On the impact of outlier filtering on the electricity price forecasting accuracy☆

Dmitriy O. Afanasyev[a,*], Elena A. Fedorova[b,a]

[a]Financial University under the Government of the Russian Federation, 49 Leningradskiy av., Moscow, Russia
[b]National Research University Higher School of Economics, 20 Myasnitskaya st., Moscow, Russia

One of the principal research questions which is still not sufficiently studied in the modern literature (Weron, 2014) is accounting for the long-term seasonal component (LTSC) in the price of electricity, specifically in the context of short-term forecasting. The several studies attempt to fill in this gap by separately modeling and forecasting of the long-term deterministic and short-term stochastic price components (see, for example, Tan et al., 2010; Keles et al., 2016; Nowotarski and Weron, 2016). The authors conclude that accounting for LTSC and a proper choice of the method and parameters for the LTSC estimation yield a significant decrease of the forecast error as compared both to the traditional ARX model and to the naïve approach (Conejo et al., 2005).

But in the mentioned forecasting researches the authors pay little attention to such an important issue as sharp short-term and, generally speaking, poorly predictable extreme values of the electricity price, i.e. spikes. These outliers usually occur due to accidents at power plants, congestions of the energy transmission networks, and climatic anomalies. In the pioneer papers of Trück et al. (2007); Janczura et al. (2013), concerning the problem of spikes, the authors note that, in the presence of such extreme price values, subtracted deterministic and stochastic components of the electricity price may be substantially distorted. As a result, the adequacy of models for such components (especially, linear models where SCARX belongs) is very doubtful, with the estimates of parameters of these models being potentially biased. In the case of in-sample modeling such biases may cause, for example, under-evaluation of value-based risk measures (such as VaR and EaR), while in the case of application of such models to the out-of-sample period (i.e., solving the forecasting problem), this may potentially corrupt the price forecast accuracy.

It is important to emphasize that the situation is worsened by the fact that in the literature there is no consensus on which price values are to be considered as spikes, and thus, which spike identification approach should be used. In the framework of in-sample modeling, there is quite a lot of approaches discussed in the literature: threshold filter on prices, standard deviation filter on prices, percentage filter on prices, recursive and moving filters on prices, Markov regime-switching models, etc. (Lapuerta and Moselle, 2001; Trück et al., 2007; Weron, 2006; Borovkova and Permana, 2006; Fanone et al., 2013; Janczura et al., 2013). But the application of these methods may lead to principally different results, with the number of observations classified as spikes (extreme values) being dramatically different for

☆The authors are grateful to Gilenko E.V. (Ph.D., St. Petersburg State University, Russia) for his valuable and stimulating comments on the paper, as well as assist for translation. This work was supported by the Russian Foundation for Basic Research (RFBR) [grant number 16-06-00237 A].

*Corresponding author. Phone: +7 926 6320115
Email addresses: dmafanasyev@gmail.com (Dmitriy O. Afanasyev), ecolena@mail.ru (Elena A. Fedorova)
URL: http://dmafanasyev.ru (Dmitriy O. Afanasyev)

the same time-series. On the one hand, it is quite natural to assume that neglecting pre-filtering of extreme values has a negative impact on the forecast accuracy. But, on the other hand, too aggressive filtering of spikes (for example, by applying a filter with a low cut-off threshold) will substantially distort the original time-series and result in predictions virtually unrelated to the original historical dynamic of the price.

Taking into account the fact that previous papers majorly focus on the effect of pre-filtering of extreme values in the framework of in-sample electricity price modeling, as well as considering a significant role that the estimates of long-term trend-seasonal and short-term seasonally stochastic components play in the problem of forecasting, in this research we pose a question of the impact of outlier filtering on the electricity price forecasting accuracy.

The basis model of our research is seasonal component autoregressive with exogenous factors (SCARX) that proposed by Nowotarski and Weron (2016) and consists of the following steps.

