Multivariate Exponential Smoothing and Dynamic Factor Model Applied to Hourly Electricity Price Analysis

Jaime CARPIO, Jesús JUAN, and Damián LÓPEZ

Dpto. Ing. de Organizació, Administración de Empresas y Estadística, Universidad Politécnica de Madrid, Spain (jaime.carpio@upm.es; jjuan@etsii.upm.es; damian.lopez@upct.es)

Thanks to its very simple recursive computing scheme, exponential smoothing has become a popular technique to forecast time series. In this work, we show the advantages of its multivariate version and present some properties of the model, which allows us to perform a dynamic factor analysis. This analysis leads to a simple methodology to reduce the number of parameters (useful when the dimension of observations is large) via a linear transformation that decomposes the multivariate process into independent univariate exponential smoothing processes, characterized by a single smoothing parameter that goes from zero (white-noise process) to one (random walk process). A computer implementation of the expectation-maximization (EM) algorithm has been built for the maximum likelihood estimation of the models. The practicality of the method is demonstrated by its application to hourly electricity price predictions in some day-ahead markets, such as Omel, Powernext, and Nord Pool markets, whose forecasts are given as examples. This article has supplementary material online.

KEY WORDS: EWMA; Forecasting; Moving average processes; State space methods; Time series.

1. INTRODUCTION

Jones (1966) introduced multivariate exponential smoothing and demonstrated its potential for forecasting multiple time series. His proposal was the multivariate generalization of the exponentially weighted moving average (EWMA) model whose usefulness and simplicity is widely proven in the literature on prediction since the earlier works of Holt (1957), Brown (1959), and Winters (1960). The simple and famous exponential smoothing recursion to make predictions is

$$\hat{y}_{t+1} = \hat{y}_t + B(y_t - \hat{y}_t),$$

(1)

where $y_t \in \mathbb{R}^n$ is the observation vector at time $t$, $\hat{y}_t$ is the predicted value for $y_t$ with the available information until instant $t-1$, and $B \in \mathbb{R}^{n \times n}$ is the matrix version of the exponential smoothing weight. The prediction errors $\varepsilon_t = y_t - \hat{y}_t$, also denoted as innovations, are a sequence of uncorrelated random variables with zero mean and steady variance matrix $\Sigma \in \mathbb{R}^{n \times n}$.

Taking (1) and writing $\hat{\varepsilon}_t = y_t - \varepsilon_t$, we obtain the multivariate IMA(1,1) process

$$y_{t+1} = y_t + \varepsilon_{t+1} - \Theta \varepsilon_t, \quad \text{var}(\varepsilon_t) = \Sigma,$$

(2)

where $\Theta = I - B$ is the $n \times n$ matrix corresponding to the MA(1) component and $I$ is the identity matrix. The multivariate exponential smoothing or IMA(1,1) is one of the basic models of the two most common approaches to time series: the Box–Jenkins vector autoregressive integrated moving average (VARIMA) models (Reinsel 1993), and the state space models (Harvey 1989). In this article, we use a restricted form of Model (2) to analyze the hourly electricity price dynamics in day-ahead markets where the observation vector $y_t$ has a dimension $n = 24$.

Electricity markets have experienced a significant transformation around the world in the last few decades. The price of electricity has gone from being set by the government every year to being entirely determined by the market every day. In most countries, the price of electricity is arranged through a daily auction for each 24 hr of the next day. This pricing mechanism introduces huge uncertainty, which means economic risk for the agents involved in the market: producers, sellers, banks, and consumers. For all of them, although in very different ways, having an accurate as possible idea of future prices is very useful.

The literature on modeling and analyzing electricity prices is abundant and is growing quickly. Most of the initial literature is in the finance area and focuses on developing realistic spot price models, analytically tractable for the purposes of derivative pricing and risk management. During the last two decades, many statistical techniques and models have been developed for forecasting wholesale electricity prices, specially for short-term price forecasting. The complexity of electricity price dynamics can be seen in Koopman, Ooms and Carnero (2007), one of the most complete studies on the subject.

Weron’s book (2006) presents an extensive review of the established approaches to modeling and forecasting electricity loads and prices. This book and many other articles summarize the factors influencing price behavior and compare different approaches, such as artificial neural networks, ARIMA models,
Looking specifically at univariate time series models, we can distinguish two different approaches in the literature. A single model is not sufficient to collect all the peculiarities of price dynamics. Conejo et al. (2005), one of the most referenced articles in this matter, used a different model for each season for the PJM Interconnections (Pennsylvania–New Jersey–Maryland). The other approach, which is preferred by several authors, is to use different models for each of the 24 hr in a day (Misiorek, Trück, and Weron 2006; Weron 2006). In some cases, differentiating between weekdays and weekends is recommended, with the result that 48 models are used (García-Martos, Rodríguez, and Sánchez 2007). The structure of these models varies from 1 hr to another. Therefore, when the literature uses the term univariate model it actually refers to multiple univariate models.

The direct application of multivariate ARIMA models requires a large number of parameters. For instance, the IMA(1,1) (2), one of the simplest models needs the estimation of a $24 \times 24$ square matrix plus the $24 \times 24$ error covariance matrix. If the day-of-the-week seasonal effect is included, a new $24 \times 24$ matrix will be required, and the multivariate final model will accumulate a huge number of total parameters. Reducing the parameter space is essential for successfully modeling multivariate time series. The problem is solved in the time series literature by dimension reduction using dynamic factor models. Dynamic factor models have been studied by Engle and Watson (1981), Peña and Box (1987), Stock and Watson (1988), Peña and Poncela (2004, 2006), among others (see these articles for further references); other works specifically devoted to the analysis of the electricity prices are Härdle and Trück (2010) and Alonso et al. (2011), to name a few.

