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

Maximum likelihood with EM algorithm

The Expectation-Maximization (EM) algorithm (Shumway and Stoffer, 2006) is an iterative procedure for computing the maximum likelihood estimation for the unknown parameters in the state space model of the paper given by:

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

As information to make parameter estimation, we have measurement vectors $\mathbf{y}_t$ for each time $t = 1, ...N$, which we can store in the set $\mathbf{Y}_N$

$$\mathbf{Y}_N = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N\}. \tag{1}$$

The state-space formulation of the EM algorithm considers compete random vectors $\boldsymbol{Z}_N = \{\boldsymbol{X}_N, \boldsymbol{Y}_N\}$ made up of the observed set $\mathbf{Y}_N$ of vectors and the unobserved state vectors $\mathbf{X}_N = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\}$. Then, the joint density function is given by

$$f(\mathbf{Z}_N|\boldsymbol{\theta}) = \prod_{t=2}^{N} f_{\mathbf{Q}}(\mathbf{x}_t|\mathbf{x}_{t-1}) \prod_{t=2}^{N} f_{\mathbf{R},\mathbf{H},\mathbf{G}}(\mathbf{y}_t|\mathbf{x}_t), \tag{2}$$

where $\boldsymbol{\theta}$ denotes the set of unknown parameters $\boldsymbol{\theta} = (\mathbf{Q}, \mathbf{R}, \mathbf{H}, \mathbf{G})$, and $\mathbf{x}_1$ is assumed to be known with zero value $\mathbf{x}_1 = 0$. Here, the log-likelihood $l_{\mathbf{Z}_N}(\boldsymbol{\theta}) = \log f(\mathbf{Z}_N|\boldsymbol{\theta})$ is preferred because information is combined by addition and it can be written as the sum of two uncoupled functions

$$l_{\mathbf{Z}_N}(\theta) = l_1(\mathbf{Q}) + l_2(\mathbf{R}, \mathbf{H}, \mathbf{G}), \tag{3}$$

where under the Gaussian assumption

$$\begin{aligned} l_1(\mathbf{Q}) &= -N \log|\mathbf{Q}| - \sum_{t=2}^{N} (\mathbf{x}_t - \mathbf{x}_{t-1})^T \mathbf{Q}^{-1} (\mathbf{x}_t - \mathbf{x}_{t-1}). \\ l_2(\mathbf{R}) &= -N \log|\mathbf{R}| - \sum_{t=2}^{N} (\mathbf{y}_t - \mathbf{H}\mathbf{x}_t - \mathbf{G}\mathbf{m}_t)^T \mathbf{R}^{-1} (\mathbf{y}_t - \mathbf{H}\mathbf{x}_t - \mathbf{G}\mathbf{m}_t). \end{aligned} \tag{4}$$

The EM algorithm provides an iterative method for finding the maximum likelihood estimators of parameter $\boldsymbol{\theta}$, by successively maximizing the conditional expectation of the complete likelihood given by (3). Each iteration of the EM algorithm consists of two steps: the first one (E step) supposes $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_j$ known (the estimated values of the parameter $\boldsymbol{\theta}$ at previous iteration j), and computes the conditional expectation

$$E[l_{\mathbf{Z}_N}(\boldsymbol{\theta})|\mathbf{Y}_N, \hat{\boldsymbol{\theta}}_j]. \tag{5}$$

The second one (M step) consists in maximizing it. The main advantage of the EM algorithm is due to the fact that the unknown parameters can be obtained from explicit regression formulas at each iteration.

E-step: computation of $E[l_{\mathbf{Z}_N}(\boldsymbol{\theta})|\mathbf{Y}_N, \hat{\boldsymbol{\theta}}_j]$

