Modeling and forecasting electricity forward prices: A DSFM approach

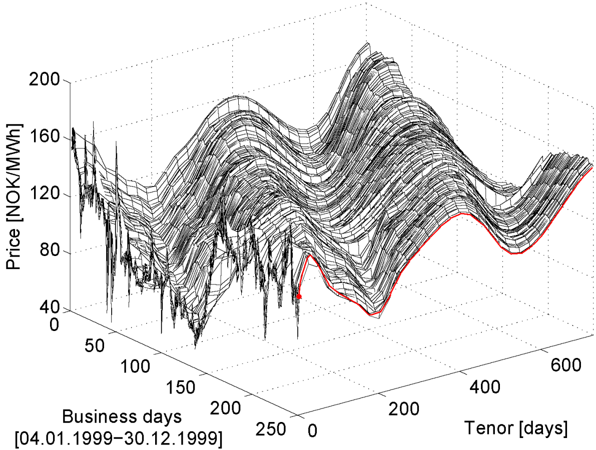
Rafał Weron*
Szymon Borak

Hugo Steinhaus Center,
Wrocław University of Technology

and

CASE-Center for Applied Statistics and Economics,
Humboldt-Universität zu Berlin

How can we model the dynamics of the electricity forward curve?



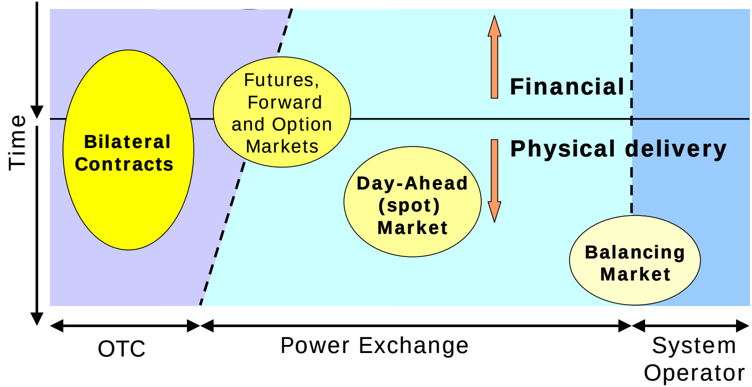
Motivation

- The electricity forward curve is a complex object with a non-trivial structure which exhibits **seasonality** and **extreme volatility at the short end**
- Our aim is to model and estimate forward curves for **trading**, **hedging** and **risk management**
- In this context the electricity forward curve acts as a very **high-dimensional** state variable
- Practice requires a **low-dimensional** representation of the curve

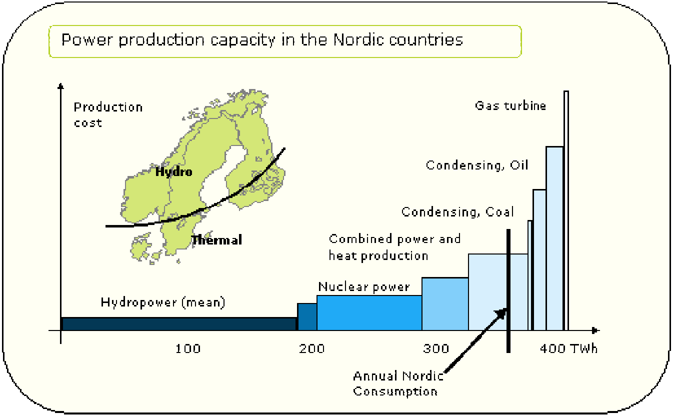
Agenda

1. Motivation ✓
2. **The Nordic power market and the forward curves**
3. The Dynamic Semiparametric Factor Model
4. Modeling and forecasting
5. Conclusions

