Importance of the long-term seasonal component in day-ahead electricity price forecasting: Regression vs. neural network models

Rafał Weron

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

*Based on a working paper with Grzegorz Marcjasz and Bartosz Uniejewski, available from RePEc: https://ideas.repec.org/p/wuu/wpaper/hsc1703.html
Markets for electricity in Europe

- N2EX (UK)
- EPEX Spot (AT, CH, DE, FR)
- OMIE (ES, PT)
- Nord Pool (DK, EST, FIN, NOR, SWE)
- APX-ENDEX (NL)
- PolPX (PL)
- OTE (CZ)
- Belpex (BE)
- EPEX Spot (AT, CH, DE, FR)
- HUPX (HU)
- PolPX (PL)
- OKTE (SK)
- GME (IT)
- Borzen (SLO)
- OPCOM (RO)
- EXAA (AT)

Rafał Weron (Wrocław, PL)
... in North America and Australia
Electricity price time series

Seasonality, mean-reversion and price spikes
The electricity ‘spot’ (day-ahead) price

- Day $D$
  - Bidding for day $D + 1$
- Day $D + 1$
  - Bidding for day $D + 2$
- Day $D + 2$
  - Bidding for day $D + 2$

24 hours of day $D + 1$

24 hours of day $D + 2$
Supply and demand, renewables and negative prices

Source: Ziel & Steinert (2016)
Prices for different load periods

Strongly correlated but seem to follow different data generating processes (DGPs)
First read on electricity price forecasting (EPF)
R.Hyndman: “this paper alone is responsible for 0.7 of the current IF$_2Y$ = 2.642” ;-)

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

Abstract

A variety of methods and ideas have been tried for electricity price forecasting over the last 15 years, with varying degrees of success. This review article examines the complexity of available solutions, their strengths and weaknesses, and threats that the forecasting tools offer or that may be encountered in the future. It looks ahead and speculates on the directions EPF will or should take or use. In particular, it postulates the need for objective comparative testing of the significance of one model's outperformance of another. An introduction to the IJF Hong award for Energy Forecasting 2013-2014

Rafał Weron (2014)

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

International Journal of Forecasting, 30(4), 1030-1081.
A look into the future of EPF

EPF directions in the next decade (according to Weron, 2014, IJF):

1. **Modeling and forecasting the trend-seasonal components**
2. Beyond point forecasts – probabilistic forecasts
3. Combining forecasts
4. Multivariate factor models
5. Guidelines for evaluating forecasts
Role of the long-term seasonal component (LTSC) for short-term EPF

- Significant prediction accuracy gains possible for linear regression models (Nowotarski & Weron, 2016, ENEECO):

<table>
<thead>
<tr>
<th>ARX</th>
<th>SCARX</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wavelet approximation</td>
</tr>
<tr>
<td></td>
<td>$S_5$</td>
</tr>
</tbody>
</table>

- Unknown effects for non-linear (e.g., ANN) models

- **Is this phenomenon more general?**
Agenda

- Introduction
  - Electricity markets and prices
  - Motivation
- Trend-seasonal components
  - Wavelets
  - The Hodrick-Prescott (HP) filter
- Case study
  - Datasets and LTSCs
  - ARX and SCARX models
  - ANNs in EPF
  - Committee machines of (SC)ANN networks
  - Results and conclusions
Wavelets

Decomposition of a signal
Wavelets

Decomposition of a signal

- Original signal
- Approximation 1 level
- Details 1 level
Wavelets

Decomposition of a signal

Original signal

Approximation 1 level

Approximation 7 level

Details 1 level

Details 7 level
Sample fits to Nord Pool data
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 $\tau_t$ for a noisy input series $y_t$:

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

Punish for:
- deviating from the original series
- roughness of the smoothed series
Sample fits to EEX and PJM data
(Weron & Zator, 2015, ENEECO)
Agenda

