3.34	3.16	3.79
February/4	2.09	2.77	2.08	**1.57**
February/5	1.89	1.93	**1.88**	2.13
May/1	5.96	**5.48**	5.95	6.74
May/2	10.88	**10.00**	10.89	11.64
May/3	7.69	**5.57**	7.69	11.22
May/4	4.05	**4.03**	4.04	5.82
May/5	2.30	**1.54**	2.30	2.96
August/1	**2.78**	2.82	2.78	5.14
August/2	2.96	2.89	2.96	**2.42**
August/3	2.09	**1.64**	2.09	3.00
August/4	1.78	1.94	**1.78**	2.08
August/5	2.34	2.38	2.33	**1.53**
November/1	2.10	2.40	2.10	**1.56**
November/2	2.23	2.30	**2.22**	2.40
November/3	1.97	2.04	**1.96**	3.71
November/4	2.34	2.67	**2.33**	2.74
November/5	2.76	**2.67**	2.75	3.05
# best	2	**7**	6	5
# better than naïve	**15**	**15**	**15**	-
mean dev. from best	0.39	**0.25**	0.39	2.70

Following Conejo et al. [3] and Misiorek et al. [7] 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 errors for the estimates are smaller than for the prices of the similar day. Surprisingly often the forecasting procedures did not pass this test. For five weeks (or 25%) of the test period the naïve method yielded the best forecasts.

The overall best approach is the TAR model which yielded the most accurate forecasts for 7 out of 20 weeks and the smallest mean deviation from the best model in each week. The latter characteristic indicates that if the TAR model was not the best one, it was not much worse than the best one. On the other hand, the naïve method led to the largest mean deviation from the best model. In other words, if it wasn’t the best then the predictions were generally much worse than those of the best model for that particular week. For instance, the naïve method’s MWEs for the third and fifth weeks of May were nearly twice larger than those of the TAR model.

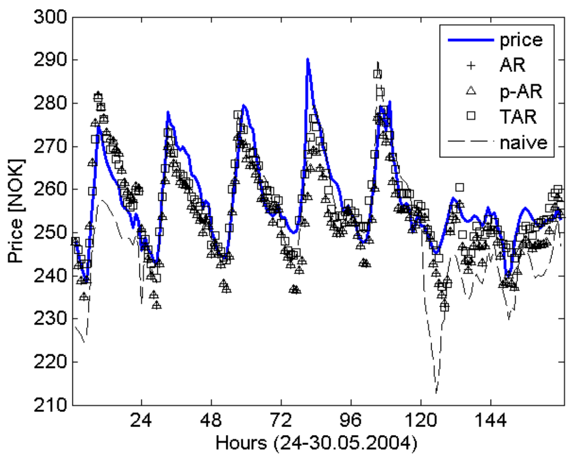

Fig. 3. Prediction results for the fourth week of May for the naïve, AR, p-AR (preprocessed ARX) and TAR models. Apparently the naïve method has problems with adapting to the changing characteristics of the price process.

We have to note, however, that the TAR model does not exhibit equally good performance for the whole test sample. It is particularly powerful during the very volatile weeks of May 2004 (see Fig. 3), but is generally inferior to the AR and p-AR models during the calmer months of February, August and November. The AR and p-AR models, on the other hand, perform almost identically, partly because the test samples do not exhibit large (upward) spikes. The p-AR model is slightly better than the original AR specification, however, their mean deviations from the best model are practically the same.

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 [5][6]. Afterwards, following [2], 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 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 “month”, 840 hourly values were determined and compared to the MCP.

Examining the deviations of the CIs from the actual MCP for the third week of February (left panels in Fig. 4), 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 and p-AR models the intra-week variation of the intervals is smaller than for the TAR model. However, the difference is not as large as for the analogous models calibrated to California power market data [7][8].

Tab. 2. 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 three considered models.

CI level	**50%**	**90%**	**99%**
	AR		
February	11.79	0.24	0.00
May	38.21	12.98	8.45
August	10.60	0.36	0.00
November	13.81	0.71	0.00
	p-AR		
February	12.02	0.24	0.00
May	38.33	12.98	8.45
August	10.12	0.36	0.00
November	13.45	0.71	0.00
	TAR		
February	6.19	0.60	0.00
May	25.48	9.05	3.69
August	4.29	0.24	0.00
November	6.43	0.48	0.00

Examining the results for the second week of May (right panels in Fig. 4; see also Tab.2), we find that the estimated 90% and especially the 99% 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 CIs fail to provide adequate estimates for the range of future spot prices. Better results are obtained for the TAR model, which roughly captures the 90% CI. However, it predicts slightly too narrow 99% intervals and significantly too wide 50% intervals.

4. CONCLUSIONS

In this paper we investigated the forecasting power of time series models for electricity spot prices. The models included linear and non-linear autoregressive time series with and without additional fundamental variables. The models were tested on a time series of hourly system prices and temperatures from the Nordic power market.

We evaluated the quality of the predictions both in terms of the Mean Weekly Error (for point forecasts) and in terms of the nominal coverage of the models to the true coverage (for interval predictions). We found that during relatively calm periods the AR and spike preprocessed AR (p-AR) models generally yielded better point forecasts than their competitors, with p-AR being slightly better than the pure AR specification. However, during volatile weeks of May 2004 the TAR model was definitely the best. Overall it also yielded the smallest mean deviation from the best model in each week.

Regarding interval forecasts we found that the estimated 90% and especially the 99% intervals of the linear models are clearly too narrow for the volatile period. Better results are obtained for the TAR model, which roughly captures the 90% CI. However, it predicts slightly too narrow 99% intervals and significantly too wide 50% intervals.