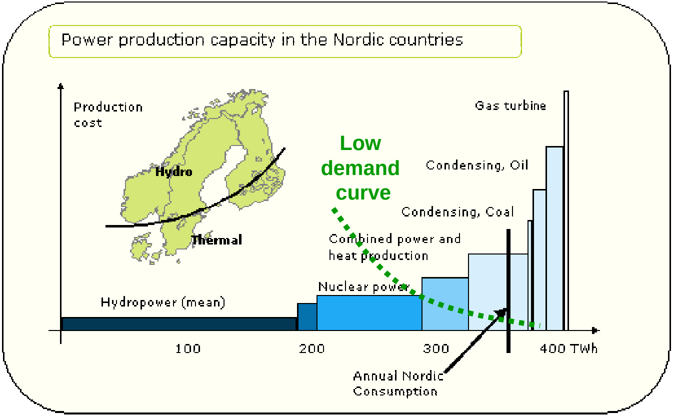
Wholesale power market structure



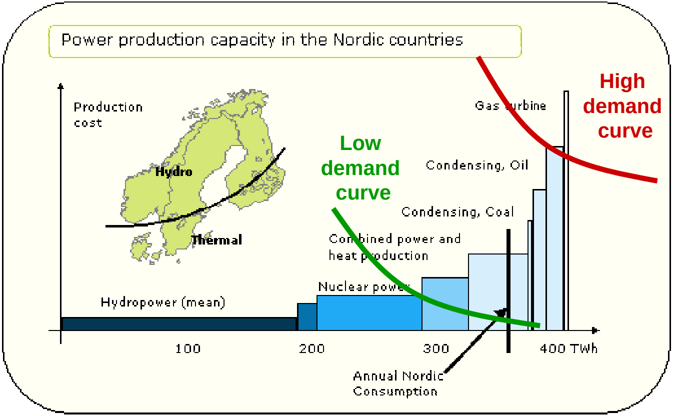
Supply stack and the market cross



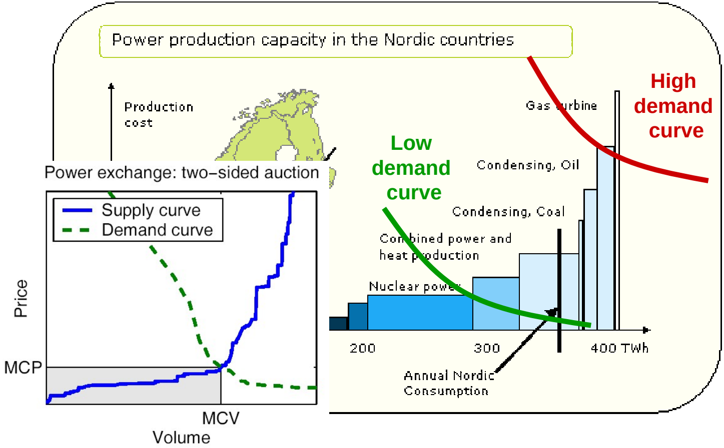
Supply stack and the market cross cont.



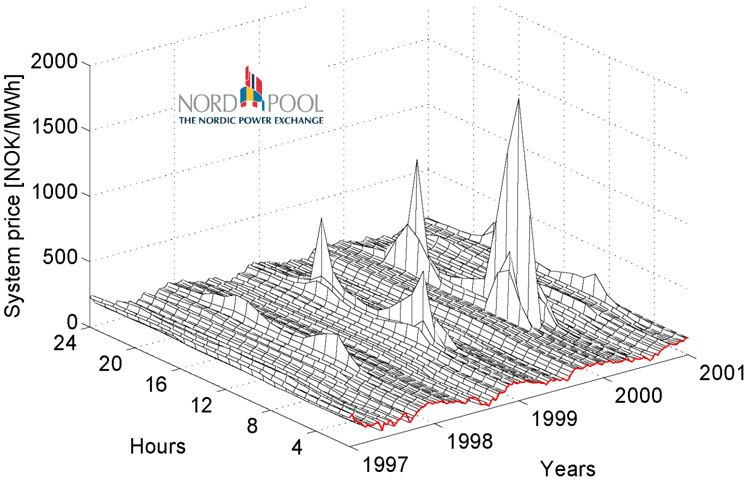
Supply stack and the market cross cont.



