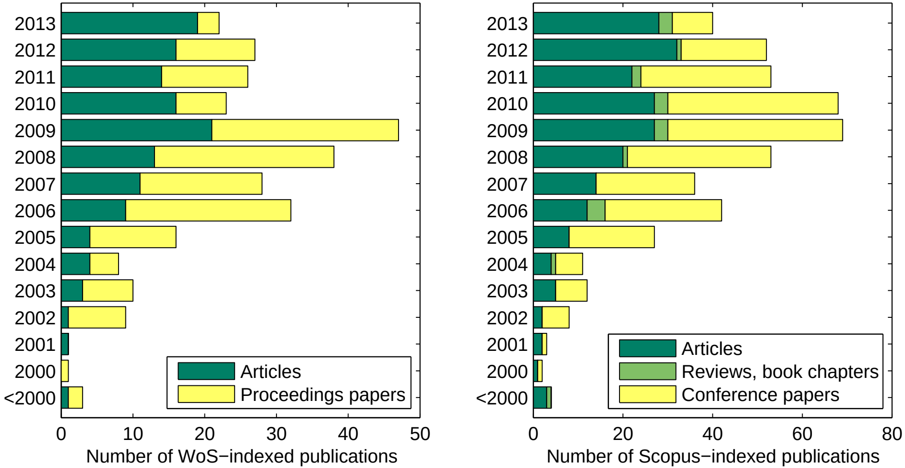
A look into the future of electricity (spot) price forecasting

Rafał Weron

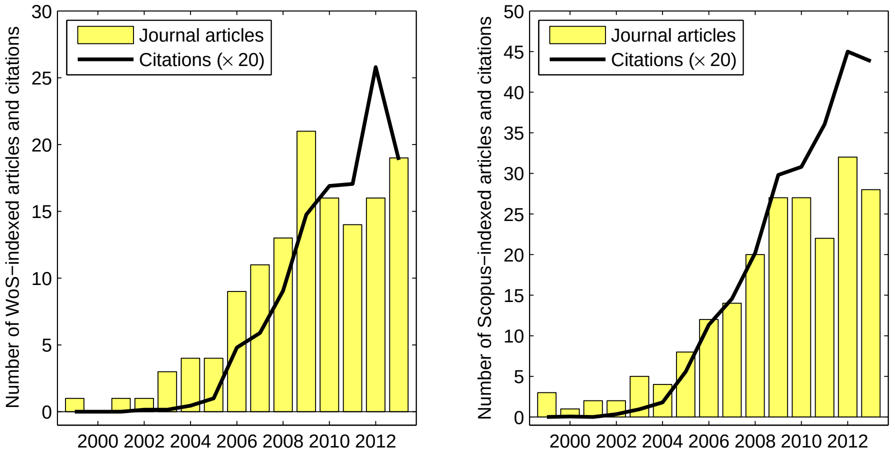
Institute of Organization and Management
Wrocław University of Technology, Poland

28 April 2014

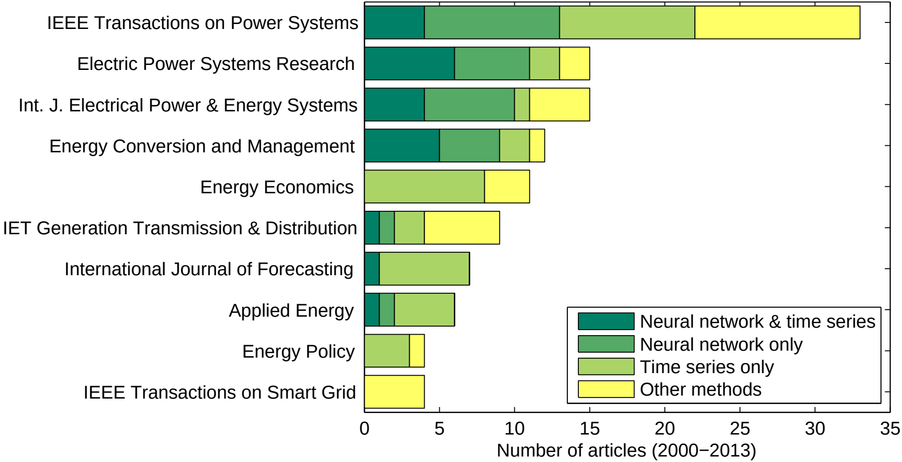
'Electricity price forecasting' (EPF) publications