To keep the model as simple as possible and, at the same time, to solve the most difficult problem, we have made some simplifications. Carefully bearing in mind the price formation process, the following aspects should be specified:

1. All the 24 hourly prices are established and obtained at the same time in only one take. Within a day the prices are not produced sequentially, so the internal correlations (dependencies) between the 24 prices do not necessarily have to follow a temporal logic. In fact, the correlations between equally distant hours are different for different pairs of hours. Due to this lack of sequentiality, it is not advisable to realize a univariate treatment of the whole hourly series but rather to study it as a multivariate one. The same observation has been made in Huisman, Huurman, and Mahieu (2007).

2. Auctions are actually only held on weekdays. The auction for Tuesday is held on Monday, for Wednesday on Tuesday, and so on until Friday when prices are set for the three following days, Saturday, Sunday, and Monday. Put another way, to set the prices for Monday agents cannot use the information from the previous Saturday and Sunday, because they do not have it. They make the Monday bid with the information available up to and including Friday. So, the data corresponding to Saturday and Sunday can be (or should be) deleted from the sequence without causing discontinuities. Given the above, it makes more sense to work with the working-day series (Monday to Friday) and remove Saturday and Sunday, which greatly simplifies the dynamics of the process. Furthermore, the study of working-day prices is the most important one due to the economic volume they represent.

According to the above two points, in this work the analysis of electricity price dynamics is performed as a 24 hr multivariate model. The autocorrelation function and the partial autocorrelation function corresponding to the differentiated pricing log series of each hour in the three markets studied show that univariate MA(1) is a suitable model for all the hours. Therefore, the 24 hourly series of logarithms of the prices can be jointly considered as a multivariate IMA(1,1) process (2). This model simplifies the identification phase of the estimation process. Although the model requires many parameters, the estimation of multivariate models presents no computational difficulty.

The layout of this article is as follows: in the next section, we will explain the role of the state space model in exponential smoothing. State space model estimation presents certain advantages and it is also possible to connect a sequence of more efficient models with fewer parameters that shed light on some features of price dynamics. In Section 3, we use the results obtained in the previous section to characterize price evolution in three different electricity markets: Powernext (France), Nord Pool (Denmark, Finland, Norway, and Sweden), and Omel (Spain). At the end of the section, we discuss the performance of the model as a function of the common factors obtained. In Section 4, we evaluate the predictive ability of the proposed model and compare it with the precision of some other univariate models. Finally, our conclusions are presented in Section 5.

2. STATE SPACE REPRESENTATION OF THE EXPONENTIAL SMOOTHING MODEL

The multivariate exponential smoothing process (Jones 1966; Harvey 1989) is commonly written in the state space form via a random walk plus noise model, whose equations are

\[
\begin{align*}
\mathbf{x}_{t+1} &= \mathbf{x}_t + \mathbf{w}_t, \quad \text{var}(\mathbf{w}_t) = \mathbf{Q}, \\
\mathbf{y}_t &= \mathbf{x}_t + \mathbf{v}_t, \quad \text{var}(\mathbf{v}_t) = \mathbf{R},
\end{align*}
\]

(3)

where we have added $\mathbf{x}_t$, the unobserved state vector; $\mathbf{w}_t$, the excitation, which is a sequence of uncorrelated random vectors with zero mean and variance matrix $\mathbf{Q}$; and $\mathbf{v}_t$, called the error, a sequence of uncorrelated random vectors with zero mean and variance matrix $\mathbf{R}$, which are also uncorrelated with $\mathbf{w}_t$. Vectors $\mathbf{x}_t, \mathbf{w}_t, \mathbf{v}_t$ have the same dimension $n$, as $\mathbf{y}_t$, and the dimension of both variance matrices $\mathbf{Q}, \mathbf{R}$ is of $n \times n$. Note, however, that each of these matrices contains $n(n + 1)/2$ parameters, while the Model (2) has $n^2 + n(n + 1)/2$ parameters: the elements of matrices $\mathbf{A}$ and $\mathbf{B}$. In the multivariate case, $n > 1$, the number of parameters of Model (2) is greater than in Model (3). As we will see in next sections, to make (2) and (3) equivalent models, it is necessary to impose certain restrictions on the matrix $\mathbf{A}$. It means that Model (2) and Model (3) are not identical; any model of type (3) can be written as (2), but not vice versa. There are several alternatives to represent the state space dynamics that
underlie the predictions given by the exponential smoothing Equation (1), for instance, the innovations vector state space model (see Hyndman 2008). It is equivalent to an unrestricted IMA(1,1) model. The use of the traditional state space model, Model (3), rather than the innovations state space model, reduces the number of parameters to be estimated and allows dynamic factor analysis as we discuss below.

However, to use the model to make predictions, the most simple equation is (1). Therefore, it is useful to obtain matrices B and Σ of (1) from matrices Q and R of Model (3). We may employ for this goal the methodology employed by Jones (1966) or Harvey (1989). However, we are going to present here a new way to obtain B and Σ from Q and R.

