Advances in forecasting of wholesale electricity prices

Rafał Weron

Department of Operations Research
Wrocław University of Technology (PWR), Poland

(Revised: 25.09.2016)

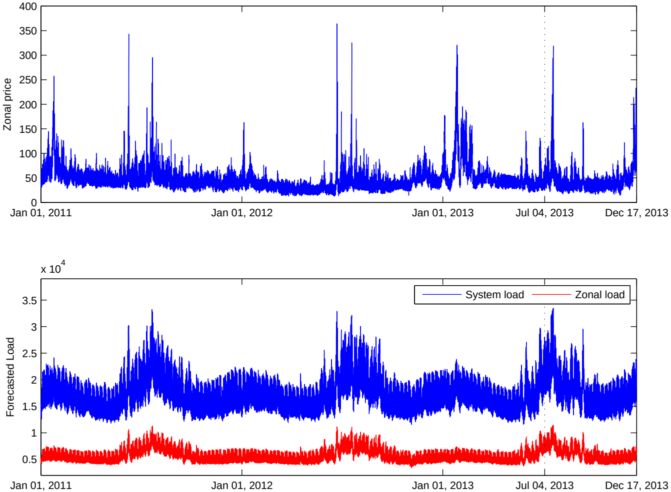
Why forecast loads and prices?

(Hong, 2015, EnergyBiz Magazine)

- A ballpark estimate of savings from a 1% reduction in MAPE for a utility with 1GW peak load:
  - \$500k/year from long-term load forecasting
  - \$300k/year from short-term load forecasting
  - \$600k/year from short-term load + price forecasting
- See Zareipour et al. (2010, TPWRS) for a more 'fundamental' study

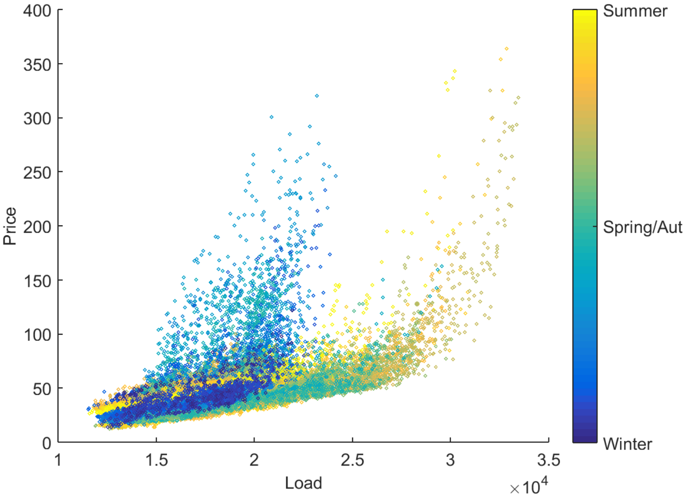
Electricity prices and loads (GEFCom2014)

Seasonality, floor reversion and price spikes

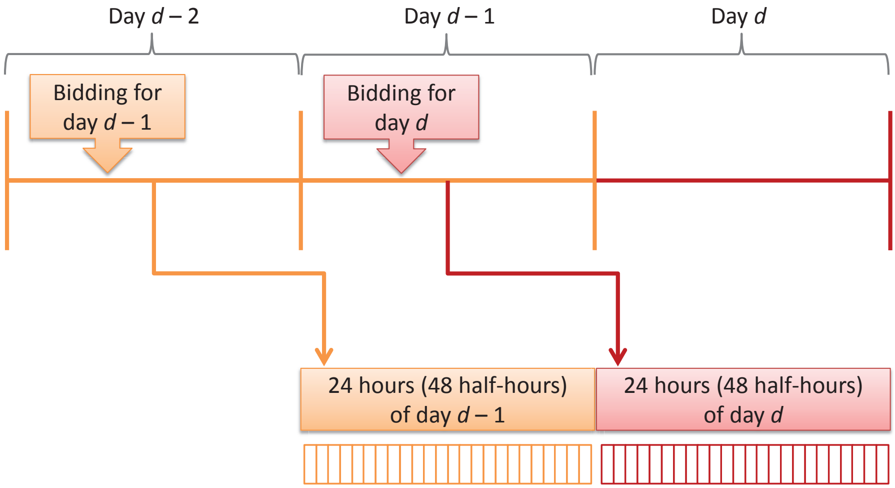


Electricity prices vs. loads (GEFCom2014)

Non-linear, time-varying dependence

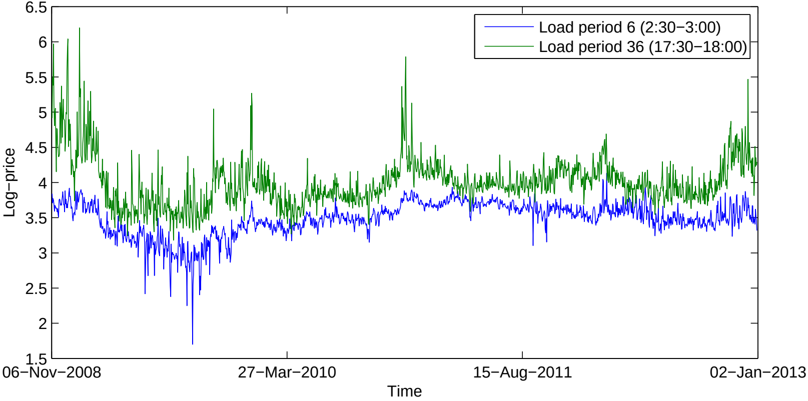


The electricity 'spot' price



Prices for different load periods

Strongly correlated but seem to follow different *data generating processes* (DGPs)



A commodity ... but a very special one

- Not storable (economically)
- Time consuming shut-down/start-up procedures for some technologies
- Extreme price changes → spikes
- Possible negative prices
- Pronounced daily and weekly cycles, annual seasonality
- Mean (floor) reversion
- Highly volatile

Recent reviews

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

ELSEVIER

2014

Review

Electricity price forecasting: A review of the state-of-the-art with a look into the future

CrossMark

Rafał Weron

Institute of Organization and Management, Wrocław University of

ARTICLE INFO

Keywords:
Electricity price forecasting
Day-ahead market
Seasonality
Autoregression
Neural network
Factor model
Forecast combination
Probabilistic forecast

ABSTI

A variety c
the last 15
complexity
and threat
looks ahea
or so. In pa
(i) the sam
testing of t

International Journal of Forecasting 32 (2016) 914–938

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

ELSEVIER

2016

Probabilistic electric load forecasting: A tutorial review

CrossMark

Tao Hong [a,*], Shu Fan [b]

ABSTRACT

Load forecasting has been a fundamental business problem since the inception of the electric power industry. Over the past 100 plus years, both research efforts and industry practices in this area have focused primarily on point load forecasting. In the most recent decade, though, the increased market competition, aging infrastructure and renewable integration requirements mean that probabilistic load forecasting has become more and more important to energy systems planning and operations. This paper offers a tutorial review of probabilistic electric load forecasting, including notable techniques, methodologies and evaluation methods, and common misunderstandings. We also underline the need to invest in additional research, such as reproducible case studies, probabilistic load forecast evaluation and valuation, and a consideration of emerging technologies and energy policies in the probabilistic load forecasting process.

International Journal of Forecasting 32 (2016) 896–913

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

ELSEVIER

2016

Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond

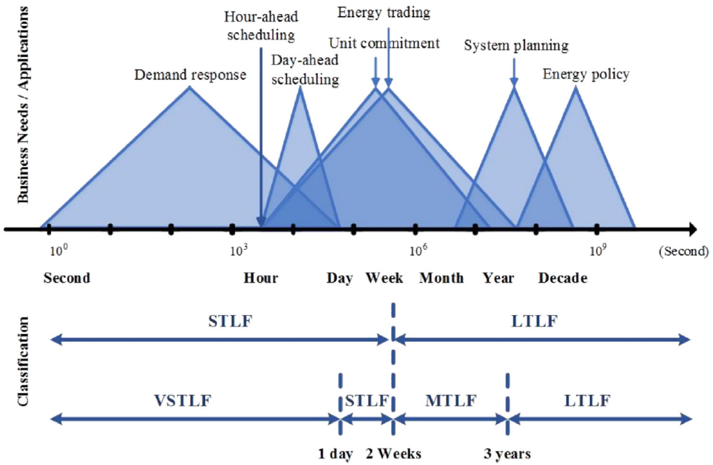
CrossMark

Tao Hong [a,*], Pierre Pinson [b], Shu Fan [c], Hamidreza Zareipour [d], Alberto Troccoli [e], Rob J. Hyndman [c]

[a] University of North Carolina at Charlotte, USA
[b] Technical University of Denmark, Lyngby, Denmark
[c] Monash Business School, Monash University, Clayton, Australia
[d] University of Calgary, Alberta, Canada
[e] World Energy & Meteorology Council, UK

Load forecasting applications and classification

(Hong & Fan, 2016, IJF)



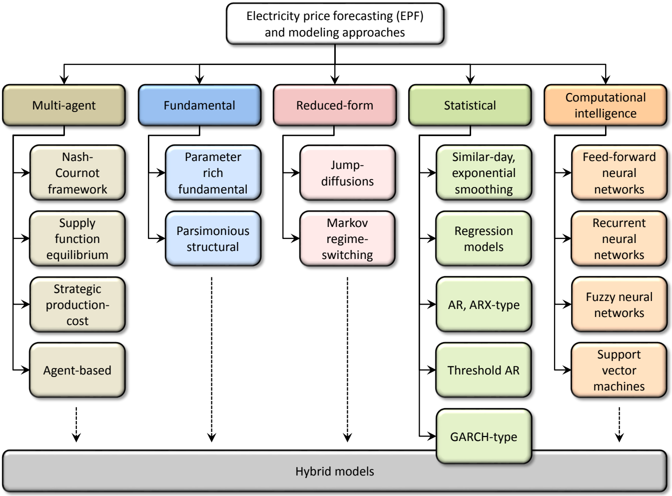
Electricity price forecasting horizons

(Weron, 2014, IJF)