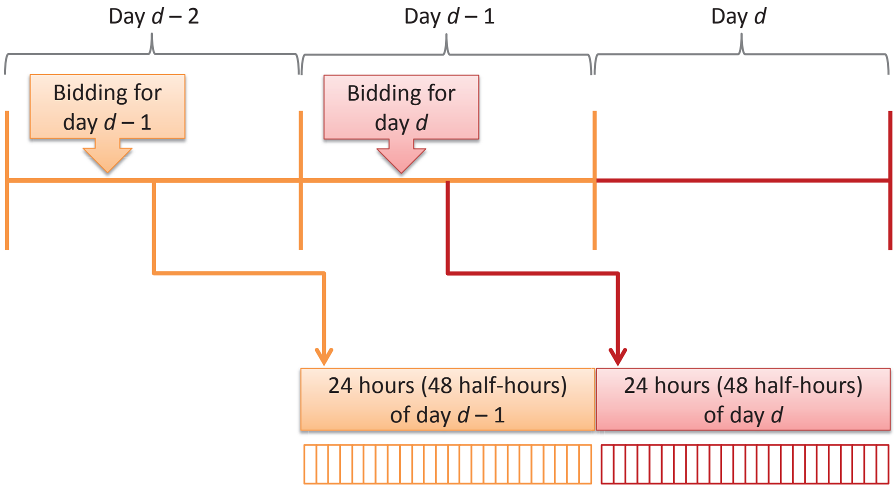
EPF journal articles and citations to those articles



Ten most popular journals



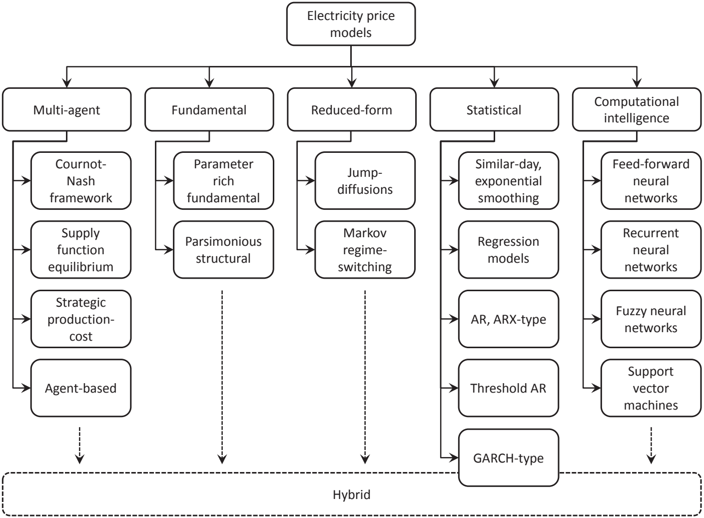
The electricity ‘spot’ price



Forecasting horizons

- 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 modeling approaches



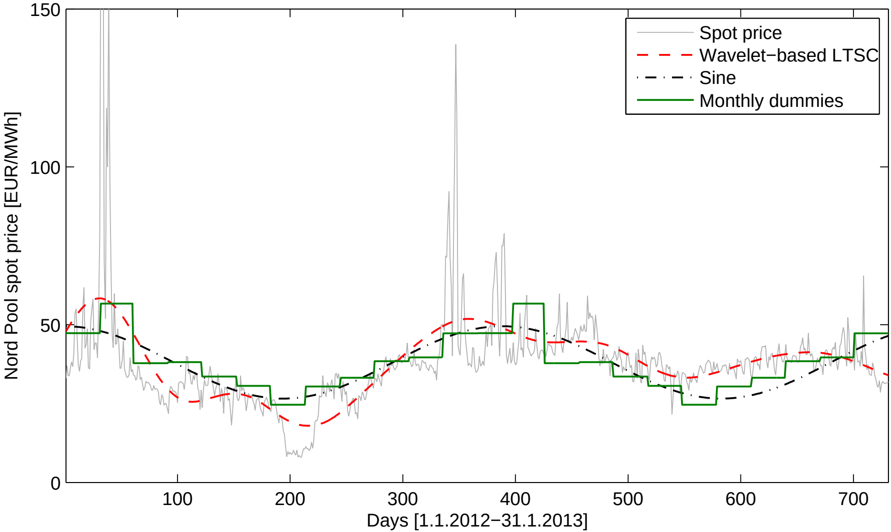
A look into the future of EPF

1. Fundamental price drivers and input variables
   - Modeling and forecasting the trend-seasonal components
   - The reserve margin and spike forecasting
2. Beyond point forecasts – probabilistic forecasts
3. Combining forecasts
   - Point forecasts
   - Probabilistic forecasts
4. Multivariate factor models
5. The need for an EPF-Competition
   - A universal test ground
   - Guidelines for evaluating forecasts