**Proposition 1.** The matrices Θ and Σ of Model (2) are uniquely determined by the parameters Q and R of Model (3) using the relations:

\[ \Theta = R(P + R)^{-1} \]

(4)

and

\[ \Sigma = P + R, \]

(5)

where P verifies

\[ P = R(P + R)^{-1}P + Q. \]

(6)

**Proof.** By considering (3),

\[ y_t - y_{t-1} = x_t - x_{t-1} + v_t - v_{t-1} = w_t + v_t - v_{t-1}, \]

it is easy to see that \( y_t - y_{t-1} \) has a covariance structure \( \Gamma(0) = Q + 2R, \Gamma(1) = -R, \) and \( \Gamma(h) = 0 \) for \( h > 2. \) The autocovariance function of \( y_t - y_{t-1} \) according to (2) is

\[ \Gamma_2(h) = \begin{cases} \Sigma + \Theta \Sigma \Theta' & h = 0 \\
-\Theta \Sigma & h = 1 \\
-\Sigma \Theta' & h = -1 \\
0 & |h| \geq 2 \end{cases} \]

which is not in general symmetric. Making equal \( \Gamma(1) \) and \( \Gamma_2(h), \) the relations between parameters in the two model forms are

\[ Q + 2R = \Sigma + \Theta \Sigma \Theta'. \]

(7)

and

\[ R = \Theta \Sigma = \Sigma \Theta'. \]

(8)

The relations (7) and (8) uniquely determine \( \Theta \) and \( \Sigma \) in terms of Q and R, or stating it in a much more useful form, we can derive the exponential smoothing parameters B = I - \( \Theta \) and \( \Sigma, \) from Q and R.

Under the definition, \( P = \Sigma - R \) (which actually, it corresponds to the steady-state error covariance matrix associated with the prediction of the state vector in the Kalman filter, see, for instance, Shumway and Stoffer (2006), we can write (8) as

\[ R = \Theta(P + R), \]

from which we get result (4).

Finally, substituting \( R = \Theta \Sigma = \Sigma \Theta' \) in (7) and taking into account that \( P(P + R)^{-1} + R(P + R)^{-1} = I, \) we obtain

\[ Q = (I - \Theta) \Sigma (I - \Theta)' = P(P + R)^{-1} \]

\[ = P - R(P + R)^{-1}P, \]

which is the result (6).

Given Q and R, the computation of P can be obtained from (6) in an iterative way. Starting by \( P_1 = Q, \) the successive application of \( P_{k+1} = R(P_k + R)^{-1}P_k + Q \) converges in a few iterations to the solution of the problem. These iterations are equivalent to the steps performed in the Kalman filter that Jones (1966) applied to obtain matrices B and Σ from (4) and (5), respectively.

**Remark 1.** For Model (3) to be equivalent to the restricted form of Model (2) requires that product \( \Theta \Sigma \) be an symmetric and positive definite matrix. This guarantees that \( R = \Theta \Sigma \) is a variance matrix. As \( \Sigma \) is a positive definite matrix, this implies that the eigenvalues of \( \Theta \) must also be positive. Throughout the treatment we have used matrix R has full rank, implying that both \( \Theta \) and \( \Sigma \) must have full rank. If the eigenvalues of \( \Theta \) are less than unity, the model is invertible. Given Q and R, the proposition allows us to obtain the invertible Model (2) equivalent to Model (3).

### 2.1 Dynamic Factor Analysis

In this section, we will show a more informative way to obtain matrices B and Σ taking advantage of the diagonalization properties of matrices Q and R. A linear transformation of vector \( y \), decomposes the multivariate process into independent univariate exponential smoothing processes. The transformation reveals interesting and meaningful features of the multivariate dynamic model and, at the same time, shows the way to reduce the number of parameters of the model.

Assuming that R is positive definite matrix, it is well known that there exists a nonsingular matrix \( T \) that simultaneously diagonalizes Q and R, so that

\[ TQT' = \Lambda \quad \text{and} \quad TRT' = I, \]

(9)

A being a diagonal matrix with all \( \lambda_i \) positive elements \( \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n), \) and I the identity matrix. These values are the eigenvalues of the matrix \( R^{-1}Q \) that we will assume to be ordered, \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n. \)

Multiplying by \( T \) the state and observation Equation (3), we get new states \( f_t = Tx_t \) and new observations \( z_t = Ty_t, \)

\[ f_{t+1} = f_t + a_t, \]

\[ z_t = f_t + e_t, \]

(10)

where noise vectors \( a_t \) and \( e_t \) are formed by independent random variables with variance matrices \( A \) and I, respectively. Equation (10) highlights the characteristics of the transformation. The new states \( f_{t,i} \) evolve according to independent random walks, each of them with different variance, \( \lambda_i. \) This value \( \lambda_i \) measures the contribution of \( f_{t,i} \) to the multivariate process variability. Following the classic factor analysis, we can call these new states dynamic factors. This dynamic factor analysis is
standardized, so the obtained eigenvalues remain invariant under change of scale and rotation of the observations $y_t$. As we will see later in the examples, among the 24 factors obtained from the electricity prices only a few of them are responsible for most of the variability.

Matrices $T$ and $A$ also provide a different way of getting $P, B,$ and $\Sigma$ of those given by Proposition 1. These results are collected in the following proposition.