- Short-term
  - From a few minutes up to a few days ahead
  - Of prime importance in day-to-day market operations
- Medium-term
  - From a few days to a few months ahead
  - Balance sheet calculations, risk management, derivatives pricing
  - Inflow of 'finance solutions'
- Long-term
  - Lead times measured in months, quarters or even in years
  - Investment profitability analysis and planning
  - Beyond the scope of this review

A taxonomy of (price) modeling approaches

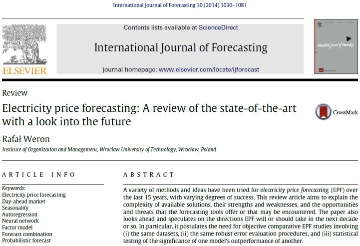
(Weron, 2014, IJF)



A look into the future of load/price forecasting

(Weron, 2014, IJF)

1. Modeling and forecasting the trend-seasonal components
2. Beyond point forecasts – probabilistic forecasts
3. Combining forecasts
   - Point forecasts
   - Probabilistic forecasts
4. Multivariate factor models
5. Guidelines for evaluating forecasts


International Journal of Forecasting 30 (2014) 1030–1081

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Review

Electricity price forecasting: A review of the state-of-the-art with a look into the future

Rafał Weron

Institute of Organization and Management, Wrocław University of Technology, Wrocław, Poland

ARTICLE INFO

Keywords:
Electricity price forecasting
Day-ahead market
Seasonality
Autoregression
Neural network
Factor model
Forecast combination
Probabilistic forecast

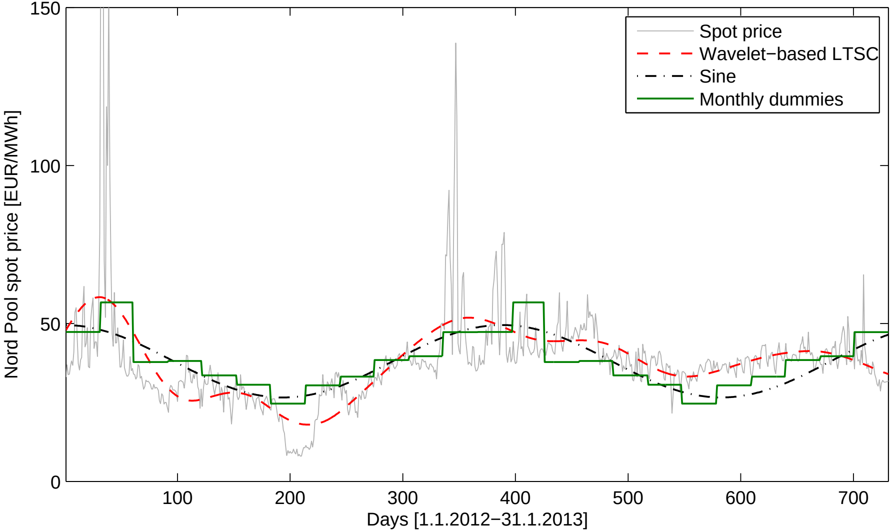
ABSTRACT

A variety of methods and ideas have been tried for *electricity price forecasting* (EPF) over the last 15 years, with varying degrees of success. This review article aims to explain the complexity of available solutions, their strengths and weaknesses, and the opportunities and threats that the forecasting tools offer or that may be encountered. The paper also looks ahead and speculates on the directions EPF will or should take in the next decade or so. In particular, it postulates the need for objective comparative EPF studies involving (i) the same datasets, (ii) the same robust error evaluation procedures, and (iii) statistical testing of the significance of one model's outperformance of another.

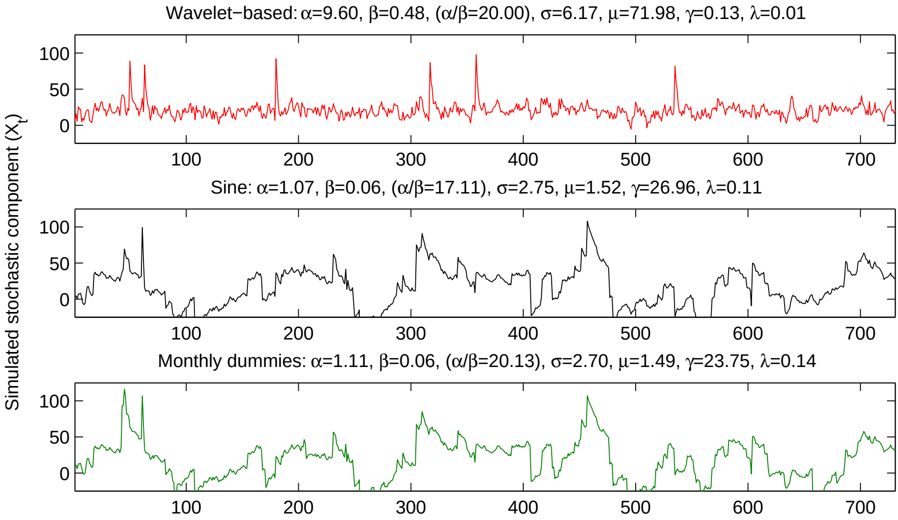
Modeling the trend-seasonal components

- Standard approach – decompose a time series of prices P_t into
  - the long-term trend-seasonal component (LTSC) T_t,
  - the short-term seasonal component (STSC) s_t,
  - and the remaining variability, error or stochastic component X_t
- The hourly/weekly STSC is usually captured by autoregression & dummies $\rightarrow$ forecasting is straightforward
- Annual seasonality is present in spot prices, but in most cases the LTSC is dominated by a more irregular cyclic component
  - Due to fuel prices, economic growth, long-term weather trends
  - See e.g. Janczura et al. (2013, ENEECO), Nowotarski, Tomczyk & Weron (2013, ENEECO)

Modeling the LTSC



Adequate seasonal decomposition is important !



Case study I

Energy Economics 39 (2013) 13–27

Contents lists available at SciVerse ScienceDirect

Energy Economics

journal homepage: www.elsevier.com/locate/eneco

Robust estimation and forecasting of the long-term seasonal component of electricity spot prices☆

Jakub Nowotarski [a,b], Jakub Tomczyk [a,b], Rafał Weron [b,*]

[a] Hugo Steinhaus Center, Institute of Mathematics and Computer Science, Wrocław University of Technology, 50–370 Wrocław, Poland
[b] Institute of Organization and Management, Wrocław University of Technology, 50–370 Wrocław, Poland

ARTICLE INFO

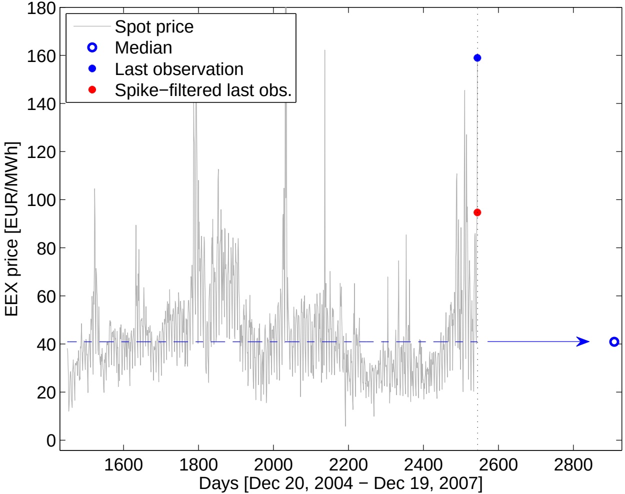
Article history:
Received 11 November 2012
Received in revised form 6 April 2013
Accepted 7 April 2013
Available online 15 April 2013

JEL classification:
C14

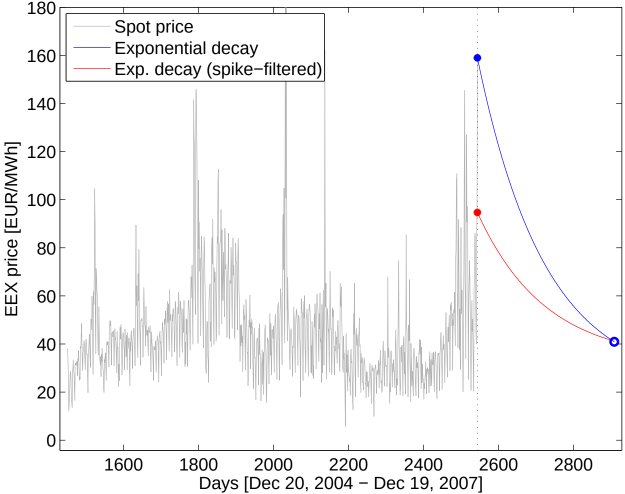
ABSTRACT

We present the results of an extensive study on estimation and forecasting of the long-term seasonal component (LTSC) of electricity spot prices. We consider a battery of over 300 models, including monthly dummies and models based on Fourier or wavelet decomposition combined with linear or exponential decay. We find that the considered wavelet-based models are significantly better in terms of forecasting spot prices up to a year ahead than the commonly used monthly dummies and sine-based models. This result questions the validity and usefulness of stochastic models of spot electricity prices built on the latter two types of LTSC models.