At the first step, the time-series of hourly electricity prices $p_t, t = 1, \dots, T$ is taken the logarithm of and is additively decomposed into a long-term seasonal (trend-cyclical) L_t and a short-term seasonally stochastic S_t components:

$$\log p_t = P_t = L_t + S_t \tag{1}$$

We use the wavelet-decomposition for LTSC estimation as in Nowotarski and Weron (2016); Xiao et al. (2017); Yang et al. (2017) and consider a wide range of the scale parameter values $m = 8, \dots, 13$.

At the second step, for the short-term seasonally stochastic component S_t parameters of the autoregressive specification proposed by Misiorek et al. (2006) are estimated:

$$\begin{aligned} S_t = \alpha_1 S_{t-24} + \alpha_2 S_{t-48} + \alpha_7 S_{t-168} + \\ + \alpha_8 m S_t + \beta_1 Z_t + \sum_{i=1,6,7} d_i D_t^i + \varepsilon_t \end{aligned} \tag{2}$$

where $S_{t-24/48/168}$ are autoregressive components; mS_t is the "price signal" equal to the minimum price of the previous day; Z_t ia a day-ahead forecast of electricity consumption before time moment t; D_t^i are dummy variables accounting for weekly seasonality ($i = 1, 6, 7$ for Monday, Saturday, and Sunday correspondingly); ε_t is a normally, independently, and identically distributed error term. The modeling is run for each of the day hour separately, i.e. there are 24 models in total.

At the third step, the L_t is assumed to be persistent in the short-term scales, i.e. its forecast value is equal to the value in the corresponding hour of the previous day. Finally, *at the fourth step*, the early obtained forecasts $\hat{S}_{T+h}$ and $\hat{L}_{T+h}$ are summed up, and the inverse logarithmic transformation is calculated which gives the final forecast $\hat{p}_{T+h}$. We will denote the model as $\mathbf{SCARX_m}$.

The choice of the SCARX-model as a basis model for our study is dictated, firstly, by the fact that this model separately estimates the long-term trend-cyclical and the short-term seasonally stochastic components of the price, with both of the components being prone to substantial influence of outliers; and, secondly, by the fact that the model, by changing the smoothing parameter, allows to obtain a whole range of quite independent models, and, thus, to have more objective results of testing on them.

We extend the SCARX-model by the following well known outlier filters (X_t – the part of S_t obtained after remove of weekly and intra-day seasonality, X_t^o – the subset of price spikes).

Threshold filter on prices (**TFP**):

$$X_t^o = \{X_t \,:\, |X_t| \geq 0.5\} \tag{3}$$

Standard deviation filter on prices (**SFP**):

$$X_t^o = \{X_t \,:\, |X_t - \overline{X}| \geq 3 \cdot \sigma\} \tag{4}$$

Recursive filters on prices (**RFP**):

$$X_{t,i}^o = \{X_t \,:\, |X_t - \overline{X}_i| \geq 3 \cdot \sigma_i\} \tag{5}$$

Moving filters on prices (**MFP**):

$$X_t^o = \bigcup_{n=1,\dots,N} \{X_{\tau_n} \,:\, |X_{\tau_n} - \overline{X}_n| \geq 1.96 \cdot \sigma_n\} \tag{6}$$

Table 1: Comparison of forecasting performance (in the WMAE sense) of SCARX-models (with wavelet smoothing parameters $m = 8, ..., 13$) with and without outlier filtering for power markets ATS Europe-Ural and ATS Siberia.