**Proposition 2.** Assuming $R$ is symmetric and positive definite matrix, there exists a nonsingular matrix $T$ such that $\Sigma = I$, and

(a) $TQ = T = \Lambda$, with $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. $\lambda_i \geq 0$.
(b) $TPT = \Omega$, with $\Omega = \text{diag}(\omega_1, \omega_2, \ldots, \omega_n)$, $\omega_i \geq 0$ and $\omega_1 = (\lambda_i + \sqrt{\lambda_i^2 + 4\lambda_i})/2$.
(c) $TST = \Delta$, with $\Delta = \text{diag}(\delta_1, \delta_2, \ldots, \delta_n)$. $\delta_i > 0$ and $\delta_i = (2 + \lambda_i + \sqrt{\lambda_i^2 + 4\lambda_i})/2$.
(d) $TBD = D$, with $D = \text{diag}(d_1, d_2, \ldots, d_n)$, $0 \leq d_i < 1$ and $d_i = (-\lambda_i + \sqrt{\lambda_i^2 + 4\lambda_i})/2$.

**Proof.** Part (a) is obtained by using the decomposition (9). Given the nonsingular matrix $T$,

\[(P + R)^{-1} = T(TPT + I)^{-1}T\]  

and substituting in (6), we have

\[P = RT(TPT + I)^{-1}TP + Q.\]

Premultiplying by $T$ and postmultiplying by $T^T$ and calling $\Omega = TPT$, we obtain

\[\Omega = (\Omega + I)^{-1}\Omega + \Lambda\]  

that can be written as $\Omega^2 - \Lambda \Omega - \Lambda = 0$. This equation has a solution that is diagonal with all positive elements, which is obtained from solving the $n$ univariate quadratic equations to give part (b).

Substituting now (11) in (4), premultiplying by $T$, postmultiplying by $T^T$, and calling $D = TB$, we have

\[D = TPT(TPT + I)^{-1} = \Omega(\Omega + I)^{-1},\]

so $D$ is also diagonal, with positive values between zero and one. The columns of $T$ are the eigenvectors of $B$. Moreover, given (12) we can write $\Omega = D + \Lambda$, so each diagonal element of $D$ is given by part (d).

Finally, the innovation variance matrix fulfills $\Sigma = P + R$, and calling $\Lambda = TST$, it can be written as

\[\Lambda = I + \Omega = I + D + \Lambda.\]

and therefore $\delta_i = 1 + \omega_i$ that proof part (c).

Proposition 1 allows us to move from Model (3) to Model (2). Knowing $Q$ and $R$, then $T$ and $A$ are obtained, and from these matrices and the results (c) and (d) of Proposition 1, we can obtain $\Sigma = T^T\Lambda(T^T)^{-1}$ and $B = T^TDT$. Moreover, now the observations $y_t$ can be seen as a linear combination $y_t = T^x_t$, of independent univariate exponential smoothing processes, and the $i$ component of the observation vector $z_{i,t}$ can be predicted by an independent equation obtained from (1)

\[\hat{z}_{i,t+1} = \hat{z}_{i,t} + d_i(z_{i,t} - \hat{z}_{i,t}).\]  

Each exponential smoothing process has a different degree of smoothing $1 - d_i$, obtained by means of part (d) of Proposition 1, whose value is between 0 and 1. When $d_i = 0$, the component $z_{i,t}$ is just a constant observed with noise, and when $d_i = 1$, the component $z_{i,t}$ became a pure random walk. This last component dictates the long-term behavior of the multivariate process. The innovations $e_{i,t} = z_{i,t} - \hat{z}_{i,t}$ correspond to random errors that are independent with variance $\delta_i$, that allow us to write (13) in a MA(1,1) form, as (2), but now also uncoupled

\[z_{i,t+1} = z_{i,t} + e_{i,t+1} - (1 - d_i)e_{i,t}, \quad \text{var}(e_{i,t}) = \delta_i,\]  

Note that in the above equations, the four values $\lambda_i, d_i, \omega_i,$ and $\delta_i$ are related and only one among them is truly independent. In proposition 1, the parameters $d_i, \omega_i,$ and $\delta_i$ are obtained from $\lambda_i$. We can also put the parameters as a function of $d_i$, since $\lambda_i$ can be written in term of $d_i$ as $\lambda_i = d_i^2/(1 - d_i)$, the following relations $\delta_i = 1/(1 - d_i), \omega_i = d_i/(1 - d_i)$ are easily deduced. Therefore, the dynamics of each independent process $z_{i,t}$ is fully characterized by a single parameter $d_i$, $0 < d_i < 1$.

### 2.2 Common Factor Model

The application of the previous analysis to the hourly electricity prices, where the time series are vectors of dimension $n = 24$, provides many $\lambda_i$ values practically equal to zero, more than in any of the three studied markets. This is a case of great interest in practice, in which the dynamics of multivariate $y_t$ of dimension $n$ is governed by a smaller number of factors $r < n$. We can split the state vector $f_t$ in (10) into two parts: $g_t$ of dimension $r$ corresponding to nonzero variance components, and $o_t$ of dimension $n - r$, corresponding to zero variance components, that is, to say, $f_t = (g_t, o_t)$, with $o_{t+1} = o_t = c$, where $c$ is a constant vector. The components of the transformed series $z_t$ in (10) are sorted in the same way, so that the last $n - r$ components are white noise with constant mean $c$. The transformed observation vector is obtained as $z_t = Ty_t$; note that $T$ is a nonsingular matrix, so $y_t = T^{-1}z_t$. To simplify the notation, we call $H = T^{-1}$, and $H$, the matrix of dimension $n \times r$ formed by the $r$ first columns of $H$, and $H_{n-r}$, the $n \times (n-r)$ matrix formed with the rest of the columns. Calling $A_r$, the diagonal matrix of dimension $r \times r$ formed by nonzero $\lambda_i$ values, then $Q = HA_rH^T$, and $T^{-1}f_t = Hg_t + H_{n-r}o_t = Hg_t + \mu_0$, and the Model (3) can be written as