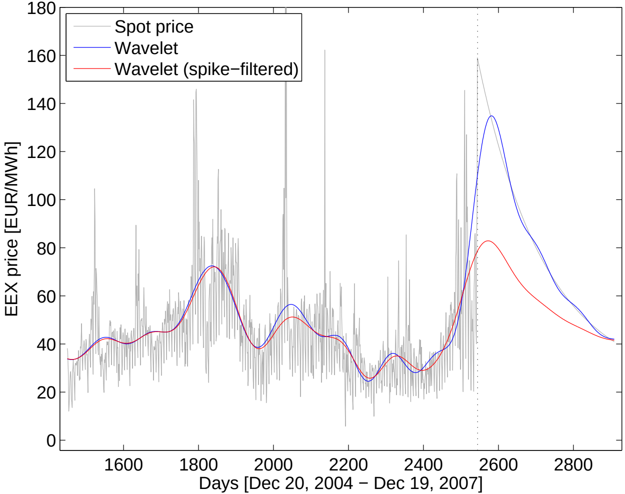
Forecasting a wavelet-based LTSC



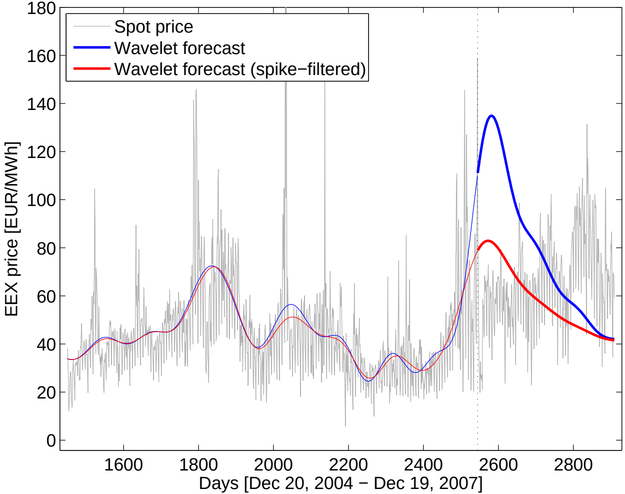
Forecasting a wavelet-based LTSC cont.



Forecasting a wavelet-based LTSC cont.

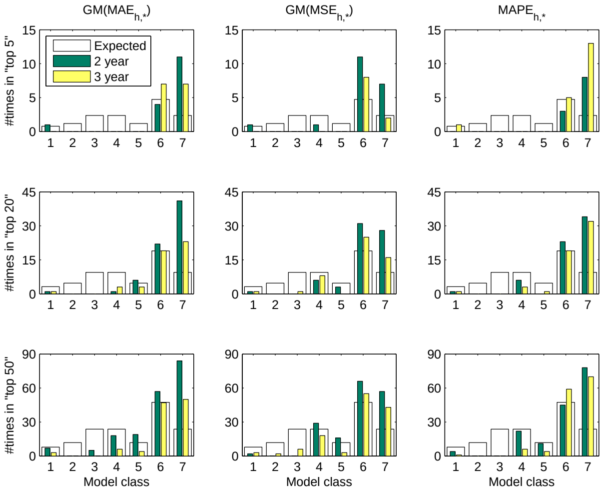


Forecasting a wavelet-based LTSC cont.



Wavelets beat sines and monthly dummies

The number of times models from a given family are ranked in the top 5, 20 and 50 of all 304 models according to $GM(MAE_{h,*})$, $GM(MSE_{h,*})$ and $MAPE_{h,*}$ for each of the six forecast horizons h = 1,...,6



Case study II

Energy Economics 48 (2015) 1–6

Contents lists available at ScienceDirect

Energy Economics

journal homepage: www.elsevier.com/locate/eneco

A note on using the Hodrick–Prescott filter in electricity markets

Rafał Weron *, Michał Zator

Department of Operations Research, Wrocław University of Technology, 50-370 Wrocław, Poland

A R T I C L E   I N F O

Article history:
Received 22 March 2014
Received in revised form 13 November 2014
Accepted 18 November 2014
Available online 1 December 2014

JEL classification:
C14
C51

A B S T R A C T

Recently, Nowotarski et al. (2013) have found that wavelet-based models for the long-term seasonal component (LTSC) are not only better in extracting the LTSC from a series of spot electricity prices but also significantly more accurate in terms of forecasting these prices up to a year ahead than the commonly used monthly dummies and sine-based models. However, a clear disadvantage of the wavelet-based approach is the increased complexity of the technique, as compared to the other two classes of LTSC models, and the resulting need for dedicated numerical software, which may not be readily available to practitioners in their work environments. To facilitate this problem, we propose here a much simpler, yet equally powerful method for identifying the LTSC in electricity spot price series. It makes use of the Hodrick–Prescott (HP) filter, a widely-recognized tool in macroeconomics.

The Hodrick-Prescott (1980, 1997) filter

A simple alternative to wavelets

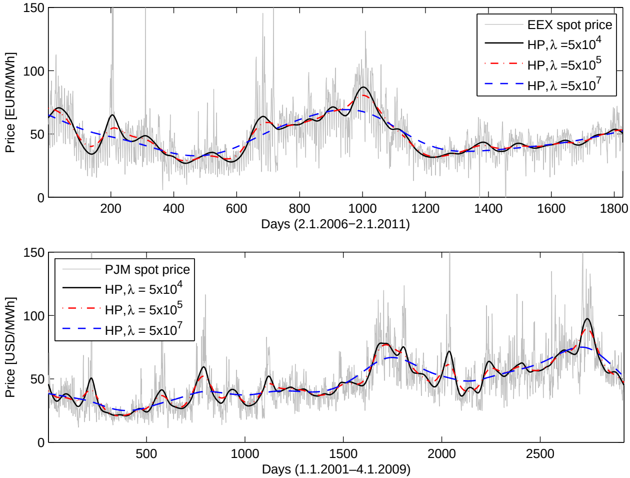
- Originally proposed for decomposing GDP into a long-term growth component and a cyclical component
- Returns a smoothed series τ_t for a noisy input series y_t:

$$\min_{\tau_t} \left\{ \sum_{t=1}^{T} (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} \Big[(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1}) \Big]^2 \right\},$$

Punish for:

- deviating from the original series
- roughness of the smoothed series

HP-smoothing for EEX and PJM spot prices



HP provides a better fit than the nominal LTSC

	Identification technique (estimated LTSC model)											
	HP filter-based ($\lambda = \dots$)							Wavelet-based				sin-EWMA
	5x10^4	10^5	5x10^5	10^6	5x10^6	10^7	5x10^7	S_5	S_6	S_7	S_8	EWMA
Nord Pool market (3 years: 01.01.2011–31.12.2013)												
S_5	18.7	33.5	60.0	68.0	89.4	103.1	138.6	0.0	73.4	83.6	174.7	92.1
S_6	19.3	11.2	0.0	1.2	23.8	42.0	87.8	49.1	10.8	16.0	134.9	43.6
S_7	38.9	29.4	10.0	4.3	12.1	29.1	81.4	71.8	29.5	0.0	140.0	34.2
S_8	92.8	81.4	56.0	45.8	24.4	16.1	0.0	132.3	83.7	40.6	10.0	55.6
sin	22.8	16.2	3.7	0.0	3.1	13.4	50.4	48.7	19.5	2.2	97.5	11.6
EEX market (5 years: 02.01.2006–02.01.2011)												
S_5	5.1	16.4	51.5	65.8	91.5	102.5	130.9	0.0	59.3	105.3	155.7	106.1
S_6	5.3	0.0	10.4	23.1	55.5	72.4	115.2	37.7	0.2	75.8	148.4	90.6
S_7	40.4	29.1	6.9	0.0	3.6	20.4	77.9	84.8	35.1	4.4	118.3	78.0
S_8	81.2	67.3	38.9	28.3	7.1	0.0	1.1	134.6	72.1	28.2	2.6	87.3
sin	10.0	4.2	0.0	2.1	13.3	22.7	52.7	41.2	14.3	17.7	75.5	47.6
PJM market (8 years: 01.01.2001–04.01.2009)												
S_5	0.0	6.9	32.1	44.1	68.9	77.9	98.0	4.1	37.0	79.8	106.3	79.7
S_6	7.2	0.8	3.7	14.6	46.0	58.4	86.2	38.9	0.0	61.6	101.1	71.9
S_7	46.4	34.8	11.7	4.2	1.6	10.2	46.5	91.0	37.5	0.0	66.1	63.4
S_8	99.4	83.8	52.1	40.1	16.4	8.7	7.0	158.9	87.2	35.8	0.0	85.4
sin	12.7	6.7	0.0	0.6	8.5	14.5	34.5	43.2	15.0	16.4	38.1	38.7

Case study III

Energy Economics 57 (2016) 228–235

Contents lists available at ScienceDirect

Energy Economics

journal homepage: www.elsevier.com/locate/eneco

On the importance of the long-term seasonal component in day-ahead electricity price forecasting

Jakub Nowotarski, Rafał Weron*

Department of Operations Research, Wrocław University of Technology, Wrocław, Poland

ARTICLE INFO

Article history:
Received 16 March 2016
Received in revised form 21 May 2016
Accepted 25 May 2016
Available online 2 June 2016

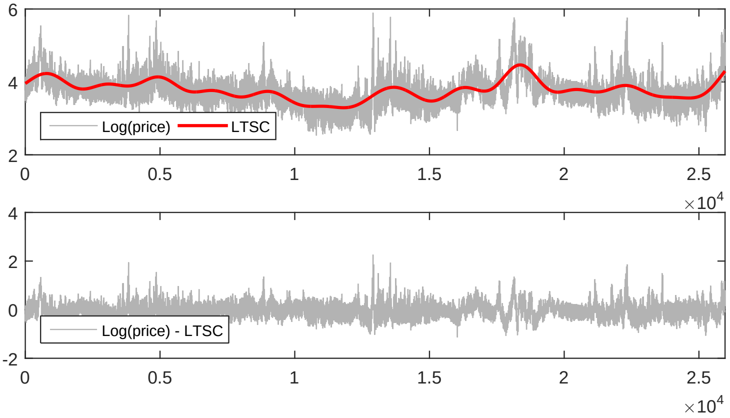
JEL classification:

ABSTRACT

In day-ahead *electricity price forecasting* (EPF) the daily and weekly seasonalities are always taken into account, but the long-term seasonal component (LTSC) is believed to add unnecessary complexity to the already parameter-rich models and is generally ignored. Conducting an extensive empirical study involving state-of-the-art time series models we show that (i) decomposing a series of electricity prices into a LTSC and a stochastic component, (ii) modeling them independently and (iii) combining their forecasts can bring – contrary to a common belief – an accuracy gain compared to an approach in which a given time series model is calibrated to the prices themselves.