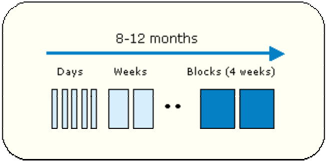
Supply stack and the market cross cont.



Seasonality, extreme volatility and spikes

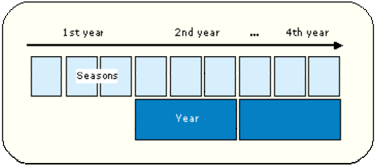


Futures and forwards at Nord Pool



Futures contracts

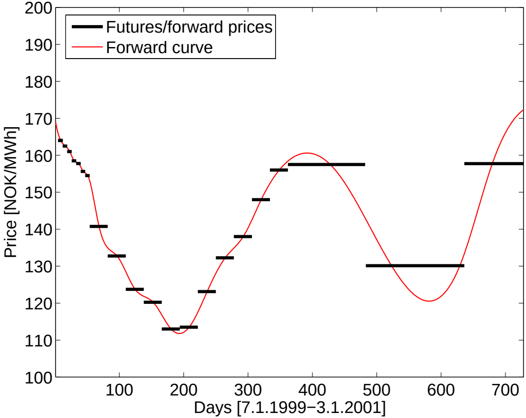
Day contracts (Dxx): 24 hours
Week contracts (GUxx): 7 days
Block contracts (GBxx): 4 weeks



Forward contracts

Winter 1 (FWV1xx): January - April
Summer (FWSOxx): May - September
Winter 2 (FWV2xx): October - December
Year (FWYRxx)

The forward curve ...



... and its dynamics

Agenda

1. Motivation ✓
2. The Nordic power market and the forward curves ✓
3. **The Dynamic Semiparametric Factor Model**
4. Modeling and forecasting
5. Conclusions

Background

- The **Dynamic Semiparametric Factor Model (DSFM)** is a principal component type approach
- Primary application: Dimension reduction
- Originally introduced for modeling implied volatility surfaces [Fengler et al. (2007)]
- Can be seen as a combination of functional principal component analysis (fPCA) and nonparametric curve estimation

The model

- The **Dynamic Semiparametric Factor Model** has the form

$$\mathbf{Y}_t = m_0(\mathbf{X}_t) + \sum_{l=1}^{L} Z_{t,l} m_l(\mathbf{X}_t) + \varepsilon_t$$

  where the data vector $\mathbf{X}_t$ has J coordinates $X_{t,j}$ (observations per day)

- $\mathbf{m}(\cdot)$ is a tuple of basis functions $(m_0, m_1, \ldots, m_L)^\top$

- $\mathbf{Z}_t = (1, Z_{t,1}, \ldots, Z_{t,L})^\top$ is a multivariate time series

The model cont.

- The functions m_l reflect the time invariant structure of $\mathbf{Y}_t$
- $\widehat{m}_l$ is a nonparametric estimator of m_l obtained **directly** from the data points $X_{t,j}$, i.e. not from some estimated functions of $X_{t,j}$ as in PCA (piecewise constant, smoothed forward curves)
- The coefficients $Z_{t,l}$ describe the dynamic behavior of the forward curves
- The whole complex system can be modeled through a typical time series analysis of the estimates $\widehat{Z}_{t,l}$

Estimation

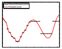
$$\mathbf{Z}_t^\top \mathbf{m}(\mathbf{X}_t) = \sum_{l=0}^{L} Z_{t,l} \sum_{k=1}^{K} a_{l,k}\psi_k(\mathbf{X}_t) = \mathbf{Z}_t^\top \mathbf{A}\psi(\mathbf{X}_t)$$

- $\psi(\cdot) = (\psi_1, \ldots, \psi_K)^\top$ is a vector of known expansion functions (e.g. B-splines)
- $\mathbf{A} \in \mathbb{R}^{(L+1)\times K}$ is a matrix of coefficients
- the smoothing parameters L (dimension of the time series; we use $L = 3, \ldots, 6$) and K (number of series expansion functions; $K = 19$ functions on 16 knots) have to be specified in advance

Estimation cont.