\[g_{t+1} = g_t + b_t, \quad \text{var}(b_t) = A_r,\]

\[y_t = Hg_t + \mu_0 + v_t, \quad \text{var}(v_t) = R,\]  

where $g_t$ is a lower-dimensional $r \times 1$ random walk vector and the variance matrix of the error term $b_t$ is $A_r$. Note that multiplying the state equation in (15) on the left by $H$, and calling $x_t = Hg_t + \mu_0$, we get the state equation of Model (3). These factors $g_t$ are called “common factors,” which represent a reduced number of “common trends” among the components of the series $y_t$ of the observation vector (Engle and Yoo 1987; Stock and Watson 1988; Harvey 1989; Reinsel 1993).

The parameter $\mu_0$ is new in the above formulation, and its presence may seem surprising. However, this vector also appears in all previous formulations, (2), (3), and (1), implicitly. Note that if one wants to apply the recurrence formula, it is necessary to provide a starting point for all of them. In Model
\(\text{(15)}, \mu_0\) represents the initial level price for each hour of the day. If we consider null the initial state vector \(g_0 = 0\). Jones (1966) recommended the first data of the observation vector \(y_1\) as an estimate of the initial level. In this work, we are going to estimate \(\mu_0\) jointly with \(Q\) and \(R\), although its value will be close to \(y_1\). Fortunately, the election of this initial vector is not crucial if the time series is large enough and here, its influence in the estimation of the known matrices, \(Q\) and \(R\), is negligible. Moreover, this constant vector \(\mu_0\) can be replaced by a vector \(\mu\), deterministically variable over time and it can be used to introduce explanatory variables or regressors as we will discuss later.

Apart from \(\mu_0\), Model (15) has three unknown matrices, the variance matrices \(\Lambda\) and \(R\), and the matrix \(H\). This model offers some advantages when \(r\), the number of common factors, is small. In that case, the number of parameters is considerably reduced, especially if \(n\) is large as it happens in the case of electricity prices. In this case of uncomplete rank of \(Q\), the observation equation of Model (15) has a direct link with the concept of co-integration (see Engle and Yoo 1987; Reinsel 1993; Escribano and Peña 1994). There exists a matrix \(V\) of dimension \(n \times (n - r)\), which fulfills \(V^*H = 0\) and therefore, multiplying the observation equation of Model (15) by \(V^*\) it follows \(V^*y\), which is a constant plus a white-noise process of dimension \(n - r\). That is, although each of the \(n\) series of vector \(y\) is nonstationary, \(n - r\) linear combinations of them are stationary. This imposes certain restrictions on price developments, establishing linear relations between them, to progress together and it limits the differences between them. On the other hand, if \(r\) is much smaller than \(n\), the number of parameters to be estimated is greatly reduced. From the standpoint of prediction avoiding overfitting can improve the predictions. As we will see in the electricity examples, the common trends lead to interesting interpretations of the process dynamics.

2.3 Special Case

Another case that has received some attention in the literature (Enns et al. 1982; Harvey 1986, 1989) is when the state variance matrix is proportional to the observation variance matrix, \(Q = \lambda R\), with \(\lambda > 0\). Interestingly, if we count the number of model parameters in this case, we get \(n(n + 1)/2\) for \(R\) and only one more for \(\lambda\) and, as this number should match the number of parameters in the exponential smoothing Equation (1), the number of degrees of freedom for matrix \(B\) is just one! The matrix of generalized eigenvalues in this case is \(\Lambda = \lambda I\), so it is trivial to prove, in accordance with the above expressions, that \(B = dI\) with \(d = (-\lambda + \sqrt{\lambda^2 + 4\lambda})/2\). So the weight matrix in (1) is diagonal with all elements equal to \(d\), the whole matrix \(B\) is condensed into a single value \(d\), as was expected according to the number of parameters and \(\hat{y}_{t+1} = \hat{y}_t + d(y_t - \hat{y}_t)\). Note that in this case, the original observations are updated independently with the same smoothing parameter. As \(Q\) is proportional to \(R\), the matrices \(P\) and \(\Sigma\) are proportional to \(R\) as well, \(P = \omega R\) and \(\Sigma = (1 + \omega)R\), where \(\omega = (\lambda + \sqrt{\lambda^2 + 4\lambda})/2\). In this case, the dynamic multivariate process is characterized by \(n\) independent random walks all with the same variance \(\lambda\). This property does not apply to electricity price data, so we will not use this model.

3. ESTIMATION AND PRICE DYNAMIC ANALYSIS

The multivariate exponential smoothing has been applied to analyze the evolution of the hourly electricity prices in three different markets: Powernext (France), Nord Pool (Denmark, Finland, Norway, and Sweden), and Omel (Spain). The observed series \(y_1, y_2, \ldots, y_N\) are taken as the logarithm of the prices, to reduce the heteroscedasticity of the time series.