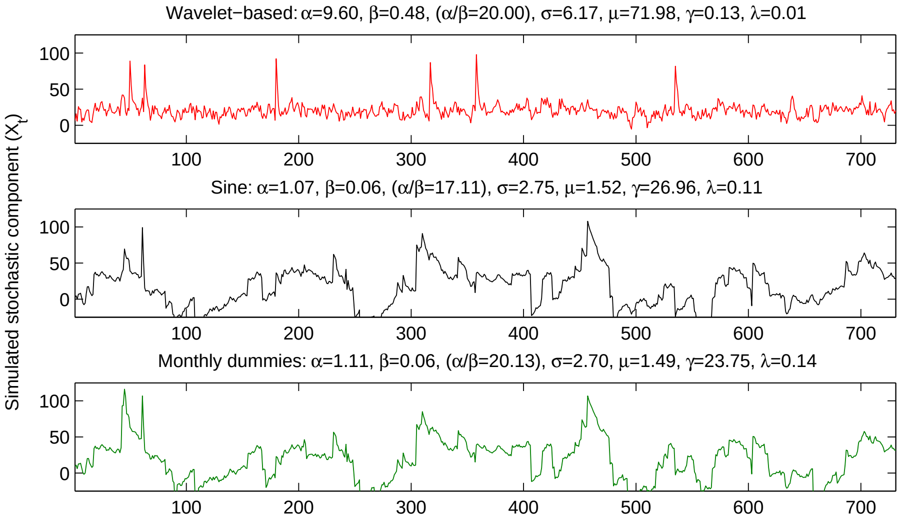
Modeling the trend-seasonal components

- 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), Nowotarski et al. (2013b)

Modeling the LTSC



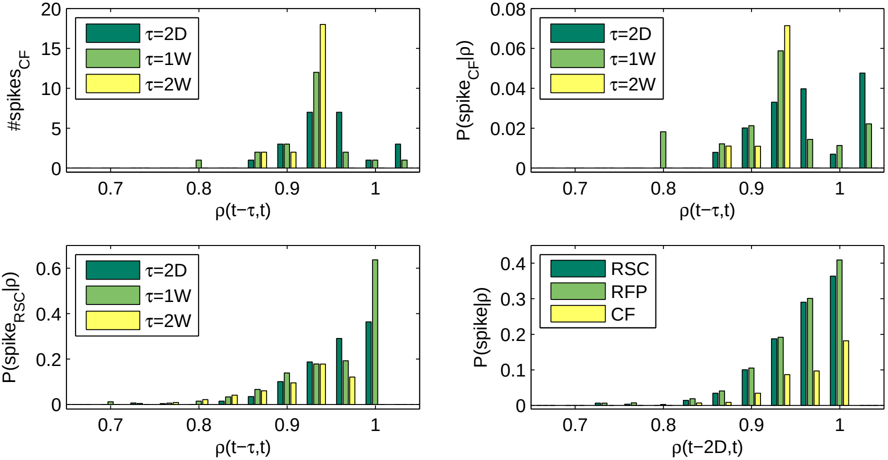
Adequate seasonal decomposition is important !