The least squares estimators $\widehat{\mathbf{Z}}_t = (\widehat{Z}_{t,0}, \ldots, \widehat{Z}_{t,L})^\top$ and $\widehat{\mathbf{A}} = \left(\widehat{a}_{l,k}\right)_{l=0,\ldots,L;k=1,\ldots,K}$ are obtained from

$$\sum_{t=1}^{T} \sum_{j=1}^{J} \left\{ Y_{t,j} - \widehat{\mathbf{Z}}_t^\top \widehat{\mathbf{A}} \psi(X_{t,j}) \right\}^2 = \min_{\widehat{\mathbf{Z}}_t, \widehat{\mathbf{A}}}$$



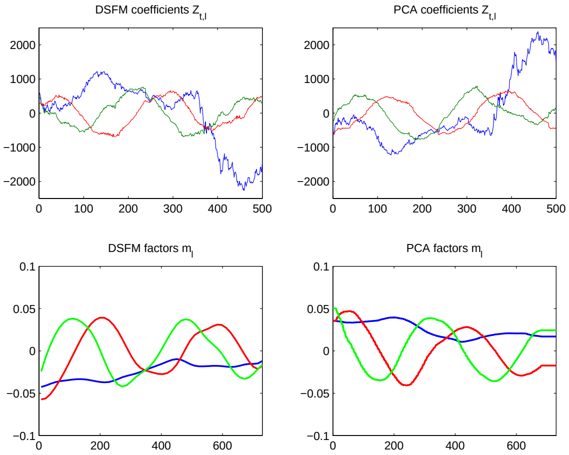
Agenda

1. Motivation ✓
2. The Nordic power market and the forward curves ✓
3. The Dynamic Semiparametric Factor Model ✓
4. **Modeling and forecasting**
5. Conclusions

The data

- Database of Nord Pool futures and forward prices from the period Jan. 4, 1999 – May 23, 2002, i.e. 843 (business) days
- A 500 day window is used for calibration
- For each day, 1, 5, 10, 25 and 125 day-ahead forward curve forecasts is computed
- This leaves us with $800 - 500 - 125 = 218$ days for which the procedure (calibration + forecast) is repeated
- Both DSFM and PCA models for various L $(= 3, \ldots, 6)$ are evaluated

Sample DSMF and PCA fits ...



... and their dynamics

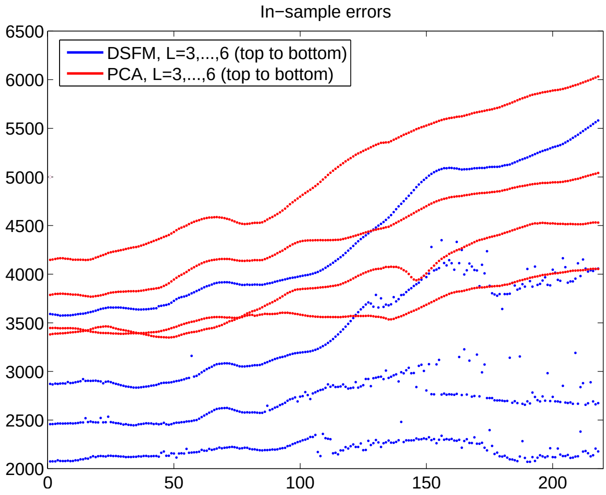
In-sample error measure

- We compute an absolute in-sample error weighted by the length of the delivery period for each contract

