Supplementary material for "Multivariate exponential smoothing and dynamic factor model applied to hourly electricity price analysis" published by Technometrics

Jaime Carpio, Jesús Juan, Damián López

Data and Matlab functions

In this paper, we have analyzed the hourly electricity prices in the Omel (Spain), Powernext (France) and Nord Pool (Denmark, Finland, Norway and Sweden) day-ahead markets. The time series of data are given in the files Omel.dat, Powernext.dat and Nordpool.dat. To read in Matlab one file, for instance the data corresponding to the Omel market, we write:

$>> \mathbf{y} = load('omel.dat');$

In the data files, we have stored the logarithm of the hourly electricity prices for each market. In particular, $\mathbf{y}$ is a matrix, where the rows are the prices for the $r = 24$ hours of the day, and the columns indicate the length of the time series, $N = 780$ week-days, equivalent to 156 weeks from January 1, 2007 to December 25, 2009.

Section 3: Analysis of price dynamics

In order to obtain the results presented in Section 3, we use the data set from January 1, 2007 to May 29, 2009 (126 weeks). Therefore, the data used to estimate are the 630 first days $\mathbf{y\_est} = \mathbf{y}(:, 1 : 630)$

1. In subsection "*3.1. Estimated the full model*", we estimate the matrices of the following model:

   $$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{x}_t + \mathbf{w}_t, & \mathrm{var}(\mathbf{w}_t) &= \mathbf{Q}. \\ \mathbf{y}_t &= \mathbf{H}\mathbf{x}_t + \boldsymbol{\mu}_0 + \mathbf{v}_t, & \mathrm{var}(\mathbf{v}_t) &= \mathbf{R}, \end{aligned}$$

   where the state vector has dimension $n = 24$. The model will be estimated with the well-known EM algorithm written in "*em_algorithm.m*". This function use the Kalman and Kalman smoother filtering programmed in *"kalman_filter2s.m"* and "*kalman_smoother2s.m*", respectively. The main program to estimate the model is "*run_models.m*", which should be run in Matlab with the following sentence:

   $>> mc = run\_models(\mathbf{y\_est}, 24, 1);$

   where 24 is the order of the model and 1 is the number of explanatory variables or regressors.

   The output variable mc is a data structure, which stores the most relevant information of the model

   $>> \mathbf{Q} = mc.Q;$

$>> \mathbf{R} = mc.R;$

$>> \mathbf{H} = mc.H;$

$>> \boldsymbol{\mu}_0 = mc.G;$

The matrices $\boldsymbol{B}$, and $\boldsymbol{\Sigma}$, related to the exponential smoothing formula (1) can be obtained from function "$dfa.m$". The nonsingular matrix $\mathbf{T}$ and the diagonal variance matrix $\boldsymbol{\Lambda}$, useful to diagonalize all matrices (as indicated in Proposition 2) are also computed by the same function:

$>> [\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{T}, \boldsymbol{\Lambda}] = dfa(\mathbf{Q}, \mathbf{H}, \mathbf{R});$

2. In subsection "*3.2. Order reduction*", we estimate the model:

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{x}_t + \mathbf{w}_t, & \mathrm{var}(w_t) &= \mathbf{Q}. \\ \mathbf{y}_t &= \mathbf{H}\mathbf{x}_t + \mathbf{G}\mathbf{m}_t + \mathbf{v}_t, & \mathrm{var}(v_t) &= \mathbf{R}. \end{aligned}$$

The state vector has length from $r = 1$ to 24. The models have been computed with function "$run\_models.m$", but now, using any order r of the model.

$>> for\ r = 1 : 24$

$model(r) = run\_models(\mathbf{y\_est}, r, 5);$

end

where 5 is the number of explanatory variables due to the different behavior for each day of the week. The estimated matrices for a model of dimension r can be obtained from Matlab with:

$>> \mathbf{Q} = model(r).Q;$

$>> \mathbf{R} = model(r).R;$

$>> \mathbf{H} = model(r).H;$

$>> \mathbf{G} = model(r).G;$

The data structure has also the following variables:

$>> m(r).AIC$: the Akaike Information Criterion given by equation (22).

$>> m(r).SIC$: the Schwarz Information Criterion given by equation (25).

$>> m(r).np$: the total number of parameters given by equation (24).

$>> m(r).loglikelihood$: the likelihood function given by equation (23) in each iteration of the EM algorithm until convergence.

For each model of order r can be also obtained the matrices $\mathbf{B}$, $\boldsymbol{\Sigma}$, $\mathbf{T}$, and $\boldsymbol{\Lambda}$ with the function "$dfa.m$".

Section 4: Forecasting accuracy

In order to obtain the results presented in Section 4, we are going to use the data set from June 1, 2009 to December 25, 2009 (30 weeks). The electricity price forecasting $\hat{\mathbf{y}}_t$ is computed by function "$forecast.m$".

$>>$ $\mathbf{yhat} = forecast(\mathbf{y}, B, G)$;

y in the last expression is the complete time series vector. However, we are only interested in the last $30 \times 5 = 150$ week-days. Thus, the predicted prices $\hat{\mathbf{y}}_t$ computed from June 1, 2009 to December 25, 2009 are $\mathbf{yhat}(:, 631 : 631 + 150)$.