

Improving electricity market price scenarios by means of forecasting factor models

M. Pilar Muñoz, Cristina Corchero and F.-Javier Heredia

*Dept. of Statistics and Operations Research
Universitat Politècnica de Catalunya, Barcelona (Spain)*

DR 2009/6

1 June 2009

Copies of this report may be downloaded at

Corresponding author: M. Pilar Muñoz
Dept. Statistics and Operations Research,
Technical University of Catalonia, Campus Nord
c. Jordi Girona 1-3, 08034 Barcelona,
tel. (0034) 934017030 fax. (0034) 934015855
email: pilar.munyo@upc.edu

Improving electricity market price scenarios by means of forecasting factor models

M. Pilar Muñoz, Cristina Corchero and F.-Javier Heredia

*Dept. of Statistics and Operations Research
Universitat Politècnica de Catalunya, Barcelona (Spain)*

DR 2009/??

1 June 2009

Copies of this report may be downloaded at

Corresponding author: M. Pilar Muñoz
Dept. Statistics and Operations Research,
Technical University of Catalonia, Campus Nord
c. Jordi Girona 1-3, 08034 Barcelona,
tel. (0034) 934017030 fax. (0034) 934015855
email: pilar.munyo@upc.edu

Improving electricity market price scenarios by means of forecasting factor models

Munoz, M.P., Corchero, C. and Heredia, F.J.

Dept. of Statistics and Operations Research, Universitat Politècnica de Catalunya

Jordi Girona 1-3, Barcelona, 08034, Spain

E-mail: pilar.munyo@upc.edu

Abstract: *In liberalized electricity markets, Generation Companies must build an hourly bid that is sent to the market operator. The price at which the energy will be paid is unknown during the bidding process and has to be forecast. In this work we apply forecasting factor models to this framework and study its suitability.*

Keywords: *Electricity market prices, short term forecasting, factor models, price scenarios*

1 Introduction

In liberalized electricity markets, a Generation Company (GenCo) must build an hourly bid that is sent to the market operator, who selects the lowest price among the bidding companies in order to match the pool load. For this reason, GenCos that participate in liberalized electricity markets around the world need to know the prices at which the energy will be paid in order to decide how to bid and how to schedule their resources for maximizing their profit. The problem is that the market price is only known once the market has been cleared, so it is needed to forecast. In fact, is not only the price needed to forecast, but also its distribution.

Our main objective is to include short-term forecasting of the electricity market spot price in an optimization model for the management of a generation company in order to obtain realistic market price scenarios in which the generation company should decide how to optimally operate and to easily update these scenarios over time. Our approach has been applied to the Iberian Electricity Market.

The problem with building electricity price scenarios has been tackled within many areas. Non-parametric statistic methods -such as clustering or bootstrapping- applied to historical data were the first and simplest approaches. The advantage of these methods is that they are easy and computationally cheap to use but, on the other hand, they do not characterize the price distribution properly.

Electricity spot prices exhibit non-constant mean and variance, daily and weekly seasonality, calendar effects on weekends and holidays, high volatility and the presence of outliers. Those characteristics do not necessarily make it easy for electricity price short-term forecasting. Several approaches have been proposed in the power system literature which basically can be classified into parametric/nonparametric, conditional homoscedastic/heteroscedastic and others, ranging from the most popular ARIMA models belonging to the class of parametric-conditional homoscedastic models to the most sophisticated ones, as for example wavelet or neural networks models. In this field, Contreras et al. (2003) use ARIMA models for forecasting day-ahead electricity prices and Conejo et al. (2005) compare the goodness of the day-ahead electricity price forecasting using ARIMA models, dynamic regression, transfer function, wavelet-transform and neural networks for the PJM market, concluding that the predictions extrapolated from dynamic regression and transfer function procedures are better than those obtained from ARIMA models; whereas wavelet models have results close to ARIMA models and neural network algorithms do not offer good forecasts. Garcia-Martos et al. (2007) decompose the Spanish hourly time series electricity prices in 24 time series and model them separately obtaining the one day-ahead forecast for each time series. However the residuals in most of the analyzed models exhibit non stationary conditional variance. To solve this problem, the classical GARCH models and their variants are used for estimating the conditional heteroscedasticity of the

electricity spot prices. Garcia et al. (2005) estimate an ARMA model with GARCH errors for the Spanish and California Electricity market showing that this combined model overcomes the predictions obtained by the classical ARIMA model. Koopman et al (2007) give a more complex version of this model, extending it to periodic dynamic long memory regression models with GARCH errors also.