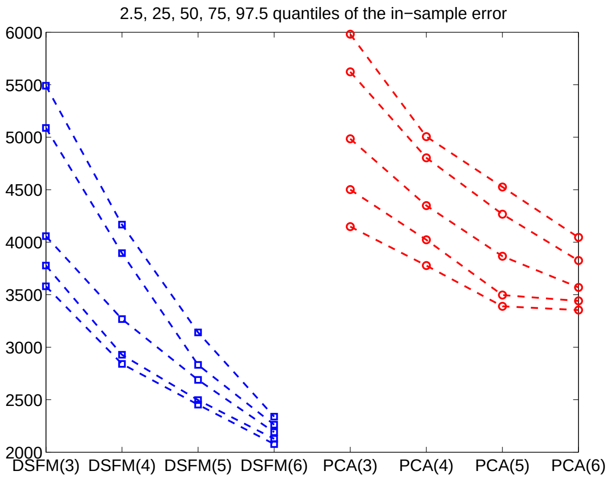
$$\epsilon_t = \sum_j |\text{Model}(X_{t,j}) - Y_{t,j}| \cdot ||I_{t,j}||$$

- where $X_{t,j}$ are the observed maturities (mid-points of the delivery periods) $j = 1, 2, \ldots$ for day t
- $Y_{t,j}$ are the respective prices
- $I_{t,j}$ are the respective time intervals with a constant price $Y_{t,j}$, such that $\bigcup_j I_{t,j} = (0, 2]$ years

In-sample errors



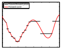
In-sample error statistics



Forecasting setup

- For a 500 day window the DSFM and PCA models (with $L = 3, \ldots, 6$ factors) are calibrated
- For $l = 2, \ldots, L$, $\widehat{Z}_{t,l}$ generally exhibit a seasonal pattern
- A sinusoidal function $g_l(t) = A_l \sin(B_l t + C_l)$ is fitted and removed yielding $\widetilde{Z}_{t,2}, \ldots, \widetilde{Z}_{t,L}$

$$
\begin{aligned}
Y_{t,j} &= m_0(X_{t,j}) + \sum_{l=1}^{L} \widehat{Z}_{t,l} m(X_{t,j}) + \varepsilon_{t,j} \\
&= m_0(X_{t,j}) + \sum_{l=2}^{L} g_l(t) m(X_{t,j}) + \sum_{l=1}^{L} \widetilde{Z}_{t,l} m(X_{t,j}) + \varepsilon_{t,j}
\end{aligned}
$$



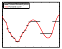
Forecasting models

- Random Walk (**RW**): forward (futures) prices from day t

$$Y^*_{t+h,j} = \widehat{m}_0(X_{t+h,j}) + \sum_{l=1}^{L} \widehat{Z}_{t,l}\widehat{m}_l(X_{t+h,j}) + \varepsilon_t$$

- Trend update (**STr** for DSFM and **PTr** for PCA): the sinusoidal trend g_l is forecasted for $l = 2, \dots, L$ and added to the forward price forecast

$$Y^*_{t+h,j} = \widehat{m}_0(X_{t+h,j}) + \sum_{l=2}^{L} g_l(t+h)\widehat{m}_l(X_{t+h,j}) + \sum_{l=1}^{L} \widetilde{Z}_{t,l}\widehat{m}_l(X_{t+h,j}) + \varepsilon_t$$



Forecasting models cont.

- Trend update from model (**STr2**, **PTr2**): Like 'Trend update', but without the error term

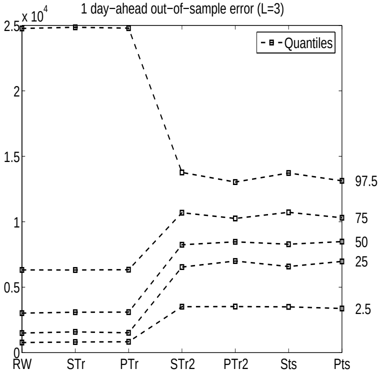
$$Y^*_{t+h,j} = \widehat{m}_0(X_{t+h,j}) + \sum_{l=2}^{L} g_l(t+h)\widehat{m}_l(X_{t+h,j}) + \sum_{l=1}^{L} \widetilde{Z}_{t,l}\widehat{m}_l(X_{t+h,j})$$