The most simple and convenient way to make predictions and calculate their uncertainties within the framework of the model presented in this work is by means of (1). Jones (1966) proposed a method to obtain \(B\), using the successive estimate of autoregressive multivariate models. However, this is not necessarily the most appropriate one at the estimation stage. The alternative is to take advantage of the enormous power and resources that state space formulation provides for estimation of matrices \(Q\) and \(R\) in Model (3) or \(\Lambda\), \(H\), and \(R\) for the reduced model of order \(r\) (15) and then, to relate them with matrices \(B\) and \(\Sigma\) of the exponential smoothing model. There are several algorithms in the literature to estimate state space models by maximum likelihood (Harvey 1989; Shumway and Stoffer 2006). Probably the easiest algorithm to implement is the one based on the expectation-maximization method (EM). The EM algorithm is a well-known procedure, which for the multivariate exponential smoothing model has a particularly neat form. The main drawback is its slow convergence near the maximum, but this drawback is not a critical problem, especially if you have a good starting solution available as is the case. An excellent description of this method can be found in Shumway and Stoffer (2006). A brief explanation of the EM algorithm and the MATLAB functions along with the time series data used in this work are available as supplementary material.

The strategy that we follow in this section is the following: first, estimating the full-order Model (3) for the three markets with the week-days from January 1, 2007, to May 29, 2009 (126 weeks of 5 days). A study of the eigenvectors and eigenvalues of the matrix \(R^{-1}Q\) allows us to analyze the electricity price dynamics. As usual, only a very few common factors account for a high proportion of the total variability of the multivariate process. The dynamic factor analysis explained in Section 2 is very useful to reduce the order of the model. A reduced final model can be reestimated using the representation (15) or simply by removing the factors found to be negligible and rewriting the model accordingly.

3.1 Estimating the Full Model

Here, we estimate the matrix \(Q\) and \(R\) for each of the three markets using the EM algorithm for Model (3). It is important to note that in the examples analyzed, the hourly residual autocorrelation functions show no significant correlation structure. In some hours, a slightly significant coefficient appears at lag 5 due to the weekly effect. (We will see some modifications in Section 3.2 to eliminate this weekly effect.) The dynamic factor analysis has been performed for the three markets and the variance \(\lambda_i\) of the unobserved factor \(f_{t,i}\) (random walk) of Model (10) are shown in Table 1. In this table, we also have the \(d_i\) values to simplify the factor analysis interpretation, although a single
parameter is sufficient to characterize the dynamics, as we can see in Proposition 2.

From data of Table 1, we can obtain the following conclusions regarding the three electricity markets.

1. According to the variances shown, Nord Pool has a completely different behavior from the other two markets. Nord Pool has a total variance of 45.19 in contrast to the values of 7.30 and 5.25 obtained for Omel and Powernext, respectively. Therefore, Nord Pool has more volatility than the other two markets.

2. The first factor is always the most important and it determines the main features of price evolution in the medium and long term. For Nord Pool, this component seems to be a random walk \((d_1 = 0.97, \text{ virtually } 1)\). However, for the other two markets, the variances of the first factor indicate a smoother evolution of the prices, with a degree of smoothing \(1 - d_1\) close to 0.2 in both cases.

3. Differences between the three markets is also observed if we evaluate the percentage of the total variability explained with the first factors. In Nord Pool, 91.4% is explained with the first two factors and 96.6% with the first three factors. However, in Powernext only 70.5% is explained with two factors, and 80.8% is explained with three factors. With Omel, the differences are even more remarkable, since the first two and three factors explain 58.2% and 72.1% of the variability, respectively.

4. The evolution of the price in Nord Pool is mostly explained (91.4% of the total variance) with the use of the first two factors, whose values \(d_1 = 0.97\) and \(d_2 = 0.92\) are similar to those of random walks. According to economic theory, the evolution of the price of a product in a perfect market should evolve like a random walk. Since Nord Pool is a market with higher trading volume, open, and competitive, this would explain, at least to a certain extent, the obtained behavior.

5. Table 1 shows a sharp change in the factor variances after the 14th factor in Nord Pool, the 15th one in Powernext, and the 16th one in Omel. Final factors present negligible variance, practically zero (values smaller than 10\(^{-9}\)), corresponding to linear combinations of the prices that remain constant with time. The model can be greatly reduced by removing these constant factors and decreasing the number of parameters to be estimated. The final number of factors to be used depends on the ultimate goal of the study. Ten factors is an unusually high number for this type of analysis; for a simplified model explaining the evolution of multivariate prices, one would choose typically two or three factors (or components). This point will be clarified in Section 3.2.

The values of the two first columns of the matrix \(H\) are plotted in Figure 1 for each market. In classic factor analysis, these 24 values are called loads or weights. Column one (solid line) represents the load of the first factor in each of the 24 hr, and they are positive in all three markets, although they are slightly

![Figure 1. Plot of the loads associated with the first and the second factors (first two columns of matrix H) for the three electricity markets analyzed in this article.](image-url)
different. A change in the first factor \( f_{1,t} \), which is the most relevant in terms of explained variability, affects all 24 hr in the same way; that means that an increment in the first factor leads to an increase in the electricity prices for all the hours of the day. The load of the second factor (dashed line), which is less important than the first one, has a different behavior depending on the market. For example, in Omel the load vector takes negative values for the early and the last hours of the day, and positive values around the hour 9. However, in Powernext the negative weights happen only at the early hours of the day, with important positive values in a peak around hour 19. These forms imply that an increment in the second factor \( f_{2,t} \) enlarges the difference between the prices of the hours with positive weights, and those hours having negative weights, leaving the price constant in the hours with negligible weight.