All of the models mentioned previously present advantages and drawbacks; nevertheless any one of them is absolutely convenient for our objective. Our approach in this work is to apply the well-known methodology of factors models to forecast electricity market prices in a short-term horizon (24 hours). In this case, the spot prices have been interpreted not as a single time series but a set of 24 time series, one for each hour, in a similar way to Munoz and Bunk (2007), Alonso et al. (2008) and Karakatsani and Bunn (2008). The factor model procedure allows us to identify common unobserved factors, which represent the relationship between the hours of a day. Despite the fact that dynamic and static factor models have been extensively used in many different frameworks (Geweke (1977), Stock and Watson (2002), Pena and Poncela (2004), Pena and Poncela (2006)), their application to short term electricity market prices forecasting has not been exploited. Previous results have shown that dynamic factor models are better for improving or extending them, but on the other hand, some authors conclude that these benefits are not sufficient to justify their use compared to the ease of estimating static factor models. For this first approach, static factor models have been applied and evaluated for the Iberian Electricity Spot prices.

The second part of our objective is to apply the results obtained from the presented forecasting technique to a stochastic optimization model for the management of a generation company operating in an electricity day-ahead market. The optimization model will be focused on the inclusion of physical derivatives products in the short-term management of a GenCo, following the model proposed in Corchero and Heredia (2009). Other approaches to similar problems involving futures contracts can be found in the works of Chen et al. (2004) and, in the medium-term optimization, Conejo et al. (2008) and Guan et al. (2008). The stochastic model is based on the representation through scenarios of the random variable involved in the problem. In our case, the stochastic variable is the day-ahead market clearing price. So, a set of scenarios for the day-ahead market clearing price will be built from the forecasting results. Once this set of scenarios is obtained, it is introduced into the optimization model, whereby the stability analysis and results are obtained.

The forecasting technique used results in a characterization of the market clearing price that is easy to estimate and to update. The results obtained in the forecasting part allow us to build realistic price scenarios that -once introduced into the model- provide suitable results for the optimization objective.

This paper is organized as follows: Section 2 contains the factor model used, the optimization procedures is in Section 3, the results in Section 4 and finally the conclusions are in Section 5.

2 Factors models

The estimation and forecasting price variables using factor analysis can be classified into two overarching groups: Static and Dynamic. The first uses Principal Component Analysis whereas the second basically formulates the model into State Space and uses the Kalman filter or the Expectation-Maximization (EM) algorithm for estimating the parameters and forecasting future values of the variable in question, the price, in our case.

In this work we use the alternative procedure *Time Series Factors Analysis (TSFA)* based on the previous work of Cattell *et al.* (1947) and described in Gilbert and Meijer (2005). The code is available in the R package *TSFA* available on CRAN. TSFA is an alternative for Standard Factor Analysis (FA) and Dynamic Factor Analysis (DFA). In the case of FA, it should not be used with

economic times series because the characteristics of the data do not agree with the assumptions of the method. Whereas TSFA differs from Dynamic Factor Analysis (DFA) in the sense that TSFA estimates parameters and predicts factor scores with few assumptions about factor dynamics; in particular, TSFA does not assume stationary covariance. DFA assumes a predetermined relationship between factors in the sense that there is an assumed *a priori* relationship between the factors at time t and the factors at time $t - 1$. If this relationship is misspecified, the factors estimated by DFA can be biased.

2.1 Factor Model Estimation

Let y_t be an M -vector of observed time series of length T and k unobserved factors ($k \ll M$) collected in the K -vector ξ . The relationship between the observed time series y_t and the ξ factors is assumed to be linear and described by equation (1)

$$(1) \quad y_t = \alpha_t + B\xi_t + \epsilon_t$$

where α_t is an M -vector of intercept parameters that can be omitted without losing generality, B is an $M \times k$ matrix parameter of loadings, assumed time-invariant, and ϵ is a random M -vector of measurement errors.