Here, given the values of parameters $\hat{\boldsymbol{\theta}}_j$ (matrices estimated at previous iteration j) and $\mathbf{Y}_N$, the only terms that remain random data in equation (4) are the states $\mathbf{X}_N$. Then, to compute the expected values of (3) with the information of $\mathbf{Y}_N$ and $\hat{\boldsymbol{\theta}}_j$ we need to provide for $t = 2, 3, .., N$ the values

$$\begin{aligned}
&\mathbf{x}_{t|N} = E[\mathbf{x}_t|\mathbf{Y}_N, \hat{\mathbf{Q}}_j, \hat{\mathbf{R}}_j, \hat{\mathbf{H}}_j, \hat{\mathbf{G}}_j],\\
&\mathbf{P}_{t|N} = E[(\mathbf{x}_t - \mathbf{x}_{t|N})(\mathbf{x}_t - \mathbf{x}_{t|N})'|\mathbf{Y}_N, \hat{\mathbf{Q}}_j, \hat{\mathbf{R}}_j, \hat{\mathbf{H}}_j, \hat{\mathbf{G}}_j],\\
&\mathbf{P}_{t,t-1|N} = E[(\mathbf{x}_t - \mathbf{x}_{t|N})(\mathbf{x}_{t-1} - \mathbf{x}_{t-1|N})'|\mathbf{Y}_N, \hat{\mathbf{Q}}_j, \hat{\mathbf{R}}_j, \hat{\mathbf{H}}_j, \hat{\mathbf{G}}_j],
\end{aligned} \tag{6}$$

and from them it is possible to compute $E[l_{\mathbf{Z}_N}(\boldsymbol{\theta})|\mathbf{Y}_N, \hat{\boldsymbol{\theta}}_j]$ as:

$$E[l_{\mathbf{Z}_N}(\boldsymbol{\theta})|\mathbf{Y}_N, \hat{\boldsymbol{\theta}}_j] = E[l_1(\mathbf{Q})|\mathbf{Y}_N, \hat{\boldsymbol{\theta}}_j] + E[l_2(\mathbf{R}, \mathbf{H}, \mathbf{G})|\mathbf{Y}_N, \hat{\boldsymbol{\theta}}_j]. \tag{7}$$

$$\begin{aligned}
E[l_1(\mathbf{Q})|\boldsymbol{Y}_N, \hat{\boldsymbol{\theta}}_j] = -N\log|\mathbf{Q}| - \mathrm{tr}\Big\{\mathbf{Q}^{-1}\sum_{t=2}^{N}\big[&\mathbf{P}_{t|N} + \mathbf{P}_{t-1|N} - \mathbf{P}_{t,t-1|N} -\\
&-\mathbf{P}'_{t,t-1|N} + \left(\mathbf{x}_{t|N} - \mathbf{x}_{t-1|N}\right)\left(\mathbf{x}_{t|N} - \mathbf{x}_{t-1|N}\right)'\big]\Big\}.
\end{aligned}$$

$$E[l_2(\mathbf{R}, \mathbf{H}, \mathbf{G})|\mathbf{Y}_N, \hat{\boldsymbol{\theta}}_j] = -N\log|\mathbf{R}| - \mathrm{tr}\left\{\mathbf{R}^{-1}\sum_{t=2}^{N}\left[\mathbf{H}\mathbf{P}_{t|N}\mathbf{H}' + \left(\mathbf{y}_t - \mathbf{H}\mathbf{x}_{t|N} - \mathbf{G}\mathbf{m}_t\right)\left(\mathbf{y}_t - \mathbf{H}\mathbf{x}_{t|N} - \mathbf{G}\mathbf{m}_t\right)'\right]\right\}. \tag{8}$$

where tr{} is the trace operator of a matrix.

M-Step: maximization of $E[l_{\mathbf{X}_N,\mathbf{Y}_N}(\boldsymbol{\theta})|\mathbf{Y}_N, \hat{\boldsymbol{\theta}}_j]$

Maximizing $E[l_{\mathbf{X}_N,\mathbf{Y}_N}(\boldsymbol{\theta})|\mathbf{Y}_N, \hat{\boldsymbol{\theta}}_j]$ with respect to the parameters $\boldsymbol{\theta}$, at iteration $j+1$, constitutes the M-step and it is analogous to the multivariate regression approach.