Models	Outlier filters						
	SRC	TFP	SFP	RFP	MFP	PFP	CFP
				ATS Europe-Ural (EU)			
ARX	4.72	–	–	–	–	–	–
$SCARX_8$	4.82	4.53 (-0.29)	4.57 (-0.25)	4.95 (0.13)	4.66 (-0.16)	5.13 (0.31)	4.58 (-0.24)
$SCARX_9$	4.73	4.46 (-0.27)	4.48 (-0.25)	4.76 (0.03)	4.60 (-0.13)	5.00 (0.27)	4.49 (-0.24)
$SCARX_{10}$	4.67	4.44 (-0.23)	4.46 (-0.21)	4.64 (-0.03)	**4.56** (-0.11)	4.90 (0.23)	**4.47** (-0.20)
$SCARX_{11}$	4.66	**4.43** (-0.23)	**4.44** (-0.22)	4.58 (-0.08)	4.60 (-0.06)	4.93 (0.27)	4.48 (-0.18)
$SCARX_{12}$	4.65	4.44 (-0.21)	4.45 (-0.20)	**4.55** (-0.10)	4.63 (-0.02)	4.90 (0.25)	4.48 (-0.17)
$SCARX_{13}$	**4.62**	4.46 (-0.16)	4.50 (-0.12)	4.76 (0.14)	4.68 (0.06)	4.89 (0.27)	4.56 (-0.06)
Summary							
Δ	–	-0.23	-0.21	0.01	-0.07	0.27	-0.18
# < SRC	–	6 (100%)	6 (100%)	3 (50%)	5 (83%)	0 (0%)	6 (100%)
				ATS Siberia (SI)			
ARX	8.84	–	–	–	–	–	–
$SCARX_8$	8.93	7.75 (-1.18)	7.72 (-1.21)	8.02 (-0.91)	8.09 (-0.84)	7.82 (-1.11)	7.75 (-1.18)
$SCARX_9$	9.07	7.58 (-1.49)	7.64 (-1.43)	7.74 (-1.33)	7.94 (-1.13)	7.86 (-1.21)	7.58 (-1.49)
$SCARX_{10}$	8.77	7.47 (-1.30)	7.53 (-1.24)	7.41 (-1.36)	7.73 (-1.04)	**7.51** (-1.26)	7.46 (-1.31)
$SCARX_{11}$	8.83	**7.42** (-1.41)	7.52 (-1.31)	**7.34** (-1.49)	7.74 (-1.09)	7.52 (-1.31)	7.41 (-1.42)
$SCARX_{12}$	9.35	7.53 (-1.82)	7.69 (-1.66)	7.45 (-1.90)	8.03 (-1.32)	7.89 (-1.46)	7.50 (-1.85)
$SCARX_{13}$	**8.72**	7.96 (-0.76)	**7.45** (-1.27)	7.36 (-1.36)	**7.71** (-1.01)	7.63 (-1.09)	**7.35** (-1.37)
Summary							
Δ	–	-1.33	-1.35	-1.39	-1.07	-1.24	-1.44
# < SRC	–	6 (100%)	6 (100%)	6 (100%)	6 (100%)	6 (100%)	6 (100%)

Comments: WMAE values (in percents) averaged over the test period are given in rows with models names. Absolute deviations of errors of the model with outlier filtering from errors of the model without outlier filtering are presented in parentheses. Minimum values of WMAE for specific filter and power market (which also are smaller than the error of ARX-model) are given in bold. Minimum values of WMAE for all filters within a specific power market are underlined. Δ is the average change of error within one specific filter. "# < SRC" is the number of SCARX-models with a filter, which produced errors smaller than the errors of the same models on the original data.

Percentage filter on prices (**PFP**):

$$X_t^o = \{X_t : X_t \le X_t^{2.5}\} \cup \{X_t : X_t \ge X_t^{97.5}\} \quad (7)$$

Besides the above-mentioned approaches to data filtering, we also propose a combined filter on prices (**CFP**), which, to the best of our knowledge, has not been applied in the previous studies on the energy economics in the context of outlier identification. CFP is based on the committee machine approach from the field of machine learning and implies classification of an observation as outlier only in the case when at least a half ($Q = 0.5$) of the basic K algorithms predicted the observation as such (Ablow and Kaylor, 1965; Tresp, 2001):

$$X_t^o = \{X_t : \frac{1}{K}\sum_{k=1}^{K} \mathbb{I}(X_t \in X_{t,k}^o) \ge Q\} \quad (8)$$

For the empirical experiment, we used data of four day-ahead electricity markets: the Europe-Ural (EU) and Siberia (SI) areas of the Russian ATS market; the biggest European market Nord Pool (NP), as well as the US power market Pennsylvania-New Jersey-Maryland (PJM). Performance evaluation was made on the basis of two independent approaches – weekly weighted mean absolute error WMAE (Table 1 are partially shown the obtained values) and a formal statistical procedure of a model confidence set (MCS) identification (Hansen et al., 2011). The obtained results allowed us to make the following interest conclusions.