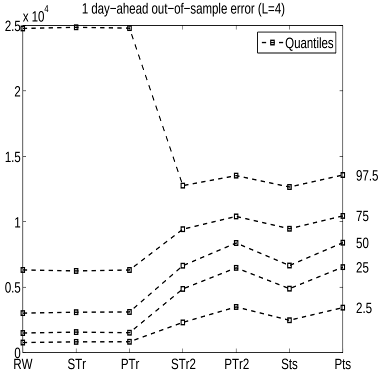
Forecasting models cont.

- Full forecast with AR(2) time series (**Sts**, **Pts**): Additionally includes AR(2) forecasts of all coefficient time series $\widehat{Z}_{t,1}, \widetilde{Z}_{t,2}, \ldots, \widetilde{Z}_{t,L}$

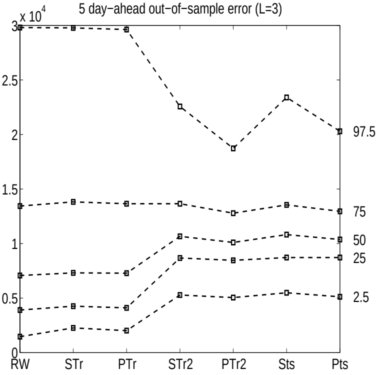
$$Y^*_{t+h,j} = \widehat{m}_0(X_{t+h,j}) + \sum_{l=2}^{L} g_l(t+h)\widehat{m}_l(X_{t+h,j}) + \sum_{l=1}^{L} \widetilde{Z}^*_{t+h,l}\widehat{m}_l(X_{t+h,j})$$

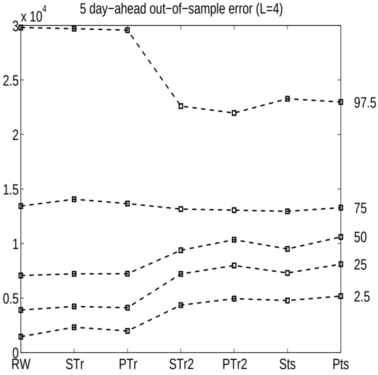
One day-ahead forecasts: $L = 3$ vs. $L = 4$