The maximum of $E[l_1(\mathbf{Q})|\mathbf{Y}_N, \hat{\mathbf{Q}}_j]$ is attained at

$$\begin{aligned}
\hat{\mathbf{Q}}_{j+1} = \frac{1}{N-1}\sum_{t=2}^{N}\Big[&\mathbf{P}_{t|N} + \mathbf{P}_{t-1|N} - \mathbf{P}_{t,t-1|N} - \mathbf{P}'_{t,t-1|N} +\\
&+ \left(\mathbf{x}_{t|N} - \mathbf{x}_{t-1|N}\right)\left(\mathbf{x}_{t|N} - \mathbf{x}_{t-1|N}\right)'\Big].
\end{aligned} \tag{9}$$

From $E[l_2(\mathbf{R}, \mathbf{H}, \mathbf{G})|\mathbf{Y}_N, \hat{\boldsymbol{\theta}}_j]$, we obtain the estimator of $\mathbf{H}$, $\mathbf{G}$, and $\mathbf{R}$:

$$\left\{\begin{aligned}
&\hat{\mathbf{H}}_{j+1}\left[\sum_{t=2}^{N}\left(\mathbf{P}_{t|N} + \mathbf{x}_{t|N}\mathbf{x}'_{t|N}\right)\right] + \hat{\mathbf{G}}_{j+1}\left[\sum_{t=2}^{N}\mathbf{x}_{t|N}\mathbf{m}'_t\right] = \left[\sum_{t=2}^{N}\mathbf{y}_t\mathbf{x}'_{t-1|N}\right],\\
&\hat{\mathbf{H}}_{j+1}\left[\sum_{t=2}^{N}\mathbf{m}_t\mathbf{x}'_{t|N}\right] + \hat{\mathbf{G}}_{j+1}\left[\sum_{t=2}^{N}\mathbf{m}_t\mathbf{m}'_t\right] = \left[\sum_{t=2}^{N}\mathbf{y}_t\mathbf{m}'_t\right].
\end{aligned}\right. \tag{10}$$

$$\hat{\mathbf{R}}_{j+1} = \frac{1}{N-1}\sum_{t=2}^{N}\left[\mathbf{H}\mathbf{P}_{t|N}\mathbf{H}' + \left(\mathbf{y}_t - \mathbf{H}\mathbf{x}_{t|N} - \mathbf{G}\mathbf{m}_t\right)\left(\mathbf{y}_t - \mathbf{H}\mathbf{x}_{t|N} - \mathbf{G}\mathbf{m}_t\right)'\right]. \tag{11}$$

To calculate $\hat{\mathbf{Q}}_{j+1}$, $\hat{\mathbf{H}}_{j+1}$, $\hat{\mathbf{G}}_{j+1}$ and $\hat{\mathbf{R}}_{j+1}$ form formula (9)-(11), we need $\mathbf{x}_{t|N}$, $\mathbf{P}_{t|N}$, and $\mathbf{P}_{t,t-1|N}$ obtained with the Kalman Smoother filter. To do so, we compute the Kalman filter forward in time and after that, apply the Kalman Smoother filter backward in time. We summarize the iterative procedure as follows:

EM algorithm

Given a maximum number of iterations $j_{\max}$*, a tolerance* ε*,*
and an initial set of parameters $\hat{\boldsymbol{\theta}}_0 = \left\{ \hat{\mathbf{Q}}_0, \hat{\mathbf{R}}_0, \hat{\mathbf{H}}_0, \hat{\mathbf{G}}_0 \right\}$.
$j = 1$.
$\delta l = 10^{10}$.
While $(\delta l > \varepsilon \;\; or \;\; j < j_{\max})$
1. *Perform the E-step:*
    Use the Kalman Filter and Kalman Smoother to obtain:
    $\mathbf{x}_{t|N}$, $\mathbf{P}_{t|N}$ *and* $\mathbf{P}_{t,t-1|N}$*, for* $t = 1, ..., N$*, with parameters* $\hat{\boldsymbol{\theta}}_j$ *known.*
2. *Perform the M-step:*
    Use the explicit formula (9), (10), and (11) to compute
    $\hat{\boldsymbol{\theta}}_{j+1} = (\hat{\mathbf{Q}}_{j+1}, \hat{\mathbf{R}}_{j+1}, \hat{\mathbf{H}}_{j+1}, \hat{\mathbf{G}}_{j+1})$ *to update the value of the estimated parameters.*
3. *Compute:*
    $\delta l = |l_{\mathbf{Z}_N}(\hat{\boldsymbol{\theta}}_{j+1}) - l_{\mathbf{Z}_N}(\hat{\boldsymbol{\theta}}_j)| / |l_{\mathbf{Z}_N}(\hat{\boldsymbol{\theta}}_{j+1})|$.
    $j = j + 1$.

end