Defining D as the difference operator, (1) becomes:

$$(2) \quad Dy_t = \tau_t + BD\xi_t + D\epsilon_t$$

and the following conditions are assumed by Gilbert and Meijer (2005):

$$(3) \quad \sum_{t=1}^T \frac{D\xi_t}{T} \xrightarrow{d} \kappa$$

$$(4) \quad \sum_{t=1}^T \frac{(D\xi_t - \kappa)(D\xi_t - \kappa)'}{T} \xrightarrow{d} \Phi$$

$$(5) \quad \sum_{t=1}^T \frac{D\epsilon_t D\epsilon_t'}{T} \xrightarrow{d} \Omega$$

The sample covariance of the differenced series Dy_t is denoted by S_{Dy} and from the previous assumptions, it follows that¹:

$$(6) \quad S_{Dy} \xrightarrow{d} \Sigma \equiv B\Phi B' + \Omega$$

Parameters are estimated by maximum likelihood, minimizing the function:

$$(7) \quad L \equiv \lg \det \Sigma + \text{tr}(\Sigma^{-1} S_{Dy})$$

2.2 Forecasting Model

The factors obtained following the previous procedure have to be implemented into a forecasting model in order to obtain the price forecasts. Many authors (such as Schumacher (2007) or Stock and Watson (2002)) describe forecasting models suitable to either dynamic or static factors and to any factor estimation methods. The one-step-ahead forecasting model is specified and estimated as a linear multiple regression model with the factors as predictors, it has the form

$$(8) \quad y_{t+1} = \beta \hat{\xi}_t + \alpha(L)y_t + \varepsilon_{t+1}$$

¹See Gibert and Mejer (2005) for a complete description

where $\tilde{\xi}_t$ are the estimation of the factors, β is the regression coefficients matrix and ε_{t+1} is the resulting forecast error. Autoregressive terms are included through the polynomial of non-negative power of the lag operator L with coefficients $\alpha(L)$. The out of the sample forecasts for y_{T+1} conditional on information until period T is given by the conditional expectation $y_{T+1|T} = \hat{\beta}\hat{\xi}_T + \hat{\alpha}(L)y_T$.

Once the forecasting has been done, the set of scenarios needed for the optimization model has to be built. The set of scenarios consists of a set of possible values for the forecast variable, in this case the electricity prices, $\lambda_t^s = \{\lambda_1^s, \dots, \lambda_T^s\}$ and its corresponding probability $P^s = P(\lambda^s) \forall s \in S$. In this work the set of scenarios is built based on the discretization of the forecast confidence interval.

3 Stochastic programming optimization model

The optimization model used to test the energy price forecasting model is a simplification of the stochastic model described in Corchero and Heredia (2009). The main objective of this work is to build a stochastic programming optimization model which includes the coordination between *Physical Futures Contracts* (PFC) and *Day-Ahead Market* (DAM) bidding following the MIBEL rules. In the MIBEL, the market operator (OMEL) demands that every generation company (GenCo) commit the quantity designated to PFC through the DAM bidding of the physical units. This commitment is made by the so called *instrumental price offer*, that is, a sale offer with a bid price of 0€/MWh (also called *price acceptant*). That regulation implies that the *Generation Company* (GenCo) has to determine its optimal bid by taking into account those instrumental price offers. Due to the algorithm the market operator uses to clear the DAM, all instrumental price offers will be matched (i.e. accepted) in the clearing process, that is, this energy shall be produced and will be remunerated at the spot price λ_t^s €/MWh. Following MIBEL's rules, if we are optimizing today we focus on tomorrow's DAM because we have to submit tomorrow's bidding. Thus, the optimization horizon is at 24-hour intervals; this set of intervals is denoted as T . The proposed short-term bidding strategies are addressed to a price-taker GenCo, that is, a GenCo without market power. The generation units to be considered are the set I of thermal units with participation in the auction process. The relevant parameters of a thermal unit are:

- Quadratic generation costs with constant, linear and quadratic coefficients, c_i^b (€), c_i^l (€/MWh) and c_i^q (€/MWh²) respectively, for the unit $i \in I$.
- \bar{P}_i and \underline{P}_i the upper and lower bound, respectively, on the energy generation (MWh) of a committed unit $i \in I$.