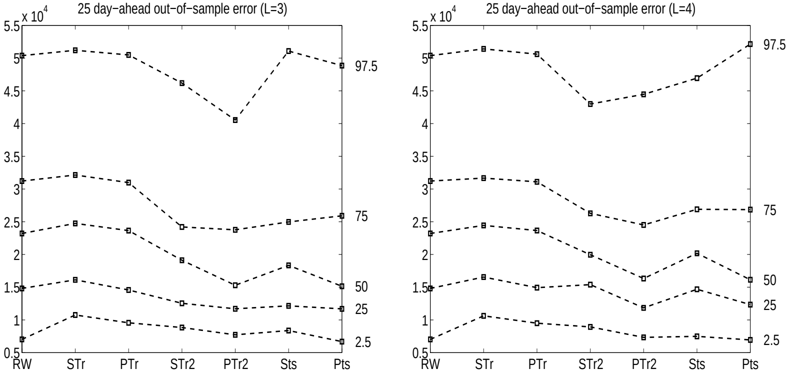


5 day-ahead forecasts: $L = 3$ vs. $L = 4$

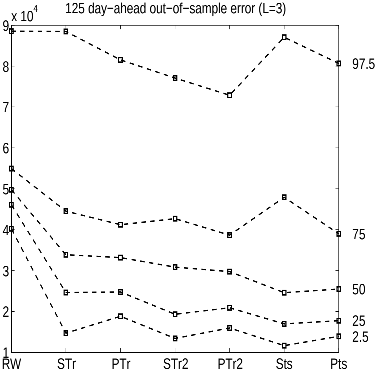


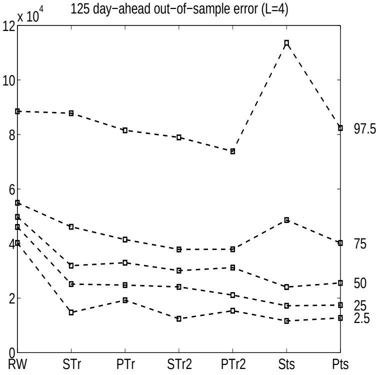


25 day-ahead forecasts: $L = 3$ vs. $L = 4$

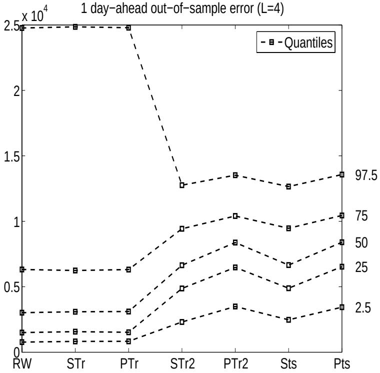


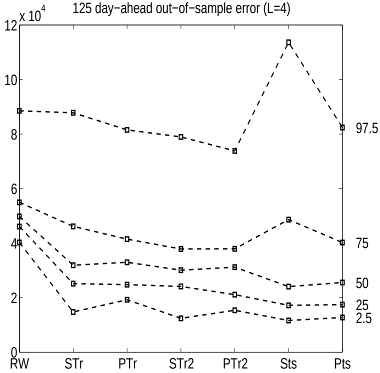
125 day-ahead forecasts: $L = 3$ vs. $L = 4$





Short vs. long term forecasts ($L = 4$)





Agenda

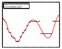
1. Motivation ✓
2. The Nordic power market and the forward curves ✓
3. The Dynamic Semiparametric Factor Model ✓
4. Modeling and forecasting ✓
5. **Conclusions**

Conclusions

- The electricity forward curve is a complex object with a non-trivial structure
- Like PCA, DSFM allows for dimension reduction
- It differs in that
  - the fits are obtained in the **local neighborhood** of forward price–maturity pairs for a given day
  - curve estimation and dimension reduction is achieved in **one single step**
- DSFM offers **superior** in-sample performance

Conclusions cont.

- There is no clear winner in the out-of-sample forecasts
- DSFM-based models are slightly **better** than their PCA-based counterparts for short and long term forecasts and **worse** for medium term predictions
- The current forward curve (RW model) is on average the **best** predictor of the curve in the next few days, it is **inferior** for medium and long term forecasts
- Larger number of basis functions (larger L) **improves** short term forecasts but leads to **increased variance** of the longer term forecasts



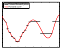
References

Borak, S., Weron, R. (2007)
Modeling and forecasting electricity forward prices using a Dynamic Semiparametric Factor Model, *Working Paper*.

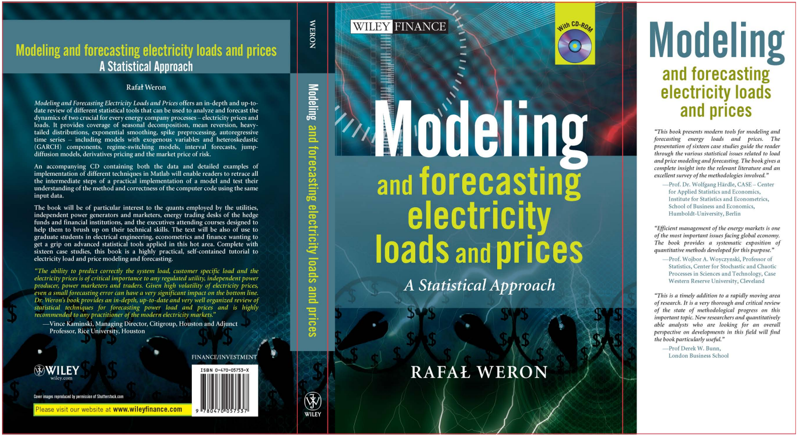
Koekebakker, S., Ollmar, F. (2005)
Forward curve dynamics in the Nordic electricity market, *Managerial Finance* 31: 74–95.