### 3.2 Order Reduction

A simple inspection of \( \lambda_i \) values in Table 1 indicates that the dynamics of multivariate process can be approximated by a number of factors \( r \) well below the dimension \( n \) of the process. Given \( r \), the estimation of this new reduced Model (15) can be done using directly the EM algorithm to obtain the three required unknown matrices \( \Lambda_r \), \( \mathbf{R} \), and \( \mathbf{H} \). The common factor Model (15) reduces significantly the number of parameters to estimate, which leads to a parsimonious model that generates effective predictions.

The common factor Model (15) is a good starting point, and serves as a nucleus of more sophisticated models that incorporate the particularities of the problem being studied. For instance, in the electricity price problem, the influence of the day of the week has an important effect in the prediction error. Analyzing the model residuals shows that Monday is the week day with the largest error, which seems to be consistent with the logic of auctions, where the average prices in the early hours of Monday are significantly larger than the rest of the days in the week. The same applies to the final hours of Fridays. We can take advantage of the constant \( \mu_0 \) in the observation equation of Model (15) and we can consider a different constant for each day of the week, from Monday to Friday. In prediction terms, the information of the day of the week improves the results, but with the disadvantage of a significant increase in the number of parameters of the model. Then, the observational equation can be slightly modified to incorporate a vector of \( p = 5 \) explanatory variables \( \mathbf{m}_t = \mathbf{Gm}_t \) or deterministic terms,

\[
\begin{align*}
\mathbf{g}_{t+1} &= \mathbf{g}_t + \mathbf{b}_t, \quad \text{var}(\mathbf{b}_t) = \Lambda_r, \\
\mathbf{y}_t &= \mathbf{H}_r \mathbf{g}_t + \mathbf{Gm}_t + \mathbf{v}_t, \quad \text{var}(\mathbf{v}_t) = \mathbf{R}_r
\end{align*}
\]

where \( \mathbf{G} \) is a new matrix with dimension \( n \times p \). Thus, creating a variable for each day of the week, \( j = 1 \) for Monday, \( j = 2 \) for Tuesday, and so on to \( j = 5 \) for Friday, we can define \( \mathbf{m}_t = (m_{1,t}, m_{2,t}, \ldots, m_{5,t})^T \) with

\[
m_{j,t} = \begin{cases} 
1, & \text{mod}(t-j,5) = 0, \\
0, & \text{mod}(t-j,5) \neq 0,
\end{cases}
\]

where \( \text{mod}(t-j,5) \) means modulus after division \( t - j \) by 5. Therefore, the \( p = 5 \) columns of matrix \( \mathbf{G} \) will be the related with the initial price of the electricity from Monday to Friday in the \( n = 24 \) hr of the day. The analysis of the residuals shows again that the model is quite suitable; now, for almost all hours of the multivariate model, the coefficients of the autocorrelation function of the residuals, associated with lag 5 (weekly effect), are no longer significant.

Given the Model (16), the next step is to select the most appropriate order \( r \). The order \( r \) can be determined by looking at the eigenvalues \( \lambda_i \). This result is consistent with a theoretical study performed by Billah, Hyndman, and Koehler (2005), where information criterion approaches are used for automatized model selection. The order \( r \) chosen is the one, which minimizes the Akaike information criterion (AIC)

\[
\text{AIC}_r = -2 \log[\hat{L}(\hat{\theta}_r, \mathbf{Y}_N)] + 2K_r,
\]

where the first term in the right-hand side of (18) is the log-maximum likelihood estimation obtained at \( \hat{\theta}_r = (\hat{\Lambda}_r, \hat{\mathbf{R}}_r, \hat{\mathbf{H}}_r, \hat{\mathbf{G}}_r) \) with the data \( \mathbf{Y}_N = \{y_1, y_2, \ldots, y_N\} \) and the second term (the “penalty” term) is a measure of the complexity of the model used, determined by the number of parameters in the model:

\[
K_r = \left(\frac{2n - r + 1}{2}\right) r + np + \frac{n(n+1)}{2}.
\]

The criterion suffers from one commonly observed drawback: it has a tendency to favor high-dimensional models. Examples illustrating this phenomenon appear in Shumway (1988, p. 169), Cavanaugh and Shumway (1997), and Linhart and Zucchini (1986), who comment that “in some cases the criterion simply continues to decrease as the number of parameters in the approximating model is increased.” Thus, to overcome the overestimation problem of AIC, Shumway (1988) recommended the use of Schwarz information criterion (SIC), where the “penalty” term in AIC is replaced by \( K_r \log(N) \). We have also considered the criterion proposed by Bai and Ng (2002) for factor models of large dimension. A small simulation study shows that these criteria applied to problems of size 24 tends to underestimate the model order.

For each market, we have analyzed all orders from \( r = 1 \) to \( r = 24 \). The AIC and SIC for the three markets are represented in Figure 2. The AIC suffers the inconveniences cited above. The SIC gives us similar results to the one obtained by simple inspection of the eigenvalues \( \lambda_i \) of Table 1, where the values corresponding to indexes above 10 or 11 are negligible.