- Introduction
  - Electricity markets and prices
  - Motivation
- Trend-seasonal components
  - Wavelets
  - The Hodrick-Prescott (HP) filter
- Case study
  - Datasets and LTSCs
  - ARX and SCARX models
  - ANNs in EPF
  - Committee machines of (SC)ANN networks
  - Results and conclusions
Datasets: GEFCom 2014

Datasets are the same as in Nowotarski & Weron (2016, ENEECO)
Datasets: Nord Pool

Datasets are the same as in Nowotarski & Weron (2016, ENEECO)
Long-Term Seasonal Components (LTSCs)

Like in Nowotarski & Weron (2016, ENEECO), we consider 18 LTSCs from two categories:

- **Wavelet filters** $S_5, S_6, \ldots, S_{14}$, ranging from ‘daily’ smoothing ($S_5 \rightarrow 2^5$ hours) up to ‘biannual’ ($S_{14} \rightarrow 2^{14}$ hours)
  - Models with wavelet filters are denoted by suffixes $-S_J$

- **HP-filters** with $\lambda = 10^8, 5 \cdot 10^8, 10^9, \ldots, 5 \cdot 10^{11}$, also ranging from ‘daily’ up to ‘biannual’ smoothing
  - Models with HP filters are denoted by suffixes $-\text{HP}_\lambda$
Benchmark: The **ARX** model

For the log-price, i.e., $p_{d,h} = \log(P_{d,h})$, the model is given by:

$$
p_{d,h} = \beta_{h,1} p_{d-1,h} + \beta_{h,2} p_{d-2,h} + \beta_{h,3} p_{d-7,h} + \beta_{h,4} p_{d-1,min} + \beta_{h,5} z_t + \sum_{i=1}^{3} \beta_{h,i+5} D_i + \varepsilon_{d,h}
$$

(1)

- $p_{d-1,min}$ is yesterday’s minimum hourly price
- $z_t$ is the logarithm of system load/consumption
- Dummy variables $D_1$, $D_2$ and $D_3$ refer to Monday, Saturday and Sunday, respectively
The SCAR modeling framework
(Nowotarski & Weron, 2016, ENEECO)

The Seasonal Component AutoRegressive (SCAR) modeling framework consists of the following steps:

1. (a) Decompose the series in the calibration window into the LTSC $T_{d,h}$ and the stochastic component $q_{d,h}$
   (b) Decompose the exogenous series in the calibration window using the same type of LTSC as for prices
2. Calibrate the ARX model to $q_t$ and compute forecasts for the 24 hours of the next day (24 separate series)
The SCAR modeling framework cont.

3. Add stochastic component forecasts $\hat{q}_{d+1,h}$ to persistent forecasts $\hat{T}_{d+1,h}$ of the LTSC to yield log-price forecasts $\hat{p}_{d+1,h}$.

4. Convert them into price forecasts of the SCARX model, i.e.,

$$\hat{P}_{d+1,h} = \exp(\hat{p}_{d+1,h})$$
Sample LTSC and stochastic component forecasts
ANNs in other EPF studies

- Variety of ANN implementations, as well as considered inputs, making it impossible to compare with commonly used methods based on linear regression.
- Several studies that acknowledge the need of removing seasonal components from time series for neural network models:
  - Andrawis et al. (2011)
  - Zhang and Qi (2005)
  - Keles et al. (2016), the only one in the context of EPF.
One hidden layer with 5 neurons and sigmoid activation functions

Inputs identical as in the ARX model

Trained using Matlab’s trainlm function, utilizing the Levenberg-Marquardt algorithm for supervised learning
Seasonal Component ANN (SCANN)

The SCANN modeling framework is a generalization of the ANN model, analogous to the SCAR framework for the ARX model:

1. Decompose the series in the calibration window into the LTSC $T_{d,h}$ and the stochastic component $q_{d,h}$
2. Decompose the exogenous series in the calibration window using the same type of LTSC as for prices
3. Calibrate the ANN model to $q_t$ and compute forecasts for the 24 hours of the next day (24 separate series)
4. Add stochastic component forecasts $\hat{q}_{d+1,h}$ to persistent forecasts $\hat{T}_{d+1,h}$ of the LTSC to yield log-price forecasts $\hat{p}_{d+1,h}$
5. Convert them into price forecasts of the SCANN model, i.e., $\hat{P}_{d+1,h} = \exp(\hat{p}_{d+1,h})$
Number of hidden neurons

There is no universally optimal number, but the errors are smallest for 4 to 6 neurons in the hidden layer.
Committee machines of (SC)ANN networks

Every forecast yields slightly different results $\Rightarrow$ two ‘model categories’ are considered:

- $\text{ANN}_1$ – the ‘expected’ result for a single $\text{ANN}$ network, an average of error scores across separate runs
- $\text{ANN}_5$ – a forecast average of 5 runs (hour-by-hour) with identical parameters, a so-called committee machine

Analogously:

- $\text{SCANN}_1$ – the ‘expected’ result for a single $\text{SCANN}$ network
- $\text{SCANN}_5$ – a committee machine of 5 SCANNs
Committee machines of (SC)ANN networks

- Real price
- Single run
- Forecast

Rafał Weron (Wrocław, PL)
Sample gains from using committee machines

- Forecast errors roughly scale as a power-law function of the number of networks in a committee machine
- We should use as large committee machines as we can ...
Sample gains cont.

... however, the time needed may be substantial, e.g., for generating forecasts for the next 24 hours:

<table>
<thead>
<tr>
<th>Model</th>
<th>ARX</th>
<th>SCARX-HP_{10^8}</th>
<th>SCARX-S_9</th>
<th>ANN_1</th>
<th>ANN_5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>8.6ms</td>
<td>13.5ms</td>
<td>37.3ms</td>
<td>7.6s</td>
<td>38.2s</td>
</tr>
</tbody>
</table>

SCANN times are omitted here, because LTSC computation is negligible compared to training the ANN.
Weekly-weighted Mean Absolute Error (WMAE)

- Following Conejo et al. (2005), Weron & Misiorek (2008) and Nowotarski et al. (2014), among others, we use:

\[
WMAE_w = \frac{1}{\bar{P}_{168}} \quad \text{MAE}_w = \frac{1}{168 \cdot \bar{P}_{168}} \quad \left| \sum_{d=\text{Sun}}^{\text{Mon}} \sum_{h=1}^{24} P_{d,h} - \hat{P}_{d,h} \right|
\]

- where \( \bar{P}_{168} = \frac{1}{168} \sum_{d=\text{Mon}}^{\text{Sun}} \sum_{h=1}^{24} P_{d,h} \)

\[
\overline{WMAE} = \frac{1}{w_{\text{max}}} \sum_{w=1}^{w_{\text{max}}} WMAE_w
\]

- where \( w_{\text{max}} = 103 \) for GEFCom and 104 for Nord Pool
Table 1: Average WMAE in percent for all 103 weeks of the GEFCom2014 out-of-sample test period (upper half) or all 104 weeks of the Nord Pool out-of-sample test period (lower half). Results for the best performing model in each row are emphasized in bold. Note, that results for the SCARX models are the same as in Uniejewski et al. (2017).