The reserve margin and spike forecasting

- *Reserve margin*, also called *surplus generation*, relates the available capacity (generation, supply), C_t, and the demand (load), D_t, at a given moment in time t
  - The traditional engineering notion: $RM = C_t - D_t$
  - Some authors prefer to work with dimensionless ratios: $\rho_t = \frac{D_t}{C_t}$ or the so-called *capacity utilization* $\mathrm{CU} = 1 - \frac{D_t}{C_t}$
- Consider $\rho(t_1, t_2) = \frac{D(t_1,t_2)}{C(t_1,t_2)}$
  - calculated at time t_1 (e.g. today) for an upcoming period t_2
  - $D(t_1, t_2)$ is the National Demand Forecast (Indicated Demand)
  - $C(t_1, t_2)$ is the predicted Generation Capacity (Indicated Generation, see www.bmreports.com)
  - See Cartea et al. (2009), Maryniak and Weron (2014)

$\rho(t_1, t_2)$ for 2003-05 (top) and 2006-12 (bottom)



Anderson and Davison (2008): $\rho = 85\%$ is the 'industrial standard' warranting a safe functioning of the power system

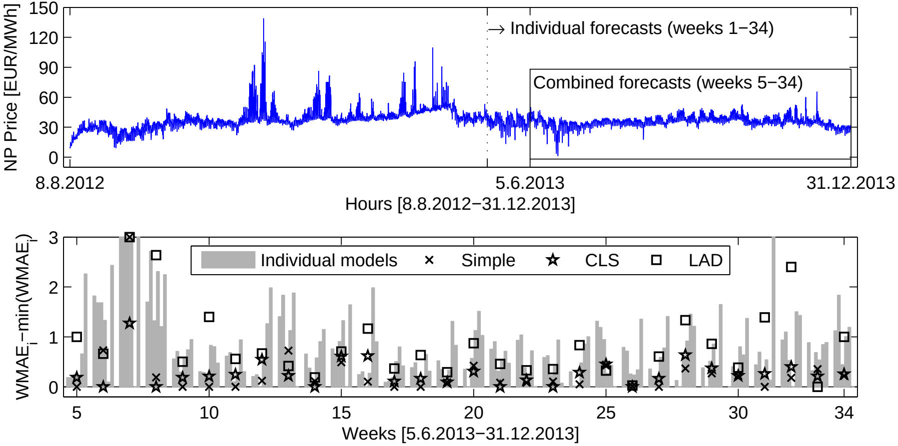
Probabilistic forecasts

- Interval forecasts (only 10 articles)
  - Zhang et al. (2003), Zhang and Luh (2005), Misiorek et al. (2006), Weron and Misiorek (2008), Zhao et al. (2008), Serinaldi (2011), Gonzalez et al. (2012), Wu et al. (2013), Khosravi et al. (2013a,b)
  - In only one paper formal statistical tests for coverage are conducted $\rightarrow$ conditional coverage of Christoffersen (1998)
- Density forecasts (only 2/3 articles)
  - Serinaldi (2011) forecasts the distribution of electricity prices, but computes and discusses only the PI
  - Huurman et al. (2012) perform density forecasting of Nord Pool spot prices and use the test of Berkowitz (2001)
  - Jonsson et al. (2014) generate prediction densities of day-ahead electricity prices in Western Denmark, but do not test them

Forecast combinations, forecast/model averaging

- The idea goes back to the 1960s
  - 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 for EPF: Bordignon et al. (2013), Nowotarski et al. (2013a), Nowotarski and Weron (2014) and Raviv et al. (2013)
- In the 'AI world': *committee machines* or *ensemble averaging*
  - 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 !

To combine or not to combine?



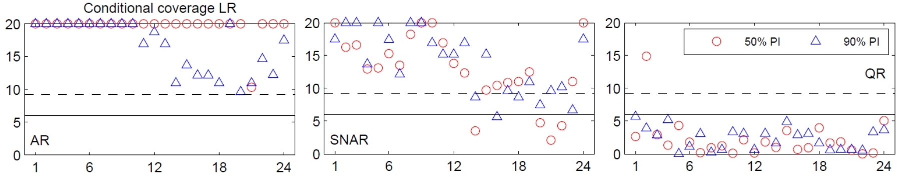
To combine or not to combine?

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

Combining interval/density EPF – only one paper

- Nowotarski and Weron (2014) propose a new method for constructing PI, which utilizes the concept of *quantile regression* (QR) and a pool of point forecasts of individual models
  - Empirical PI from combined forecasts do not yield gains
  - QR-based PI are more accurate than those of the benchmark (AR) and the best individual model (SNAR)



Factor models

- All hourly prices P_{kt}, $k = 1, ..., 24$, co-move and depend on a small set of common factors $\mathbf{F}_t = [F_{1t}, ..., F_{Nt}]'$
- The individual series P_{kt} can be modeled as a linear function of N principal components $\mathbf{F}_t$ and stochastic residuals ν_{kt}:

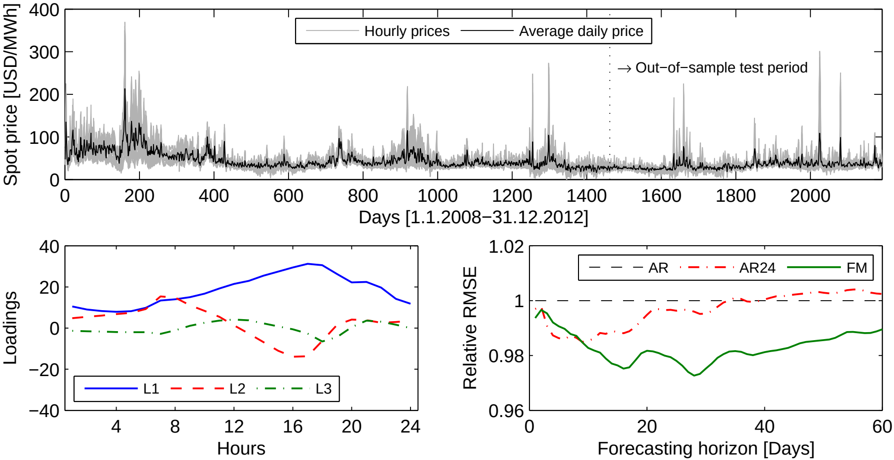
$$P_{kt} = \boldsymbol{\Lambda}_k \mathbf{F}_t + \nu_{kt}, \tag{1}$$

where the *loads* (or *loadings*) $\boldsymbol{\Lambda}_k = [\Lambda_{k1}, ..., \Lambda_{kN}]$ describe the relation between the factors $\mathbf{F}_t$ and the panel variables P_{kt}

  - See e.g. Bai (2003), Stock and Watson (2002)
  - It is natural to assume that the common factors follow a VAR(p) model, see e.g. Maciejowska and Weron (2014)

Forecasting PJM Dominion Hub daily spot prices

Using the information contained in hourly prices



Factor models and EPF

- Applications of multivariate models to EPF are very recent
  - Chen et al. (2008), Härdle and Trück (2010)
- In the last two years, an increased inflow of 'multivariate EPF papers' can be observed
  - Garcia-Martos et al. (2012), Peña (2012), Vilar et al. (2012), Elattar (2013), Miranian et al. (2013), Wu et al. (2013)
- The idea originating in macroeconometrics of using disaggregated data for forecasting of aggregated variables
  - Liebl (2013), Maciejowska and Weron (2013, 2014), Raviv et al. (2013)

The need for an EPF-Competition

- Many of the published results seem to contradict each other
  - Misiorek et al. (2006) report a very poor forecasting performance of a MRS model, while Kosater and Mosler (2006) reach opposite conclusions for a similar MRS model but a different market and mid-term forecasting horizons
  - On the other hand, Heydari and Siddiqui (2010) find that a regime-switching model does not capture price behavior correctly in the mid-term
- Cross-category comparisons are even less conclusive and more biased
  - Typically advanced statistical techniques are compared with simple AI methods, see e.g. Conejo et al. (2005a), and vice versa, see e.g. Amjady (2006)

A universal test ground

- This calls for a comprehensive and thorough study involving
  1. the same datasets
  2. the same robust error evaluation procedures
  3. statistical testing of the significance of the outperformance of one model by another
- Like the Makridakis or M-Competitions for economic forecasting
- Global Energy Forecasting Competition 2014 includes a 'price forecasting' track this year, see www.drhongtao.com/gefcom

Guidelines for evaluating forecasts

- A selection of the better performing measures
  - weighted-MAE, like the weekly-weighted WMAE
  - seasonal MASE (*Mean Absolute Scaled Error*)
  - RMSSE (*Root Mean Square Scaled Error*)

  should be used exclusively or in conjunction with the more popular ones (MAPE, RMSE)
- Statistical testing for the significance of the difference in forecasting accuracy of the models
  - The Diebold and Mariano (1995) test; for uses and abuses see Diebold (2013)
  - The *model confidence set* approach of Hansen et al. (2011)

Bibliography

Based on an invited paper prepared for the *International Journal of Forecasting*. A restricted working paper version is available for download from: http://ideas.repec.org/p/wuu/wpaper/hsc1402.html