Fengler, M., Härdle, W., Mammen, E. (2007)
A semiparametric factor model for implied volatility surface dynamics, *Journal of Financial Econometrics* 5(2): 189–218.
See also: SFB 649 Discussion Paper 2005-020.

Borak, S., Härdle, W., Mammen, E., Park, B.U. (2007)
Time Series Modelling with Semiparametric Factor Dynamics, SFB 649 Discussion Paper 2007-023.



Read more ...


Modeling and forecasting electricity loads and prices
A Statistical Approach
Rafał Weron
Modeling and Forecasting Electricity Loads and Prices offers an in-depth and up-to-date review of different statistical tools that can be used to analyze and forecast the dynamics of two crucial for every energy company processes – electricity prices and loads. It provides coverage of seasonal decomposition, mean reversion, heavy-tailed distributions, exponential smoothing, spike preprocessing, autoregressive time series – including models with exogenous variables and heteroskedastic (GARCH) components, regime-switching models, interval forecasts, jump-diffusion models, derivatives pricing and the market price of risk.
An accompanying CD containing both the data and detailed examples of implementation of different techniques in Matlab will enable readers to retrace all the intermediate steps of a practical implementation of a model and test their understanding of the method and correctness of the computer code using the same input data.
The book will be of particular interest to the quants employed by the utilities, independent power generators and marketers, energy trading desks of the hedge funds and financial institutions, and the executives attending courses designed to help them to brush up on their technical skills. The text will be also of use to graduate students in electrical engineering, econometrics and finance wanting to get a grip on advanced statistical tools applied in this hot area. Complete with sixteen case studies, this book is a highly practical, self-contained tutorial to electricity load and price modeling and forecasting.
"The ability to predict correctly the system load, customer specific load and the electricity prices is of critical importance to any regulated utility, independent power producer, power marketers and traders. Given high volatility of electricity prices, even a small forecasting error can have a very significant impact on the bottom line. Dr. Weron's book provides an in-depth, up-to-date and very well organized review of statistical techniques for forecasting power load and prices and is highly recommended to any practitioner of the modern electricity markets."
—Vince Kaminski, Managing Director, Citigroup, Houston and Adjunct Professor, Rice University, Houston
FINANCE/INVESTMENT
WILEY
wiley.com
ISBN 0-470-05753-X
Cover images reproduced by permission of Shutterstock.com
Please visit our website at www.wileyfinance.com
WERON
Modeling and forecasting electricity loads and prices
WILEY
WILEY FINANCE
With CD-ROM
Modeling and forecasting electricity loads and prices
A Statistical Approach
RAFAŁ WERON
Modeling and forecasting electricity loads and prices
"This book presents modern tools for modeling and forecasting energy loads and prices. The presentation of sixteen case studies guide the reader through the various statistical issues related to load and price modeling and forecasting. The book gives a complete insight into the relevant literature and an excellent survey of the methodologies involved."
—Prof. Dr. Wolfgang Härdle, CASE – Center for Applied Statistics and Economics, Institute for Statistics and Econometrics, School of Business and Economics, Humboldt-University, Berlin
"Efficient management of the energy markets is one of the most important issues facing global economy. The book provides a systematic exposition of quantitative methods developed for this purpose."
—Prof. Wojbor A. Woyczynski, Professor of Statistics, Center for Stochastic and Chaotic Processes in Sciences and Technology, Case Western Reserve University, Cleveland
"This is a timely addition to a rapidly moving area of research. It is a very thorough and critical review of the state of methodological progress on this important topic. New researchers and quantitatively able analysts who are looking for an overall perspective on developments in this field will find the book particularly useful."
—Prof Derek W. Bunn, London Business School