The problem to be solved is in trying to maximize the expected value of the benefits coming from the DAM, which, at each scenario $s \in S$ can be calculated as the difference between the incomes from the matched energy, $\lambda_t^{d,s} p_{it}^s$, and the generations costs ($c_i^b + c_i^l p_{it}^s + c_i^q (p_{it}^s)^2$). This maximization must be done satisfactorily, for each PFC $j \in J$, a delivering of L_j MWh, and the DAM rules. The first-stage decision variables of this model are q_{it} , the energy of the instrumental price offer, that is, the energy bid by unit i to the t^{th} auction of the DAM at 0€/MWh, and variable f_{itj} , the energy delivered by the thermal unit i to the PFC j at period t . The second-stage variables are p_{it}^s , the matched energy of thermal i at the t^{th} auction of the DAM under scenario s . The mathematical expression of this two-stage stochastic programming problem is:

$$\begin{aligned}
(9) \quad & \underset{p,q,f}{\text{maximize}} \sum_{t \in T} \sum_{i \in U_t} \sum_{s \in S} P^s \left[\lambda_t^{d,s} p_{it}^s - (c_t^b + c_i^l p_{it}^s + c_i^q (p_{it}^s)^2) \right] \\
& \text{s.t.} \\
(10) \quad & \sum_{i \in U_{tj}} f_{itj} = L_j \quad t \in T, j \in F \\
(11) \quad & q_{it} \geq \sum_{j \in F_i} f_{itj} \quad i \in U_t, t \in T \\
(12) \quad & \underline{P}_i \leq q_{it} \leq p_{it}^s \leq \bar{P}_i \quad i \in U_t, t \in T, s \in S \\
(13) \quad & f_{itj} \geq 0 \quad i \in U_t, t \in T, j \in F
\end{aligned}$$

The sets appearing in this formulation are: the subset of PFC in which unit t participates (F_t); the set of thermal units that participates in contract j (T_j); the energy that has to be settled for contract j (L_j); the set of thermal units assigned to PFC j which are operating at period t (U_{tj}) and, finally, the periods where thermal unit t is operating (U_t). Constraint (10) ensures that the energy of the j^{th} PFC, L_j , will be completely dispatched among all the committed units. Constraints (11) formulate the MIBEL's rule that forces the energy of the future contracts to be bid through the instrumental price offer. Constraints (12) express the relations between the instrumental price offer q_{it} , the matched energy at scenario s , p_{it}^s , and the minimum and maximum generation levels \underline{P}_i and \bar{P}_i respectively. Finally, the nonnegativity on variables f_{itj} are expressed in (13).

4 Results

In this section both the results of the validation of the proposed forecasting method and its application to the optimization process are shown.

The variable to be forecasted is the Iberian Day-Ahead Market electricity prices. This is hourly data and the data set used corresponds to the work days from January 1st, 2007 to March 30th, 2008. This data is available at OMEL's site (www.omel.es). As has been described, the electricity prices have been interpreted not as a single time series but as a set of 24 time series, one for each hour. This 24 time series must be summarized by a small number of factors.

Following Gilbert and Meijer (2005) the number of factors is fixed based on the eigenvalues of the sample correlation matrix of indicators, in our case the number of significant factors is three. The loading matrix obtained is represented in Figure and its relationship can be observed in the the boxplot of hourly prices shown in Figure 2. The behavior of the prices throughout one day has a particular profile, with hours called "base hours", in which the price is low and there is lower variance. There are also hours called "peak hours", in which there are higher prices and high variance. The first factor separates clearly between night and day and it can be observed that the profile of the daily hours loads (between 8 a.m. and 8 p.m.) is similar to the profile of the prices during these hours. The second factor gives positive loads to the base hours and the third to the peak hours.

