Electricity Demand Probabilistic Forecasting

With Quantile Regression Averaging

Bidong Liu, Jakub Nowotarski, Tao Hong, Rafał Weron

Department of Operations Research, Wrocław University of Technology, Poland
Big Data Energy Analytics Laboratory, University of North Carolina at Charlotte

Riverside, 24.06.2015

Based on:
Bidong Liu, Jakub Nowotarski, Tao Hong and Rafal Weron, Probabilistic Load Forecasting via Quantile Regression Averaging on Sister Forecasts, IEEE Transactions on Smart Grid, forthcoming
This work was supported by funds from NCN (Poland) through grant no. 2013/11/N/HS4/03649
Motivation: probabilistic forecasts

- Stochastic nature of forecasting
- Assessment of future uncertainty
- Ability to plan different strategies for the range of possible outcomes indicated by the probabilistic forecast
- Variability of the electricity demand becoming a challenge to the utility industry
- → useful in practice (risk management and decision-making)
Motivation: combining forecasts

- Similar to portfolio diversification and management
- Availability of various models/experts’ predictions
- No single best forecasting method
- Generally believed to improve forecast accuracy
Motivation: load forecasting

- Interval/density forecast, combining not so popular in load forecasting
- Combine point predictions for probabilistic forecasting → opportunity to leverage existing research
- Use methodology proved to work well (J. Nowotarski and R. Weron (2014), T. Hong, B. Liu, and P. Wang (2015))
- Relative simplicity of the two key components
Background: **Point** forecast averaging

\[ f_{c} = \sum_{i=1}^{N} w_{i} f_{i} \]

Individual forecasts

- \( f_{1} \)
- \( f_{2} \)
- \( \ldots \)
- \( f_{N} \)

Weights estimation

Combined forecast

- \( f_{C} \)
Background: **Interval** forecast averaging

- For point forecasts:  
  \[ f_c = \sum_{i=1}^{M} w_i f_i \] 
  (e.g. a linear regression model)

- For interval forecasts the above formula may not hold

- A linear combination of \( \alpha \)-th quantiles is not an \( \alpha \)-th quantile of a linear combination of random variables

  \[ q_c^\alpha \neq \sum_{i=1}^{M} w_i q_i^\alpha \]

- \( \rightarrow \) A possibility for development of new approaches
Background: quantile regression
Background: quantile regression
Proposed model: Quantile Regression Averaging

\[
\min_{w_t} \left[ \sum_t (q - 1_{p_t < \hat{f}_{t,w_t}})(p_t - \hat{f}_t w_t) \right]
\]

Individual point forecasts

\[f_1, f_2, \ldots, f_N\]

Quantile regression

Combined interval forecast (2 quantiles)

\[f_C\]
Methodology: sister models and sister forecasts

- **Motivation**: variable selection is core in regression model for 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
- **Sister forecasts** are generated from sister models
- The family of sister recency effect models:

\[
\hat{y}_t = \beta_0 + \beta_1 M_t + \beta_2 W_t + \beta_3 H_t + \beta_4 W_t H_t + f(T_t) + \\
+ \sum_d f(\tilde{T}_{t,d}) + \sum_{\text{lag}} f(T_{t-\text{lag}}),
\]
Methodology: the data (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
Methodology: 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
Methodology: evaluation of forecasts

- Pinball loss function for 99 percentiles

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

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

\[ W_t = \begin{cases} \delta_t & \text{for } p_t \in [L_{t|t-1}, U_{t|t-1}], \\ \delta_t + \frac{2}{\alpha}(L_{t|t-1} - p_t) & \text{for } p_t < L_{t|t-1}, \\ \delta_t + \frac{2}{\alpha}(p_t - U_{t|t-1}) & \text{for } p_t > U_{t|t-1}, \end{cases} \]

where \( \delta_t = U_{t|t-1} - L_{t|t-1} \) is the interval’s width
Results: validation period

- 7 QRA models, 8+1 individual models
- 4 lengths of calibration period
- One year of validation to pick up best (model, length) pairs
- → QRA models are dominantly better than the benchmark models
## Results: test period

<table>
<thead>
<tr>
<th>Model class</th>
<th>Pinball</th>
<th>Winkler (50%)</th>
<th>Winkler (90%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QRA(8,183)</td>
<td>2.85</td>
<td>25.04</td>
<td>55.85</td>
</tr>
<tr>
<td>Ind(1,91)</td>
<td>3.22</td>
<td>26.35</td>
<td>56.38</td>
</tr>
<tr>
<td>BI(-,365)</td>
<td>3.00</td>
<td>26.38</td>
<td>57.17</td>
</tr>
<tr>
<td>Direct</td>
<td>3.19</td>
<td>26.62</td>
<td>94.27</td>
</tr>
<tr>
<td>Vanilla</td>
<td>8.00</td>
<td>70.51</td>
<td>150.0</td>
</tr>
</tbody>
</table>
Discussion

- **Resolution** – log-transform caused intervals to be wider in peak hours

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

- **Extensions**
  - Sister forecasts eg. for machine learning methods
  - QRA for expert forecasts
Summary

- QRA – a new technique the load forecasting literature
- Practical value (1) – input to QRA from point forecasts
- Practical value (2) – the sister forecasts are easy to generate
- Publicly available data (GEFCom2014)
- **Accurate** – confirmed by the pinball loss function and Winkler scores
Questions?

jakub.nowotarski@pwr.edu.pl
Methodology: sister models and sister forecasts

\[ \hat{y}_t = \beta_0 + \beta_1 M_t + \beta_2 W_t + \beta_3 H_t + \beta_4 W_t H_t + \text{calendar effects} + \text{temp. dependence} \]

\[ f(T_t) + \sum_{d} f(\tilde{T}_{t,d}) + \sum_{\text{lag}} f(T_{t-\text{lag}}), \]

where:

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

\[ \tilde{T}_{t,d} = \frac{1}{24} \sum_{\text{lag}=24d-23}^{24d} T_{t-\text{lag}} \]
Extension: Factor Quantile Regression Averaging

Individual point forecasts

\[ f_1 \]
\[ f_2 \]
\[ \ldots \]
\[ f_N \]

PCA

Quantile regression

\[ F_1 \]
\[ \ldots \]
\[ F_K \]

Combined interval forecast (2 quantiles)

K factors extracted from a panel of point forecasts, \( K<N \)

B. Liu, J. Nowotarski, T. Hong & R. Weron
Price forecasting results: case study 1

J. Nowotarski and R. Weron (2014, Computational Statistics)
Price forecasting results: case study 2
K. Maciejowska, J. Nowotarski and R. Weron (2015, IJF)