LTSC and short-term price forecasting

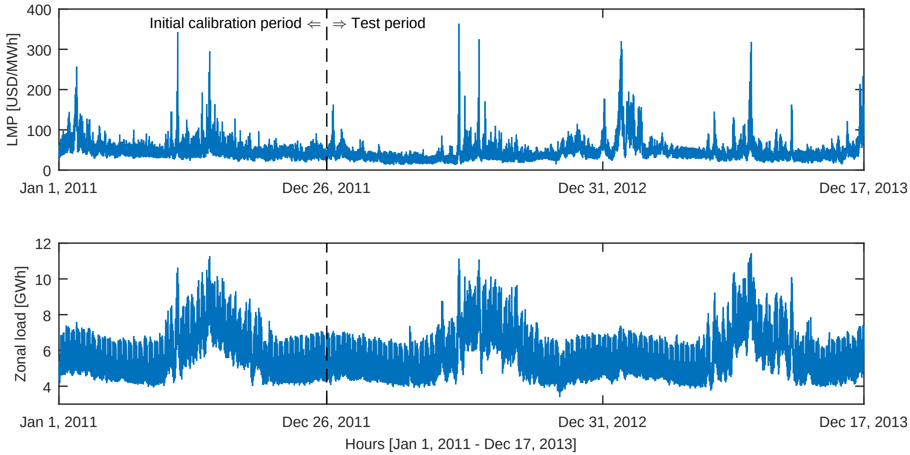
- Can the long-term trend-seasonal component impact short-term (day-ahead) electricity price forecasts?



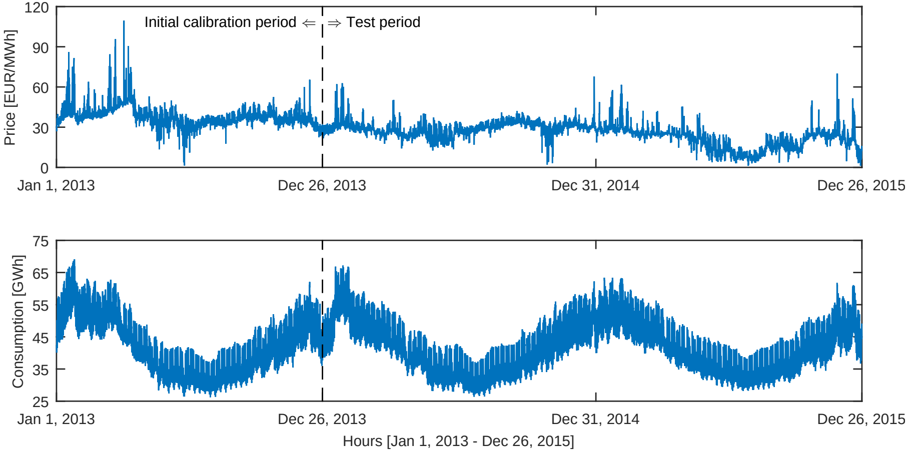
The setup

- Two 2-year long, hourly test periods
  - GEFCom2014
  - Nord Pool
- Two autoregressive model structures for day-ahead forecasting
- Two well-performing LTSC model classes
  - Wavelets
  - The Hodrick-Prescott filter
- Two models combining 24 hour-ahead extrapolation of an estimated LTSC with the forecasts of autoregressive models

The data: GEFCom2014



The data: Nord Pool (2013-2015)



The benchmarks: Naïve and ARX

- The **naïve** (or persistent) benchmark
  - $\widehat{P}_{t+h|t} = P_{t+h-24}$ for Tuesday to Friday
  - $\widehat{P}_{t+h|t} = P_{t+h-168}$ for Saturday to Monday
- **ARX** for the log-price $p_t = \log P_t$, originally proposed by Misiorek et al. (2006, SNDE):

$$p_t = \phi_1 p_{t-24} + \phi_2 p_{t-48} + \phi_7 p_{t-168} + \phi_8 mp_t + \psi_1 z_t + \sum_{i=1}^{3} d_i D_i + \varepsilon_t$$

  - mp_t is the minimum of the previous day's 24 hourly log-prices
  - D_1, D_2, D_3 are dummies for Monday, Saturday and Sunday

The benchmarks: mARX

- Multi-day ARX (**mARX**), an extension of ARX used in GEFCom2014 by Maciejowska & Nowotarski (2016, IJF):

$$
\begin{aligned}
p_t \;=\; & \left(\sum_{i=0}^{3} \phi_{1,i} D_i\right) p_{t-24} + \phi_2 p_{t-48} + \phi_3 D_1 p_{t-72} + \phi_7 p_{t-168} \\
& + \phi_8 m p_t + \psi_1 z_t + \sum_{i=1}^{3} d_i D_i + \varepsilon_t
\end{aligned}
$$

  - Uses different model structures for different days of the week, not only different parameter sets
  - $D_0 \equiv 1$ and $D_1 p_{t-72}$ accounts for the autoregressive effect of Friday's prices on the prices for the same hour on Monday

The Seasonal Component AR (SCAR) model

1. Decompose log-prices p_t into a LTSC, T_t, and a stochastic component with short-term (weekly) periodicities, X_t
2. Model X_t using one of the ARX or mARX models
3. Model T_t:
   - Using one of the 10 wavelet smoothers ($S_5, \ldots, S_{14}$) or one of the 8 HP filters ($\lambda = 1 \times 10^8, \ldots, 5 \times 10^{11}$)
   - Compute a persistent day-ahead prediction: $\hat{T}_{t^*+1} \equiv T_{t^*-23}$, ..., $\hat{T}_{t^*+24} \equiv T_{t^*}$, where t^* is the time index of the last observation in the calibration window
4. Compute SCAR forecast:

$$\widehat{p}_{t+h|t} = \widehat{T}_{t+h|t} + \widehat{X}_{t+h|t} = T_t + \widehat{X}^{ARX}_{t+h|t}$$

GEFCom2014: Average WMAE over all weeks

GEFCom2014

				SCARX					
				Wavelet approximation					
S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}
13.530	13.686	12.466	11.558	11.378	11.264	11.263	**<u>11.112</u>**	**11.221**	11.245
				HP filter λ					
	1×10^8	5×10^8	1×10^9	5×10^9	1×10^{10}	5×10^{10}	1×10^{11}	5×10^{11}	
	11.775	11.586	11.527	11.425	11.396	11.376	11.362	11.287	
				mSCARX					
				Wavelet approximation					
S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}
13.482	13.647	12.233	11.379	**11.216**	**11.213**	11.312	**<u>10.901</u>**	**10.976**	**11.130**
				HP filter λ					
	1×10^8	5×10^8	1×10^9	5×10^9	1×10^{10}	5×10^{10}	1×10^{11}	5×10^{11}	
	11.580	11.414	11.369	11.347	11.381	11.548	11.612	11.598	
				Benchmarks					
Naïve	**ARX**	**mARX**							
14.716	11.232	11.252							

WMAE errors smaller (better) than those of the ARX and mARX benchmarks are in **bold**. <u>Underlined</u> are the results for the best performing model in each part of the table.

Note, that in Nowotarski & Weron (2016) the WMAE error for the Naïve benchmark was mistakenly given as 20.475. With the correct value of WMAE the Naïve benchmark is still much worse than any of the other models

Nord Pool: Average WMAE over all weeks

Nord Pool

SCARX									
Wavelet approximation									
S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}
9.949	9.988	8.598	**8.389**	**8.309**	**8.332**	**8.417**	**8.453**	**8.463**	**8.475**
HP filter λ									
	1×10^8	5×10^8	1×10^9	5×10^9	1×10^{10}	5×10^{10}	1×10^{11}	5×10^{11}	
	8.665	8.697	8.718	8.760	8.766	8.766	8.757	8.729	
mSCARX									
Wavelet approximation									
S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}
9.954	10.049	8.558	8.286	**8.157**	**8.154**	**8.331**	8.471	8.428	8.361
HP filter λ									
	1×10^8	5×10^8	1×10^9	5×10^9	1×10^{10}	5×10^{10}	1×10^{11}	5×10^{11}	
	8.516	8.504	8.513	8.526	8.530	8.561	8.578	8.644	
Benchmarks									
Naïve	**ARX**	**mARX**							
9.661	8.500	8.341							

WMAE errors smaller (better) than those of the ARX and mARX benchmarks are in **bold**. Underlined are the results for the best performing model in each part of the table.

The Naïve benchmark is better than the (m)SCARX models with the most volatile LTSCs (S_5, S_6), but much worse than any of the other models. Note, that in Nowotarski & Weron (2016) the WMAE error for the Naïve benchmark was mistakenly given as 12.663

Diebold-Mariano test

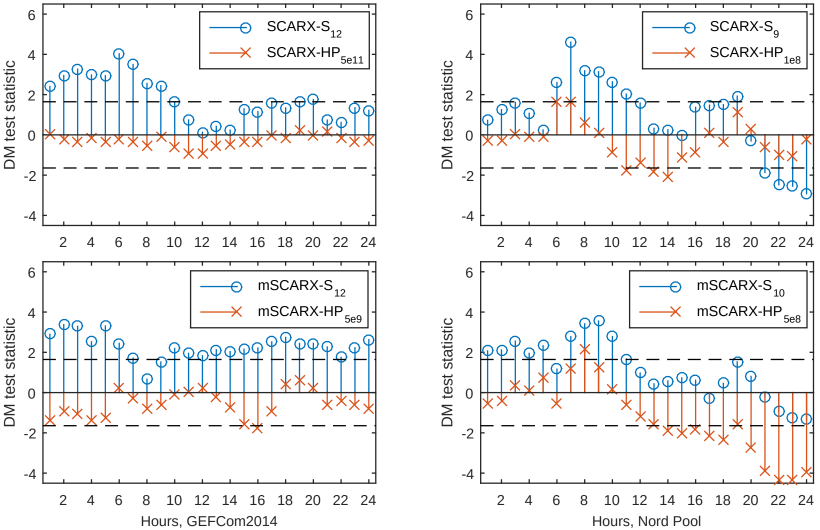
- DM tests using absolute errors of the model forecast:

$$L(\varepsilon_t) = |\varepsilon_t| = |P_t - \hat{P}_t|$$

- For each best performing SCAR-type model in its class:
    - **SCARX-S**$_{12}$, **-HP**$_{5\times10^{11}}$, **mSCARX-S**$_{12}$, **-HP**$_{5\times10^{9}}$ for GEFCom2014,
    - **SCARX-S**$_{9}$, **-HP**$_{1\times10^{8}}$, **mSCARX-S**$_{10}$, **-HP**$_{5\times10^{8}}$ for Nord Pool,
- For each hour independently calculate the loss differential series:

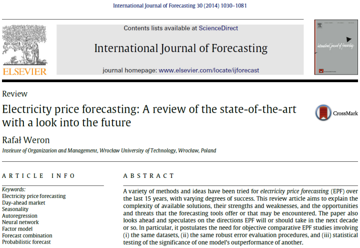
$$d_t = L(\varepsilon_t^{model}) - L(\varepsilon_t^{benchmark})$$

DM test vs. the (m)ARX benchmarks



A look into the future of load/price forecasting

1. Modeling and forecasting the trend-seasonal components
2. Beyond point forecasts – probabilistic forecasts
3. Combining forecasts
   - Point forecasts
   - Probabilistic forecasts
4. Multivariate factor models
5. Guidelines for evaluating forecasts



International Journal of Forecasting 30 (2014) 1030–1081

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

ELSEVIER

Review

Electricity price forecasting: A review of the state-of-the-art with a look into the future

Rafał Weron

Institute of Organization and Management, Wrocław University of Technology, Wrocław, Poland

ARTICLE INFO

Keywords:
Electricity price forecasting
Day-ahead market
Seasonality
Autoregression
Neural network
Factor model
Forecast combination
Probabilistic forecast

ABSTRACT

A variety of methods and ideas have been tried for *electricity price forecasting* (EPF) over the last 15 years, with varying degrees of success. This review article aims to explain the complexity of available solutions, their strengths and weaknesses, and the opportunities and threats that the forecasting tools offer or that may be encountered. The paper also looks ahead and speculates on the directions EPF will or should take in the next decade or so. In particular, it postulates the need for objective comparative EPF studies involving (i) the same datasets, (ii) the same robust error evaluation procedures, and (iii) statistical testing of the significance of one model's outperformance of another.

Beyond point forecasts

- Variability of the electricity demand becoming a challenge to the utility industry in the smart grid era
- Extreme variability of electricity prices
- Ability to plan different strategies for the range of possible outcomes indicated by the probabilistic forecast
- Useful in practice → risk management and decision-making

- GEFCom2012 (point) → GEFCom2014 (probabilistic forecasts)

GEFCom2012

- Two tracks
- Participants
  - 2000+ entries
  - 200+ teams
  - 30+ countries

GEFCom2014

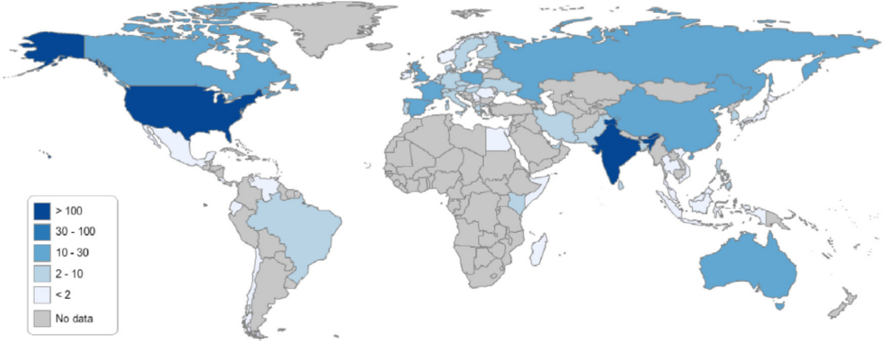
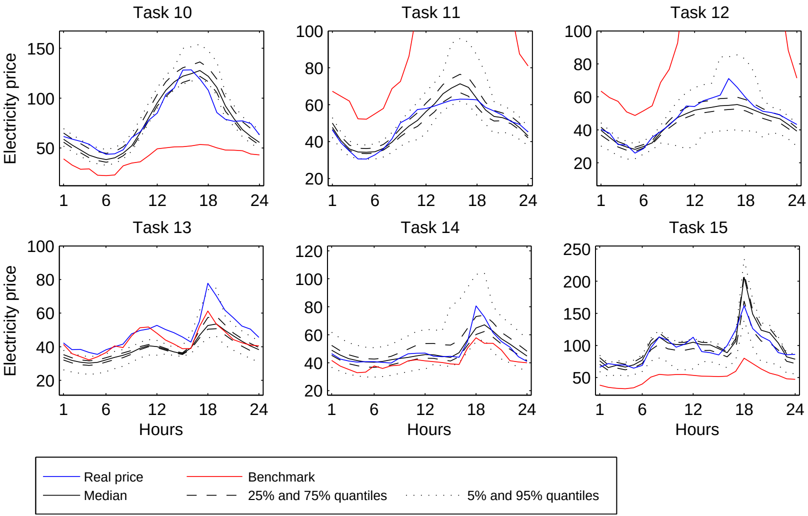


Fig. 1. Geographic distribution of GEFCom2014 contestants (581 people from 61 countries).

GEFCom2014

- Incremental data sets released on weekly basis
- Price Track:
  - 287 contestants
  - Submit 99 quantiles for 24h load periods of the next day

GEFCom2014 Price Track



GEFCom2014 Price Track

Top winning teams

1. Pierre Gaillard, Yannig Goude, Raphaël Nedellec (EDF R&D, F)
2. Katarzyna Maciejowska, Jakub Nowotarski (Wrocław UT, PL)
3. Grzegorz Dudek (Częstochowa UT, PL)
4. Zico Kolter, Romain Juban, Henrik Ohlsson, Mehdi Maasoumy (C3 Energy, USA)
5. Frank Lemke (KnowledgeMiner Software, D)

GEFCom2014 Price Track: (1st and) 2nd place for QRA!

A look into the future of load/price forecasting

1. Modeling and forecasting the trend-seasonal components
2. Beyond point forecasts – probabilistic forecasts
3. Combining forecasts
   - Point forecasts
   - Probabilistic forecasts
4. Multivariate factor models
5. Guidelines for evaluating forecasts

International Journal of Forecasting 30 (2014) 1030–1081

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

ELSEVIER

Review

Electricity price forecasting: A review of the state-of-the-art with a look into the future

Rafał Weron

Institute of Organization and Management, Wrocław University of Technology, Wrocław, Poland

ARTICLE INFO

Keywords:
Electricity price forecasting
Day-ahead market
Seasonality
Autoregression
Neural network
Factor model
Forecast combination
Probabilistic forecast

ABSTRACT

A variety of methods and ideas have been tried for *electricity price forecasting* (EPF) over the last 15 years, with varying degrees of success. This review article aims to explain the complexity of available solutions, their strengths and weaknesses, and the opportunities and threats that the forecasting tools offer or that may be encountered. The paper also looks ahead and speculates on the directions EPF will or should take in the next decade or so. In particular, it postulates the need for objective comparative EPF studies involving (i) the same datasets, (ii) the same robust error evaluation procedures, and (iii) statistical testing of the significance of one model's outperformance of another.

Forecast combinations, forecast/model averaging

- The idea goes back to the 1960s to the seminal papers of Bates and Granger (1969) and Crane and Crotty (1967)
- In electricity markets:
  - Electricity demand or transmission congestion forecasting (Bunn, 1985a; Bunn and Farmer, 1985; Løland et al., 2012; Smith, 1989; Taylor, 2010; Taylor and Majithia, 2000)
  - Only recently applied in the context of electricity price forecasting (EPF): Bordignon et al. (2013, ENEECO), Nowotarski et al. (2014, ENEECO), Weron (2014, IJF) and Raviv et al. (2015, ENEECO)

See also

Argo   Issue n. 9 - Winter 2016   Energisk.org

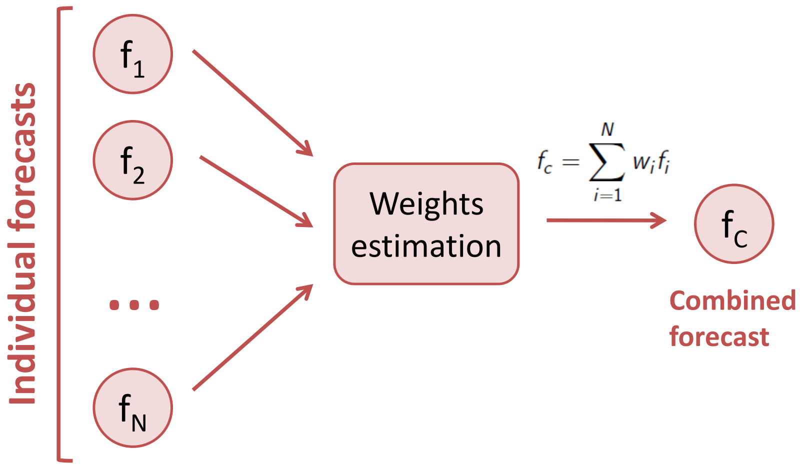
To Combine or not to Combine?

Recent Trends in Electricity Price Forecasting

Jakub NOWOTARSKI
Rafał WERON

Essentially everyone agrees nowadays that electricity spot price forecasting is of prime importance to the energy business. A variety of methods and ideas have been tried over the years, with varying degrees of success. Yet, despite this diversity of models, it is impossible to select one single, most reliable approach. In this article[1] the authors argue that *combining forecasts* – also known as *averaging forecasts*, *aggregating experts*, *committee machines* or *ensemble averaging* – is an idea worth considering. Using publicly available data from the Global Energy Forecasting Competition 2014 and four commonly used time series models, they show that for both point and probabilistic forecasts the quality of predictions can be improved if combined.