The forecasting model is based on these factors. The estimation of the 24 regression models is made with a subset of the available data. The R^2 and the Mean Square Error (MSE) of the forecast regression model for each hour are shown in Table 1. The estimated model is used to forecast the next 5 days. In Figure 2, the following are plotted: the real price (red line), the forecast price (black line) and the forecast confidence interval (blue line) used to build the set of scenarios. This forecasting procedure has been compared to the ARIMA model used in previous works (Corchero and Heredia (2009)) and it has been observed that the results in terms of MSE are equivalent.

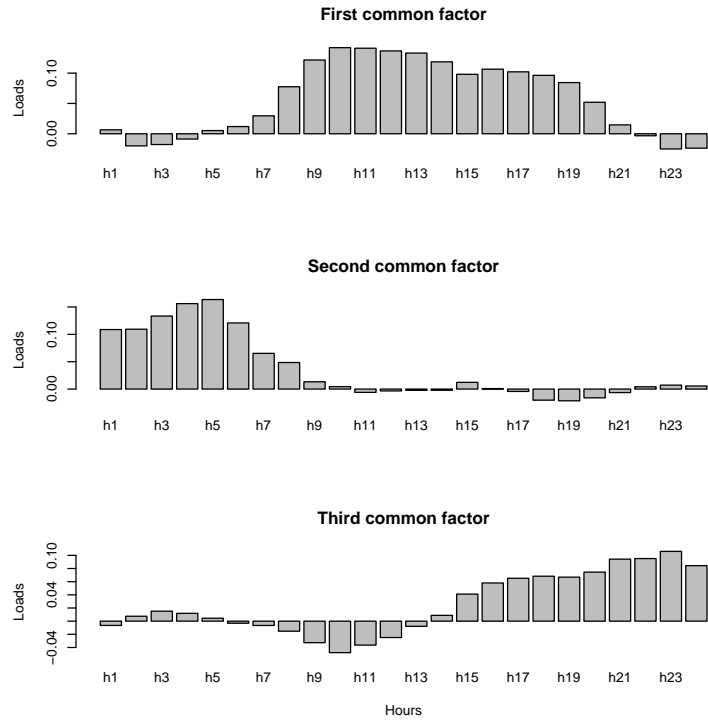


Figure 1: *Loads of the common factors*

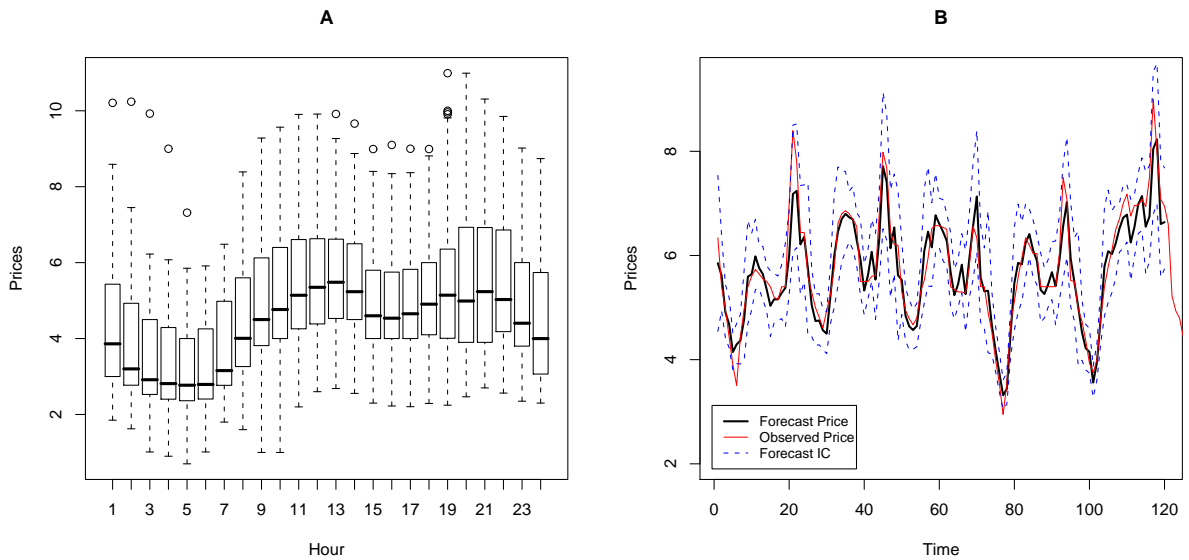


Figure 2: A. *Iberian Day-Ahead Electricity Market, (January 1st, 2007 - March 30th, 2008)*, B. *One-step-ahead forecast prices*