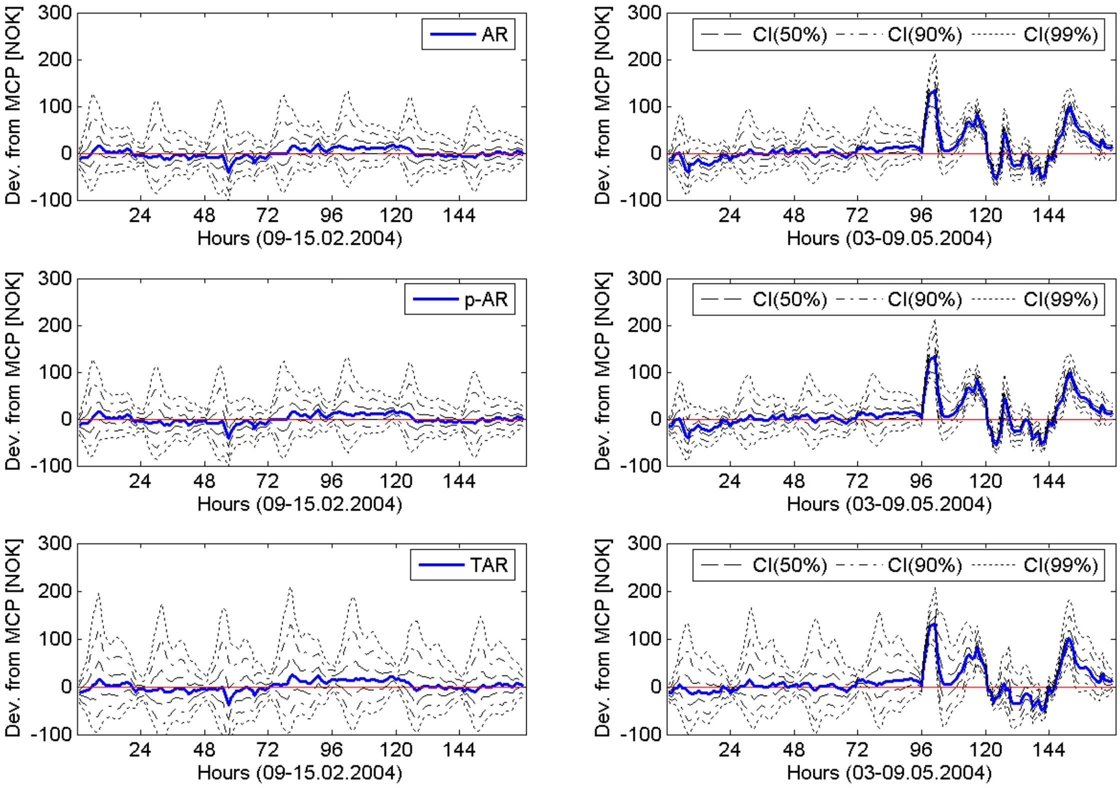

Fig. 4. Deviation of the day-ahead point forecasts and their respective 50%, 90% and 99% two-sided confidence intervals (CI) from the actual market clearing price (MCP) for three models: AR (*top panels*), p-AR (*middle panels*) and TAR (*bottom panels*), and for two weeks of the test period: the third week of February and the second week of May.

Moreover, we found that during relatively calm periods 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% CIs deviations from the actual MCP are rarely high enough to exclude the price from the interval. This is in contrast to the results for the California power market [7][8], where the TAR model yielded acceptable interval forecasts for the whole test sample. A possible reason for such a behavior could be temporal dependence (or "non-whiteness") in the model residuals. Whether this is true has yet to be investigated.

REFERENCES

[1] D.W. Bunn, "Forecasting loads and prices in competitive power markets", *Proc. IEEE* 88, 2000, 163-169.

[2] P. Christoffersen and F.X. Diebold, "How relevant is volatility forecasting for financial risk management", *Review of Economics and Statistics* 82, 2000, 12-22.

[3] A.J. Conejo at al., "Forecasting electricity prices for a day-ahead pool-based electric energy market", *International Journal of Forecasting* 21, 2005, 435-462.

[4] E.A. Feinberg and D. Genethliou, "Load forecasting" In: *Applied Mathematics for Restructured Electric Power Systems: Optimization, Control, and Computational Intelligence*, J.H. Chow et al. (eds.), Springer, 2005.

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

[6] B.E. Hansen, "Inference in TAR models", *Studies in Nonlinear Dynamics and Econometrics* 2,1997, 1-14.

[7] A. Misiorek at al., "Point and Interval Forecasting of Spot Electricity Prices: Linear vs. Non-Linear Time Series Models", *Studies in Nonlinear Dynamics and Econometrics*, 2006, *forthcoming*.

[8] A. Misiorek and R. Weron, "Interval forecasting of spot electricity prices", *Proceedings of the EEM-06 Conference*, Warszawa, 2006.

[9] M. Shahidehpour at al., *Market Operations* in Electric Power Systems: Forecasting, Scheduling, and Risk Management, Wiley, 2002.

[10] R. Weron and A. Misiorek, "Short-term electricity price forecasting with time series models: A review and evaluation" In: *Complex Electricity Markets*, Mielczarski, W. (ed.), Łódź, 2006.

[11] R. Weron, *Modeling and forecasting electricity loads and prices: A statistical approach*, Wiley, Chichester, 2006, *forthcoming*.

[1] This work was supported in part by KBN grant 4-T10B-030-25.