Firstly, outlier pre-filtering of the original data in many cases lets achieve substantial forecasting accuracy gain. For example, in the Siberia area of the Russian ATS market the obtained decrease in forecasting error lies in the range of 1-1.5% of the

Table 2: Minimum averaged WMAE for the most precise SCARX-models, as found by on-grid optimization of parameter values for filters TFP, SFP, RFP, and MFP for power markets ATS Europe-Ural ($SCARX_{13}$), ATS Siberia ($SCARX_{13}$), Nord Pool ($SCARX_{10}$), and Pennsylvania-New Jersey-Maryland ($SCARX_{12}$).

Market	Outlier filters					
	SRC	TFP	SFP	RFP	MFP	CFP*
EU	4.62	4.44 (0.75 / 0.3%)	4.45 (6.5 / 0.2%)	4.44 (9.0 / 0.2%)	4.46 (1.000 / 0.3%)	4.56
SI	8.72	7.43 (0.75 / 1.9%)	7.36 (2.0 / 1.8%)	7.33 (2.0 / 2.2%)	7.71 (0.950 / 4.2%)	7.35
NP	8.38	8.23 (0.75 / 0.5%)	8.23 (4.5 / 0.4%)	8.25 (6.0 / 0.3%)	8.23 (0.999 / 0.9%)	8.31
PJM	10.60	10.57 (1.50 / 0.1%)	10.58 (6.5 / 0.0%)	10.58 (6.0 / 0.0%)	10.60 (1.000 / 0.1%)	11.21

Comments: the optimal parameter values and the corresponding average fraction of removed outliers for the specified WMAE are given in parentheses. SRC denotes the results for the original, unfiltered data. CFP* is the combined filter based on the individual filters with unoptimized parameters.

average weekly price. This is a very good result since it potentially may bring significant financial saving to the market participants.

Secondly, among the studied filters with a priori set parameters, the most stable positive results were demonstrated by filters TFP (with the threshold for the logarithmic price being equal to 0.5), SFP (with the threshold being equal to three standard deviations), and CFP (the majority vote rule). TFP provides accuracy gain in 63% of the cases, SFP – in 67%, and CFP – in 63%. In the context of MCS-based ranking, increase of probability for a model to get into the MCS occurs for these filters in 58%, 54%, and 54% cases, correspondingly.

Thirdly, the main cause of difference in the results of application of the filters lies not in the difference of their algorithms, but in the improper a priori choice of the parameters of these filters. If a grid-search to find forecast error-minimizing value of the filter parameter is run, then, concerning the obtained forecasting accuracy, the difference between the filters will be negligible. In particular, filters RFP and MFP, after finding the optimal values of their parameters, allowed to obtain the final forecast errors only slightly different from the ones of TFP and SFP for each of the markets (see Table 2).

Fourthly, the proposed combined filter based on the committee machine is a quite competitive alternative to the individual methods both in the case of a priori set parameters, and in the case of grid-search optimized values of these parameters. In the former case, CFP plays a role of a more objective method which lowers the risk of an a priori choice, taking into account the results of each of the independent filters. In the latter case, the grid-search procedure requires substantial computational resources and cannot be applied to the future, practically unavailable data, while CFP uses as the basis algorithms an un-optimized filters and is not resource demanding, but at the same time results in forecasting accuracy which is comparable to the best accuracy of individual filters after optimization of their parameters (see column CFP* in Table 2).

Finally, *fifthly*, although the above-given conclusions seem inspiring, still, in rare case both the SCARX-model, and outlier pre-filtering may worsen the results of comparatively simple autoregressive models. An example of such case, among the considered power markets, is the US PJM market, for which no advantages of application of the SCARX-model and outlier pre-filtering were obtained (both for the case of a priori set filter parameters, and for the case of finding their optimized values). This is why the decision on application of the model and pre-filtering should be taken with care.