Hour	1	2	3	4	5	6	7	8	9	10	11	12
R^2	99.1	95.3	97.1	99.8	99.8	97.6	96.0	99.6	99.7	99.8	96.3	98.3
MSE	0.017	0.004	0.003	0.003	0.002	0.002	0.003	0.008	0.008	0.004	0.003	0.001
Hour	13	14	15	16	17	18	19	20	21	22	23	24
R^2	99.9	97.7	99.8	99.9	99.9	97.1	99.7	96.6	94.2	99.7	99.7	95.1
MSE	0.002	0.002	0.004	0.002	0.002	0.002	0.006	0.005	0.007	0.007	0.007	0.005

Table 1: *Summary of the forecast models for each hour*

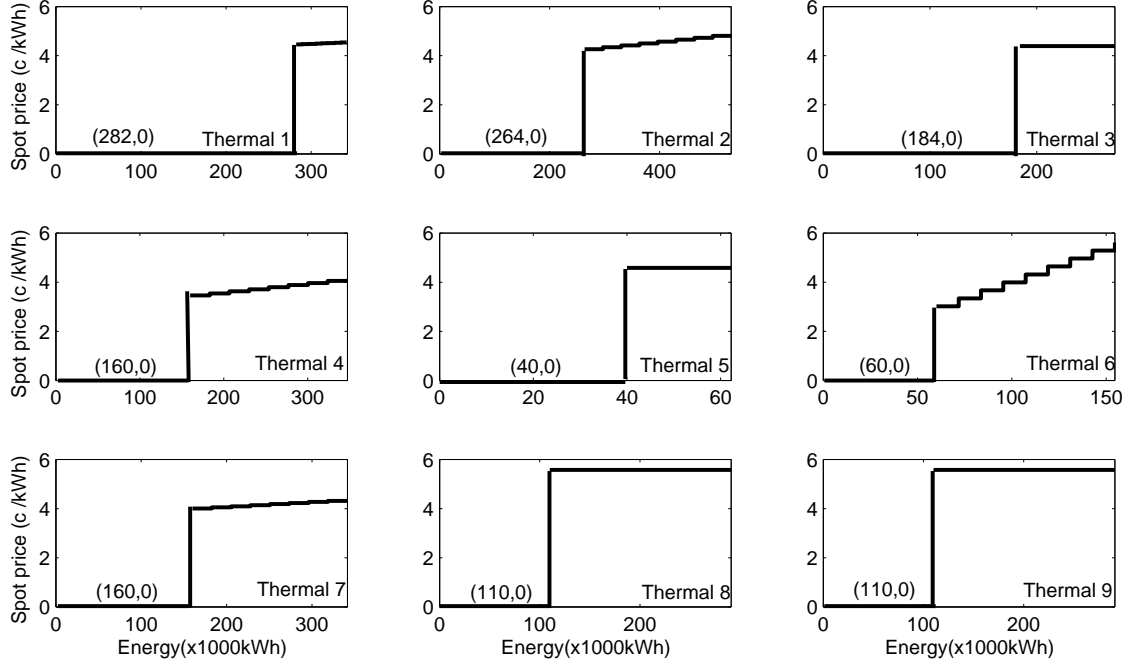


Figure 3: *Bidding curve for each unit at hour 20*

Finally, the set of scenarios has been built and introduced into the optimization model. Figure 3 shows the bidding curves obtained for each thermal unit at hour 20, built from the optimal value of the first stage variables (see Corchero and Heredia (2009)).

5 Conclusions

The forecast procedure based on factor models gives suitable results. These results are equivalent to the ones obtained through an ARIMA model but the advantage of the procedure presented in this work lies in its simplicity. The forecast model is easier to implement and to interpret than an ARIMA one. To build an ARIMA model for the electricity prices, a profound knowledge of times series identification is necessary, whereas such profound knowledge is not necessary for using this presented procedure. This advantage facilitates the implementation of the models automatically, so that companies can use it regularly.