Point forecast averaging: The idea



Case study IV

International Journal of Forecasting 30 (2014) 1030–1081

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Review

Electricity price forecasting: A review of the state-of-the-art with a look into the future

Rafał Weron

Institute of Organization and Management, Wrocław University of Technology, Wrocław, Poland

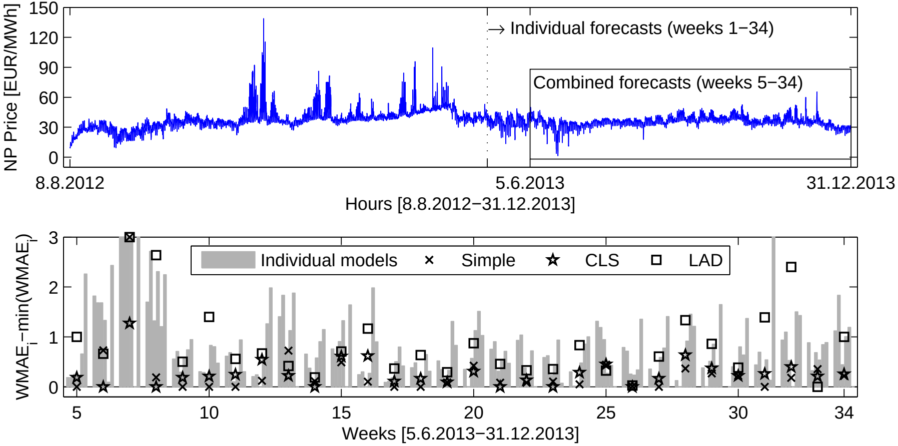
A R T I C L E   I N F O

Keywords:
Electricity price forecasting
Day-ahead market
Seasonality
Autoregression
Neural network
Factor model
Forecast combination
Probabilistic forecast

A B S T R A C T

A variety of methods and ideas have been tried for *electricity price forecasting* (EPF) over the last 15 years, with varying degrees of success. This review article aims to explain the complexity of available solutions, their strengths and weaknesses, and the opportunities and threats that the forecasting tools offer or that may be encountered. The paper also looks ahead and speculates on the directions EPF will or should take in the next decade or so. In particular, it postulates the need for objective comparative EPF studies involving (i) the same datasets, (ii) the same robust error evaluation procedures, and (iii) statistical testing of the significance of one model's outperformance of another.

Combining price forecasts



Summary of results

Summary statistics for 6 individual and 3 averaging methods: $\overline{\mathrm{WMAE}}$ is the mean value of WMAE for a given model (with standard deviation in parentheses), # *best* is the number of weeks a given averaging method performs best in terms of WMAE, and finally *m.d.f.b.* is the mean deviation from the best model in each week. The out-of-sample test period covers 30 weeks (5.6.2013–31.12.2013).

	Individual models						Forecast combinations		
	AR	TAR	SNAR	MRJD	NAR	FM	Simple	CLS	LAD
$\overline{\mathrm{WMAE}}$	5.03	5.07	4.77	4.98	4.88	5.36	4.47	**4.29**	4.92
	(3.40)	(3.53)	(3.26)	(3.17)	(1.62)	(3.17)	(2.87)	(1.88)	(2.41)
# *best*	1	3	4	1	2	4	**8**	6	1
m.d.f.b.	1.01	1.05	0.75	0.96	0.86	1.34	0.45	**0.27**	0.89

Interval forecast averaging

- For point forecasts: $f_c = \sum_{i=1}^{N} w_i f_i$ (e.g. a linear regression model)
- For interval forecasts the above formula does not hold
- A linear combination of q-th quantiles is **not** the q-th quantile of a linear combination of random variables

$$x_c^q \neq \sum_{i=1}^{N} w_i x_i^q$$

- $\rightarrow$ Need for development of new approaches

Quantile Regression Averaging (QRA) defined

Comput Stat (2015) 30:791–803
DOI 10.1007/s00180-014-0523-0

CrossMark

ORIGINAL PAPER

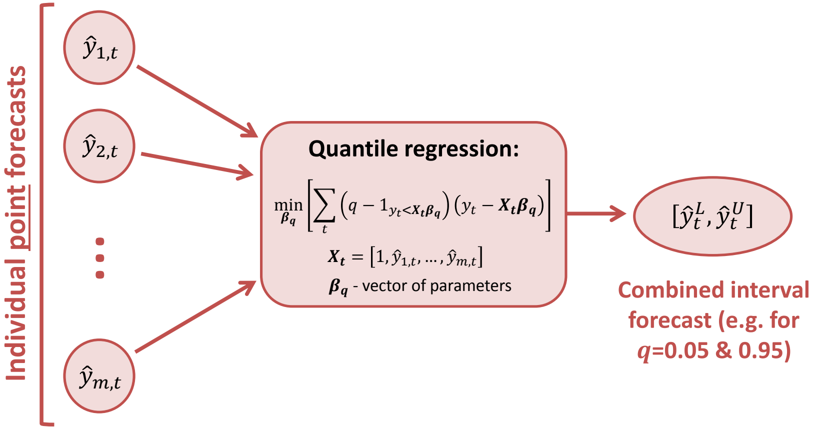
Computing electricity spot price prediction intervals using quantile regression and forecast averaging

Jakub Nowotarski · Rafał Weron

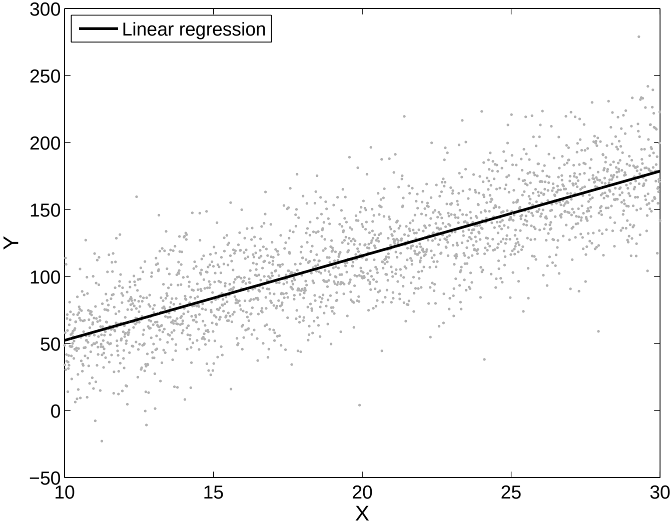
Received: 31 December 2013 / Accepted: 6 August 2014 / Published online: 19 August 2014

Abstract We examine possible accuracy gains from forecast averaging in the context of interval forecasts of electricity spot prices. First, we test whether constructing empirical prediction intervals (PI) from combined electricity spot price forecasts leads to better forecasts than those obtained from individual methods. Next, we propose a new method for constructing PI—Quantile Regression Averaging (QRA)—which utilizes the concept of quantile regression and a pool of point forecasts of individual (i.e. not combined) models. While the empirical PI from combined forecasts do not provide significant gains, the QRA-based PI are found to be more accurate than those of the best individual model—the smoothed nonparametric autoregressive model.

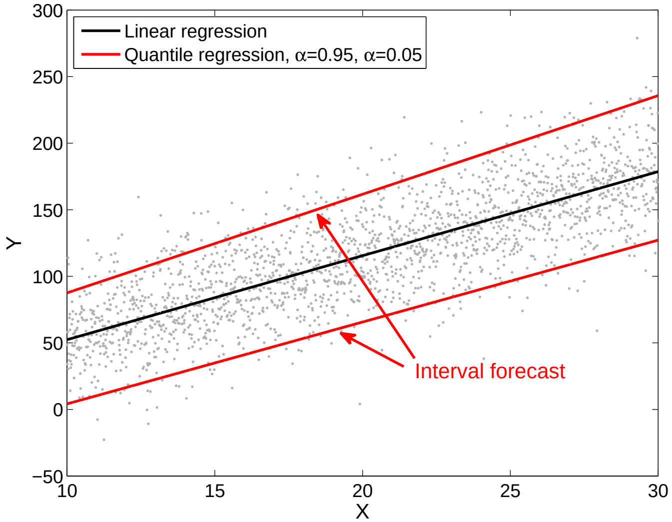
Quantile Regression Averaging: The idea



Quantile regression



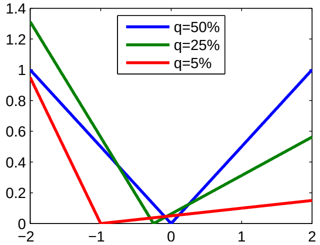
Quantile regression



How does the score function look like?

For vector $\boldsymbol{X_t} = [1, \hat{y}_{1,t}, ..., \hat{y}_{m,t}]$ of point forecasts, i.e. explanatory variables, weights $\boldsymbol{\beta_q}$ are estimated by minimizing:

$$\min_{\boldsymbol{\beta_q}} \left[\sum_{\{t: y_t \geq \boldsymbol{X_t}\boldsymbol{\beta_q}\}} q|y_t - \boldsymbol{X_t}\boldsymbol{\beta_q}| + \sum_{\{t: y_t < \boldsymbol{X_t}\boldsymbol{\beta_q}\}} (1-q)|y_t - \boldsymbol{X_t}\boldsymbol{\beta_q}| \right]$$



Case study V

978-1-4799-6095-8/14/$31.00 ©2014 IEEE

Merging quantile regression with forecast averaging to obtain more accurate interval forecasts of Nord Pool spot prices

Jakub Nowotarski
Institute of Organization and Management
Wrocław University of Technology
Wrocław, Poland
Email: jakub.nowotarski@pwr.wroc.pl

Rafał Weron
Institute of Organization and Management
Wrocław University of Technology
Wrocław, Poland
Email: rafal.weron@pwr.wroc.pl

***Abstract*—We evaluate a recently proposed method for constructing prediction intervals, which utilizes the concept of quantile regression (QR) and a pool of point forecasts of different time series models. We find that in terms of interval forecasting of Nord Pool day-ahead prices the new QR-based approach significantly outperforms prediction intervals obtained from standard, as well as, semi-parametric autoregressive time series models.**

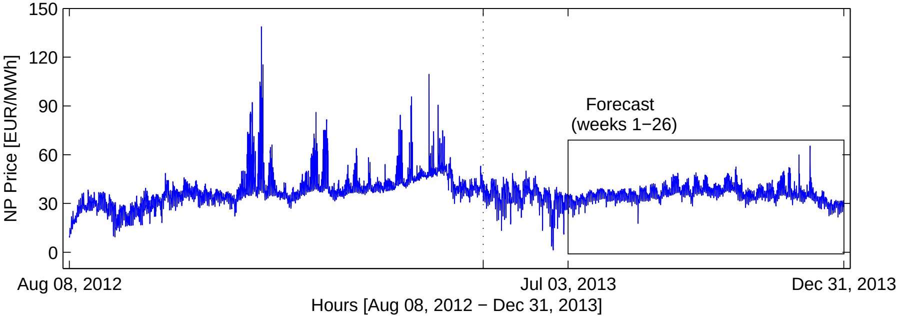
tions we are interested in PI, i.e. intervals which contain the true values of future observations with specified probability, not in confidence intervals.