Although the results of this study shed some light on certain aspects of short-term electricity price forecasting, still there arise a number of new questions. Specifically, combining of forecast is a promising technique of increasing forecasting accuracy (Bates and Granger, 1969; Nan, 2009). But, as it was shown in a number of previous papers, not all the averaging methods allow to obtain good results (Bordignon et al., 2013; Nowotarski et al., 2014). From this perspective, it would be challenging to understand whether pre-filtering helps in solving of existing here problems. Another challenging question that arises in the context of the current research is which is a more efficient tool for obtaining accuracy gain – combining of filters or combining of forecasts on filtered data? We leave these and other questions for further research.

References

Ablow, C., Kaylor, D., 1965. A committee solution of the pattern recognition problem. IEEE Transactions on Information Theory 11, 453–455.

Bates, J., Granger, C., 1969. The combination of forecasts. Journal of the Operational Research Society 20 (4), 451–468.

Bordignon, S., Bunn, D., Lisi, F., Nan, F., 2013. Combining day-ahead forecasts for British electricity prices. Energy Economics 35, 88–103.

Borovkova, S., Permana, F. J., 2006. Modelling electricity prices by the potential jump-diffusion. In: Shiryaev, A. N., Grossinho, M. R., Oliveira, P. E., Esquível, M. L. (Eds.), Stochastic Finance. Springer US, Boston, MA, pp. 239–263.

Conejo, A., Plazas, M., Espínola, R., Molina, A., 2005. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. IEEE Trans. Power Syst. 20, 1035–1042.

Fanone, E., Gamba, A., Prokopczuk, M., 2013. The case of negative day-ahead electricity prices. Energy Economics 35, 22–34.

Hansen, P., Lunde, A., Nason, J., 2011. The model confidence set. Econometrica 79, 453–497.

Janczura, J., Trück, S., Weron, R., Wolff, R., 2013. Identifying spikes and seasonal components in electricity spot price data: A guide to robust modelings. Energy Economics 38, 96–100.

Keles, D., Scelle, J., Paraschiv, F., Fichtner, W., 2016. Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks. Applied Energy 162, 218–230.

Lapuerta, C., Moselle, B., 2001. Recommendations for the Dutch Electricity Market. The Brattle Group Report, London.

Misiorek, A., Trück, S., Weron, R., 2006. Point and interval forecasting of spot electricity prices: Linear vs. non-linear time series models. Studies in Nonlinear Dynamics and Econometrics 10 (3), Article 2.

Nan, F., 2009. Forecasting next-day electricity prices: From different models to combination. PhD Thesis Universita degli Studi di Padova.
URL `http://paduaresearch.cab.unipd.it/2147/`

Nowotarski, J., Raviv, E., Trück, S., Weron, R., 2014. An empirical comparison of alternate schemes for combining electricity spot price forecasts. Energy Economics 46, 395–412.

Nowotarski, J., Weron, R., 2016. On the importance of the long-term seasonal component in day-ahead electricity price forecasting. Energy Economics 57, 228–235.

Tan, Z., Zhang, J., Wang, J., Xu, J., 2010. Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models. Applied Energy 87 (11), 3606–3610.

Tresp, V., 2001. Committee machines. In: Hu, Y., Hwang, J.-N. (Eds.), Handbook for Neural Network Signal Processing. CRC Press, pp. 135–152.

Trück, S., Weron, R., Wolff, R., 2007. Outlier treatment and robust approaches for modeling electricity spot prices. MPRA Paper No. 4711. Hugo Steinhaus Center, Wroclaw University of Technology.
URL `http://mpra.ub.uni-muenchen.de/4711/1/MPRA_paper_4711.pdf`

Weron, R., 2006. Modeling and Forecasting Electricity Loads and Prices: a Statistical Approach. Wiley, Chichester.

Weron, R., 2014. Electricity price forecasting: A review of the state-of-the-art with a look into the future. International Journal of Forecasting 30, 1030–1081.

Xiao, L., Shao, W., Yu, M., Ma, J., Jin, C., 2017. Research and application of a hybrid wavelet neural network model with the improved cuckoo search algorithm for electrical power system forecasting. Applied Energy 198, 203–222.

Yang, Z., Ce, L., Lian, L., 2017. Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods. Applied Energy 190, 291–305.