From the optimization point of view, the improved forecasts have been used to successfully generate a set of scenarios to feed the stochastic optimization model (9)-(13). This set of scenarios is also built in an automatic way from the forecast confidence interval. The optimal solution of this

model provides the optimal bid to be sent to the market operator for each thermal unit at each of the 24 day-ahead market auctions.

REFERENCES

- Alonso A.M., Garcia-Martos C., Rodriguez J., Sanchez M.J. (2008). Seasonal Dynamic Factor Analysis and Bootstrap Inference: Application to Electricity Market Forecasting. *Working Paper 08-14. Statistics and Econometric Series 06*, Universidad Carlos III de Madrid.
- Contreras J., Espinola R., Nogales J., Conejo A.J. (2003). ARIMA Models to Predict Next-Day Electricity Prices. *IEEE Transactions on Power Systems*, 18(3): 1014-1020.
- Conejo A.J., Contreras J., Espinola R., Plazas M.A. (2005). Forecasting electricity prices for a day-ahead pool-based electric energy market. *International Journal of Forecasting*, 21: 435-462.
- Conejo, A.J., Garcia-Bertrand, R., Carrion, M., Caballero, A., de Andres, A. (2008). Optimal involvement in futures markets of a power producer. *IEEE Transactions on Power Systems* 23(2): 703-711.
- Corchero C., Heredia F.J. (2009). A Stochastic Programming Model for the Thermal Optimal Day-Ahead Bid Problem with Physical Futures Contracts. Submitted to *European Journal of Operations Research*. DR 2009/03 DEIO, UPC. <http://hdl.handle.net/2117/2795>
- Chen, X., He, Y., Song, Y. H., Nakanishi, Y., Nakahishi, C., Takahashi, S., Sekine, Y. (2004). Study of impacts of physical contracts and financial contracts on bidding strategies of gencos. *Electrical Power & Energy Systems*, 26: 715-723.
- Garcia R.C., Contreras J., van Akkeren M., Garcia J.B.C. (2005). A GARCH Forecasting Model to Predict Day-Ahead Electricity Prices. *IEEE Transactions on Power Systems*, 20(2): 867-874.
- Garcia-Martos C., Rodriguez J., Sanchez M.J. (2007). Mixed Models for Short-Run Forecasting of Electricity Prices: Application for the Spanish Market. *IEEE Transactions on Power Systems*, 22(2): 544-552.
- Geweke J. (1977). The dynamic factor analysis of economic time series. *Latent variables in socio-economic models*. Amsterdam: North Hollan.
- Gilbert P.D., Meijer E. (2005). Time Series Factor Analysis with an Application to Measuring Money. *Research Report No. 05F10*. University of Groningen, SOM Research School.
- Guan, X., Wu, J., Gao, F., Sun, G. (2008). Optimization-based generation asset allocation for forward and spot markets. *IEEE Transactions on Power Systems*, 23(4): 1796-1807.
- Karakatsani N.V., Bunn DW. (2008). Forecasting Electricity Prices: The Impact of Fundamentals and Time-Varying Coefficients. *International Journal of Forecasting*, 24: 764-785.
- Koopman S.J., Ooms M., Carnero M.A. (2007). Periodic Seasonal Reg-ARFIMA-GARCH Models for Daily Electricity Spot Prices. *Journal of the American Statistical Association*, 102 (477): 16-27.
- Munoz M.P., Bunk D.W. (2007). Covariates of Stochastic Volatility in Electricity Prices. *International Symposium on Business and Industrial Statistics*. University of Azores.
- Pena D., Poncela P. (2004). Forecasting with Nonstationary Dynamic Factor Models. *Journal of Econometrics*, 119: 291-321.
- Pena D., Poncela P. (2006). Nonstationary Dynamic Factor Analysis. *Journal of Statistical Planning and Inference*, 136: 1237 - 1257.
- R Development Core Team (2008). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing* (<http://www.R-project.org>)
- Shumacher C. (2007). Forecasting German GDP Using Alternative Factor Models Based on Large Datasets. *Journal of Forecasting*, 26: 271-302.
- Stock J., Watson M.W. (2002). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association*, 97 (460): 1167-1179.