From a practical point of view, PI provide additional information on price forecasts. High volatility and uncertainty of electricity price forecasts may frequently deviate from the true price levels. In fact, possible errors in point predictions

QRA at work

- Six individual point forecasting models:
  - Autoregression (AR)
  - Threshold AR (TAR)
  - Semi-parametric AR (SNAR)
  - Mean-reverting jump diffusion (MRJD)
  - Non-linear AR neural network (NAR)
  - Factor model (FM)

The data: Nord Pool (2012-2013)



- **Seven** months for calibration of individual models
- **Four** weeks for calibration of quantile regression
- **26** weeks for evaluation of interval forecasts

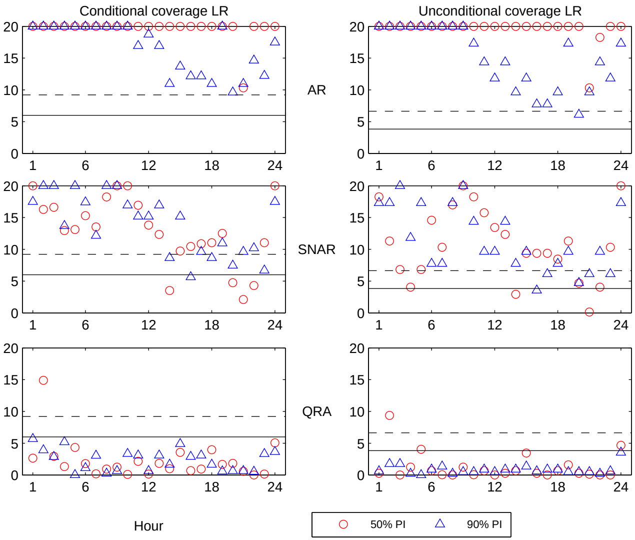
Evaluation of forecasts

- 50% and 90% two-sided day-ahead prediction intervals
- Two benchmark models: AR and SNAR
- Christoffersen's (1998, IER) test for unconditional and conditional coverage
- The focus on the sequence: $I_t = \begin{cases} 1 & y_t \in [\hat{y}_t^L, \hat{y}_t^U] \\ 0 & y_t \notin [\hat{y}_t^L, \hat{y}_t^U] \end{cases}$
- <u>Conditional Coverage test</u> (UC + independece) Asymptotically $\chi^2(2)$ | <u>Unconditional Coverage test</u> Asymptotically $\chi^2(1)$

Results: Unconditional coverage

PI	AR	SNAR	**QRA**
		Unconditional coverage	
50%	77.50	61.93	**49.77**
90%	97.53	96.41	**89.33**
		Mean width (STD of interval width)	
50%	4.55 (1.34)	2.76 (0.61)	**2.23 (0.81)**
90%	11.14 (3.31)	9.33 (2.45)	**6.78 (2.20)**

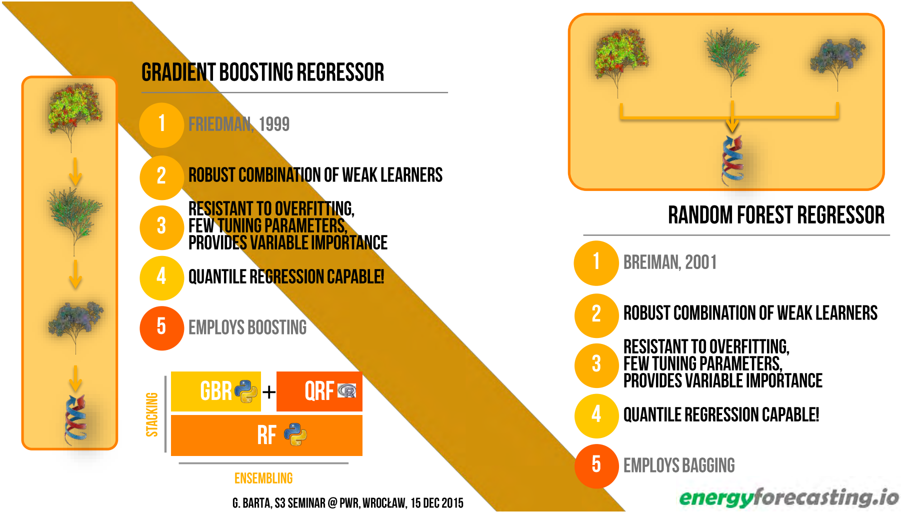
Results: Christoffersen's test



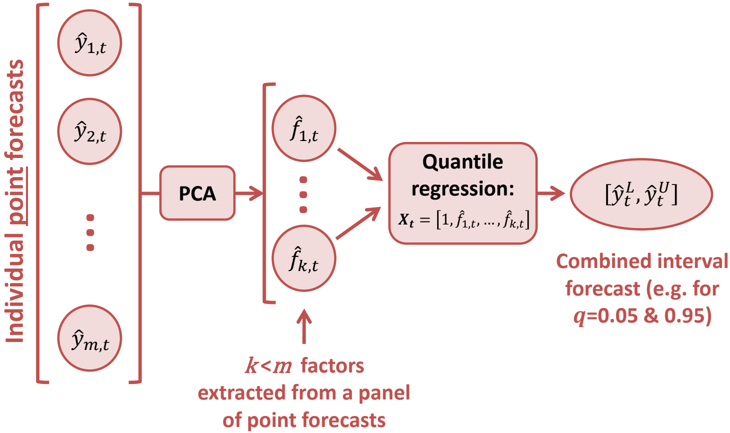
In the 'AI world' ...

- *Committee machines*, *ensemble averaging*, *expert aggregation*:
  - Guo and Luh (2004) combine a RBF network (23 inputs and six clusters) and a MLP (55 inputs and eight hidden neurons) to compute daily average on-peak electricity price for New England
  - Forecast combinations and committee machines seem to *evolve independently*, with researchers from both groups not being aware of the parallel developments !

GEFCom2014 Solar and Wind Tracks: 2nd place



FQRA: When the number of predictors is large



Case Study VI

International Journal of Forecasting 32 (2016) 957–965

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Probabilistic forecasting of electricity spot prices using Factor Quantile Regression Averaging

Katarzyna Maciejowska[a,b], Jakub Nowotarski[a], Rafał Weron[a,*]

[a] Department of Operations Research, Wrocław University of Technology, Wrocław, Poland
[b] CERGE-EI, Prague, Czech Republic

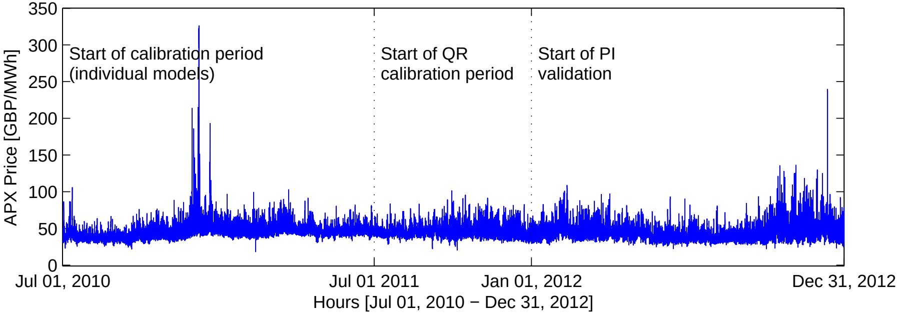
ARTICLE INFO

Keywords:
Probabilistic forecasting
Prediction interval
Quantile regression
Factor model
Forecasts combination
Electricity spot price

ABSTRACT

We examine possible accuracy gains from using factor models, quantile regression and forecast averaging to compute interval forecasts of electricity spot prices. We extend the Quantile Regression Averaging (QRA) approach of Nowotarski and Weron (2014a), and use principal component analysis to automate the process of selecting from among a large set of individual forecasting models that are available for averaging. We show that the resulting Factor Quantile Regression Averaging (FQRA) approach performs very well for price (and load) data from the British power market. In terms of unconditional coverage, conditional coverage and the Winkler score, we find the FQRA-implied prediction intervals to be more accurate than those of either the benchmark ARX model or the QRA approach.