<table>
<thead>
<tr>
<th></th>
<th>Naïve</th>
<th>ARX</th>
<th>ANN₁</th>
<th>ANN₅</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEFCom2014</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Benchmarks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARX</td>
<td>12.226</td>
<td>11.106</td>
<td>10.849</td>
<td>10.732</td>
</tr>
<tr>
<td>ANN₁</td>
<td>11.438</td>
<td>11.085</td>
<td>11.216</td>
<td>11.363</td>
</tr>
<tr>
<td>ANN₅</td>
<td>10.598</td>
<td>10.516</td>
<td>10.627</td>
<td>10.547</td>
</tr>
</tbody>
</table>

**SCARX / SCANN with wavelet approximation of price and load**

<table>
<thead>
<tr>
<th></th>
<th>S₅</th>
<th>S₆</th>
<th>S₇</th>
<th>S₈</th>
<th>S₉</th>
<th>S₁₀</th>
<th>S₁₁</th>
<th>S₁₂</th>
<th>S₁₃</th>
<th>S₁₄</th>
</tr>
</thead>
</table>

**SCARX / SCANN with HP filter on price and load (λ)**

<table>
<thead>
<tr>
<th></th>
<th>10⁸</th>
<th>5 · 10⁸</th>
<th>10⁹</th>
<th>5 · 10⁹</th>
<th>10¹⁰</th>
<th>5 · 10¹⁰</th>
<th>10¹¹</th>
<th>5 · 10¹¹</th>
</tr>
</thead>
</table>
Average WMAE for Nord Pool

Table 1: Average WMAE in percent for all 103 weeks of the GEFCom2014 out-of-sample test period (upper half) or all 104 weeks of the Nord Pool out-of-sample test period (lower half). Results for the best performing model in each row are emphasized in bold. Note, that results for the SCARX models are the same as in Uniejewski et al. (2017).

<table>
<thead>
<tr>
<th></th>
<th>Naïve</th>
<th>ARX</th>
<th>ANN₁</th>
<th>ANN₅</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Benchmarks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANN₅</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*SCARX / SCANN with wavelet approximation of price and load*
Aggregate results of SCANN performance

GEFCom2014

Note: Step 1(b) is important (green vs. yellow)!
The Diebold-Mariano test (1995)

- We define the error function as

\[ L(\varepsilon_d) = ||\varepsilon_d||_1 = \sum_{h=1}^{24} |P_{d,h} - \hat{P}_{d,h}| \]

- For each pair of models we compute the loss differential

\[ D_d = L(\varepsilon^\text{model}_X) - L(\varepsilon^\text{model}_Y) \]

- Hypothesis \( H_0: E(D_d) \leq 0 \), \textit{model}_X outperforms \textit{model}_Y

- Reversed hypothesis \( H_0^R: E(D_d) \geq 0 \), \textit{model}_Y outperforms \textit{model}_X
Diebold-Mariano test results

### GEFCom2014

<table>
<thead>
<tr>
<th>Model</th>
<th>1: over 24h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive</td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td></td>
</tr>
<tr>
<td>SCANN-S</td>
<td></td>
</tr>
<tr>
<td>SCANN-HP</td>
<td></td>
</tr>
<tr>
<td>ARX</td>
<td></td>
</tr>
<tr>
<td>SCARX-S</td>
<td></td>
</tr>
<tr>
<td>SCARX-HP</td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td></td>
</tr>
<tr>
<td>SCANN-S</td>
<td></td>
</tr>
<tr>
<td>SCANN-HP</td>
<td></td>
</tr>
</tbody>
</table>

### Nord Pool

<table>
<thead>
<tr>
<th>Model</th>
<th>1: over 24h</th>
<th>1: over 24h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCANN-S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCANN-HP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARX</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCARX-S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCARX-HP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCANN-S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCANN-HP</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Rafał Weron (Wrocław, PL)
Conclusions

- Using Seasonal Component ANN (SCANN) models can yield statistically significant improvement over the ANN benchmark
  - $\text{SCANN}_5$ returns 0.72–0.99% lower WMAE than $\text{ANN}_5$
- The accuracy gains from using LTSC are greater in ANN models than in regression models
  - $\text{SCARX}$ models yield only a 0.35–0.80% improvement in WMAE vs. the $\text{ARX}$ benchmark
- Forecast averaging is crucial in outperforming the $\text{SCARX}$ model
  - $\text{SCANN}_5$ yields 0.21–0.36% lower WMAE