FQRA in action



- **32** individual forecasting models
- **One year** for calibration of individual models
- **Half a year** for calibration of quantile regression
- **One year** for evaluation of interval forecasts

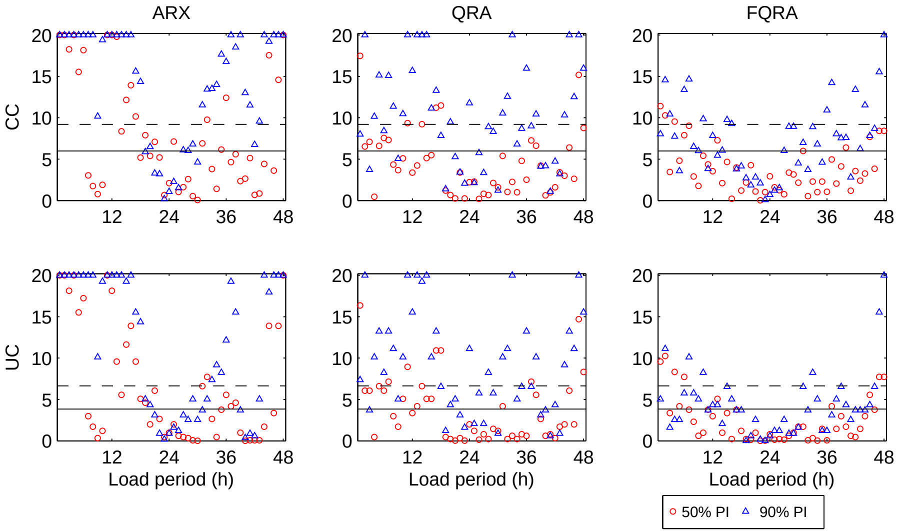
Evaluation of forecasts

- 50% and 90% two-sided day-ahead prediction intervals
- Three methods: QRA, FQRA and ARX (benchmark)
- Christoffersen's (1998) test for unconditional and conditional coverage
- Winkler score for a symmetric $(1-\alpha)\times 100\%$ prediction interval:

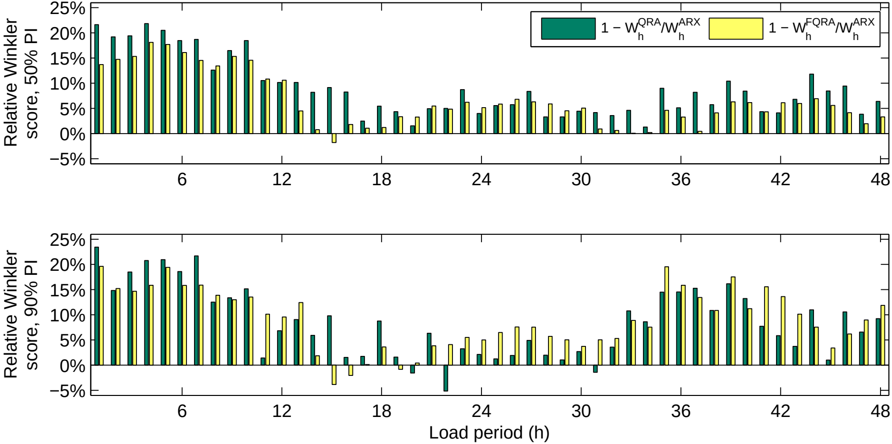
$$W_t = \begin{cases} \delta_t & \text{dla} \quad y_t \in [\hat{y}_t^L, \hat{y}_t^U], \\ \delta_t + \frac{2}{\alpha}(\hat{y}_t^L - y_t) & \text{dla} \quad y_t < \hat{y}_t^L, \\ \delta_t + \frac{2}{\alpha}(y_t - \hat{y}_t^U) & \text{dla} \quad y_t > \hat{y}_t^U, \end{cases}$$

where $\delta_t = \hat{y}_t^U - \hat{y}_t^L$ is the interval's width

Results: Christoffersen's test



Results: Winkler score



Bonus: Case Study VII

IEEE TRANSACTIONS ON SMART GRID 1

Probabilistic Load Forecasting via Quantile Regression Averaging on Sister Forecasts

Bidong Liu, Jakub Nowotarski, Tao Hong, and Rafał Weron

***Abstract*—The majority of the load forecasting literature has been on point forecasting, which provides the expected value for each step throughout the forecast horizon. In the smart grid era, the electricity demand is more active and less predictable than ever before. As a result, probabilistic load forecasting, which provides additional information on the variability and uncertainty of future load values, is becoming of great importance to power systems planning and operations. This paper proposes a practical methodology to generate probabilistic load forecasts by performing quantile regression averaging on a set of sister point forecasts. There are two major benefits of the proposed approach. It can leverage the development in the point load forecasting literature over the past several decades and it does not rely so much on high-quality expert forecasts, which are rarely achievable in load forecasting practice. To demonstrate the effectiveness of the proposed approach and make the results reproducible to the load forecasting community, we construct a case study using the publicly available data from the Global Energy Forecasting Competition 2014. Compared with several benchmark methods, the proposed approach leads to dominantly better performance as measured by the pinball loss function and the Winkler score.**

and planning of the power systems. Variability and uncertainty associated with the electricity demand is becoming a challenge to the utility industry. As a result, more and more decision making processes in the utility industry rely on probabilistic load forecasts. Typical applications of probabilistic load forecasting include stochastic unit commitment, probabilistic price forecasting, probabilistic transmission planning, and so forth [1], [2]. In the microgrid environment, probabilistic load forecasting is rather crucial, because the demand at individual household level or even distribution feeder level is quite volatile due to various demand response programs and feeder reconfiguration activities.

The load forecasting literature has focused on point forecasting, with researchers trying to forecast the expected value of future load using various techniques, primarily statistical techniques (such as regression models, exponential smoothing, and time series models), and artificial intelligence techniques (such as neural networks and support vector machines) [3]–[7].

Combining sister load forecasts

- Variable selection may be difficult in load forecasting
- **Sister models** – constructed by different subsets of variables with overlapping components
  - Here: 2 or 3 years for calibration and 4 ways of partitioning training and validation periods

$$\hat{y}_t = \beta_0 + \beta_1 M_t + \beta_2 W_t + \beta_3 H_t + \beta_4 W_t H_t + f(T_t) + \\ + \sum_d f(\tilde{T}_{t,d}) + \sum_{lag} f(T_{t-lag}),$$

- **Sister forecasts** are generated from sister models

Sister models

$$\hat{y}_t = \beta_0 + \overbrace{\beta_1 M_t + \beta_2 W_t + \beta_3 H_t + \beta_4 W_t H_t}^{\text{calendar effects}} + \overbrace{f(T_t)}^{\text{temp. dependence}} + \underbrace{+ \sum_d f(\tilde{T}_{t,d}) + \sum_{lag} f(T_{t-lag})}_{\text{recency effect}},$$

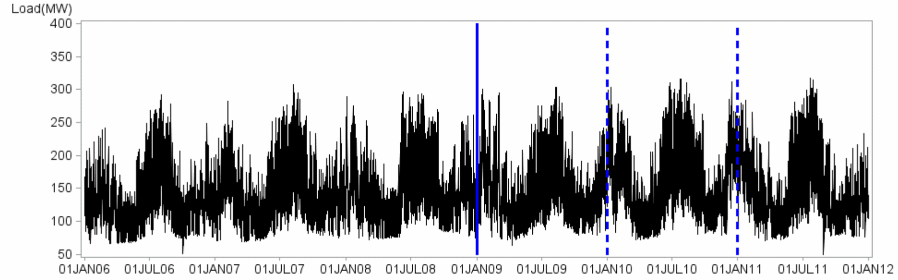
where:

$$f(T_t) = \beta_5 T_t + \beta_6 T_t^2 + \beta_7 T_t^3 + \beta_8 T_t M_t + \beta_9 T_t^2 M_t + \\ + \beta_{10} T_t^3 M_t + \beta_{11} T_t H_t + \beta_{12} T_t^2 H_t + \beta_{13} T_t^3 H_t$$

$$\tilde{T}_{t,d} = \frac{1}{24} \sum_{lag=24d-23}^{24d} T_{t-lag}$$

The data

(from the load forecasting track of GEFCom2014)



- **2 or 3 years** for calibration of sister (individual) models
- **1 year** for validation of sister (individual) models (variable selection)
- **1 year** for validation of probabilistic forecasts (best models selection)
- **1 year** for testing probabilistic forecasts

Benchmarks

- Two naive benchmarks
  - Scenario generation from historical weather data, no recency effect (Vanilla)
  - Quantiles interpolated from 8 individual forecasts (Direct)
- Benchmarks from individual models
  - 8 individual models (Ind) with residuals' distribution
  - Best Individual (BI) individual model according to MAE

Evaluation of forecasts

- Pinball loss function for 99 percentiles (as in GEFCom2014)

$$P_t = \begin{cases} (1-q)(\hat{y}_t^q - y_t), & y_t < \hat{y}_t^q \\ q(y_t - \hat{y}_t^q), & y_t \geq \hat{y}_t^q \end{cases}$$

- Winkler score for 50% and 90% two-sided day-ahead PI:

$$W_t = \begin{cases} \delta_t & \text{dla} \quad y_t \in [\hat{y}_t^L, \hat{y}_t^U], \\ \delta_t + \frac{2}{\alpha}(\hat{y}_t^L - y_t) & \text{dla} \quad y_t < \hat{y}_t^L, \\ \delta_t + \frac{2}{\alpha}(y_t - \hat{y}_t^U) & \text{dla} \quad y_t > \hat{y}_t^U, \end{cases}$$

where $\delta_t = \hat{y}_t^U - \hat{y}_t^L$ is the interval's width

Results: Test period

Model class	Pinball	Winkler (50%)	Winkler (90%)
QRA(8,183)	**2.85**	**25.04**	**55.85**
Ind(1,91)	3.22	26.35	56.38
BI(-,365)	3.00	26.38	57.17
Direct	3.19	26.62	94.27
Vanilla	8.00	70.51	150.0

- Sister forecasts easy to generate
- No need for independent expert forecasts
- Simple way to leverage from point to probabilistic forecasts

Where are we now?

(Hong et al., 2016, IJF)

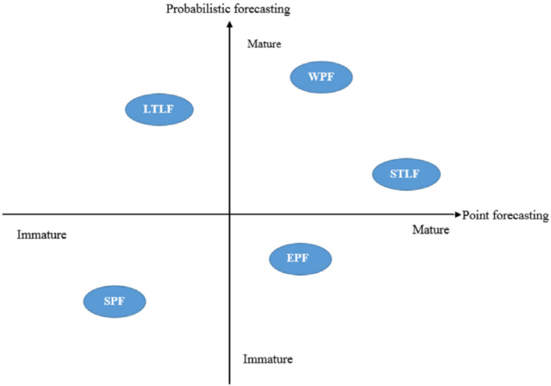


Fig. 12. Maturity quadrant of the energy forecasting subdomains (SPF: solar power forecasting; LTLF: long term load forecasting; EPF: electricity price forecasting; WPF: wind power forecasting; STLF: short term load forecasting).

Take-home message(s)

- Combining point forecasts is a robust technique, generally improving the performance
- The new trend is probabilistic forecasting
  - See: *Recent advances in electricity price forecasting: A review of probabilistic forecasting* (RePEc working paper)
- Combining interval (or density) forecasts is more tricky than combining point forecasts
- QRA is a simple way to leverage from point to probabilistic forecasts
- ... do not forget about the importance of getting the seasonal components right
- ... forecast evaluation is a critical issue