The AIC selected model is optimal if we are interested in minimizing prediction error. A simple and commonly used way to check the goodness of a time series model is proving its ability to predict future observations. This is discussed in the next section and the results show that the order selected by both procedures is similar.

### 4. Forecasting Accuracy

When a time series model is selected, economists and statisticians would like to evaluate the technique for its ability to forecast accurately. As mentioned previously, the models were estimated with the data corresponding to the week-days from January 1, 2007 to May 29, 2009 (126 weeks of 5 days). To evaluate the forecasting performance of our model, we have set aside a subset of our data to serve as a test (from June 1, 2009 to
December 25, 2009, 30 weeks of 5 days). To get an idea of the accuracy of model predictions, Figure 3 shows the forecasts in two very different weeks in Powernext market and their 95% prediction intervals (the prediction intervals are based on the Gaussian distribution of model errors, $\varepsilon_t = y_t - \hat{y}_t \sim N(0, \Sigma)$).

The left part of the figure shows a very acceptable adjustment, which corresponds to a week when prices maintained a stable profile. In the right part of the figure, forecasting is much worse than in the top part, because prices experienced a large change from 1 day to another. However, the proposed method is very robust; Monday in right panel of Figure 3 is a clear outlier day and the exponential smoothing model gives a wrong prediction; nevertheless, after a short period of time (2 or 3 days), the model behaves again as expected, ignoring the false signal. By the end of the week shown at the bottom of Figure 3, the prediction follows a pattern once more. These different behaviors are repeated in all three analyzed markets. We think that this robustness is a remarkably good property of our model.

We have use to make predictions the multivariate exponential smoothing Model (16) for different orders from $r = 3$ to $r = 24$. Moreover, we have also considered, as other authors did, another naive models, which are also a special case of our full model: (i) the model $\hat{y}_t = \mathbf{G}m_t$ is a pure regression model in which the dynamic part is eliminated $r = 0$. (ii) The model $\hat{y}_t = y_{t-1}$, which predicts for tomorrow the values observed today, it corresponds to $\mathbf{B} = \mathbf{I}$ (the identity matrix) in (1), and in the state space form (3) the model implies $\mathbf{R} = 0$.

In Table 2, we compare the predictions obtained with different models using the root mean square error, $\text{RMSE} = \sqrt{\sum_{t=1}^{N} \varepsilon_t^2} / N$. In this table, we have removed the 1% of the highest hourly values of $\varepsilon_{i,t}$ to avoid the outlying observations in the evaluation of the performance of the model.

Once more, the model with $r = 10$ appears to be the best one, in agreement with the eigenvalues analysis of Section 2.1 and the SIC criterion used in the above section. Both naive models show worse behavior in all three markets, as expected (these two models were included among the candidates in the optimization process). The good performance of model $\mathbf{B} = \mathbf{I}$ may seem surprising. This result has been mentioned by other authors (Conejo et al. 2005; Weron 2006; Alonso et al. 2011), and is consistent with the estimated models, and with the eigenvalue analysis of the full model: in the context of electricity price markets, the achievement of a 10% improvement in predicting implies a considerable economic benefit. The model assumptions have been checked by residual analysis. Apart from the existence of outliers, the analysis shows some other minor
anomalies such as changes in the process variance. However, these small deviations do not substantially modify the main conclusions obtained in relation to the evolution of mean electricity prices.

We also want to present a comparative study of our methodology with two other short-term electricity price forecasting techniques, to analyze the predictive accuracy of our method. These models are: (A) a complex seasonal ARIMA model for the hourly price series as proposed by Conejo et al. (2005). (B) The other approach, which is preferred by several authors, is to use different ARIMA models for each of the 24 hr in a day, used by Misiorek, Trück, and Weron (2006) and García-Martos, Rodríguez, and Sánchez (2007).

To compare our predictions with the results obtained with the univariate models used for short-term hourly price forecasting, we use the mean week error (MWE) which is defined as

$$MWE_s = \frac{1}{120} \sum_{i=1}^{5+r} \sum_{j=1}^{24} |y_{i,j} - \hat{y}_{i,j}|, \quad s = 1, \ldots, 30,$$

(20)

where $y_{i,j}$, $\hat{y}_{i,j}$ are the observed and predicted hourly log-price and $\bar{y}$ is the mean log-price for a week. Figure 4 shows MWE corresponding to the 30 forecasted weeks in each market; please note that we performed the comparison in logarithmic scale to simplify the graphs. In this analysis, we kept all data as recorded, without removing any of the outliers, to illustrate the large variations observed between weeks. As described in this article, and also mentioned extensively in the literature, some material.

5. CONCLUSIONS

Short-term prediction (1 day ahead) of electricity prices is an issue of great importance in the liberalized electricity market. Electric utilities and large consumers apply it daily to make decisions about their strategy at the spot market. During the last few years, several models have been proposed, which are based essentially on the univariate analysis of the hourly price series. This strategy presents some drawbacks, as mentioned in the first section of this work. The most important drawback is the huge complexity of the models, necessitating different equations for different hours or for different seasons of the year.

A multivariate model is presented in this work. It is widely used for predicting in some fields, and is known as exponential smoothing. The model has excellent properties regarding robustness and versatility, and its application is easy. Among multivariate models, it is the simplest one to estimate. The EM algorithm has been used to estimate the parameters by maximum likelihood. The application to vectors of dimension 24, as in hourly electricity prices, presents the need of a large number of parameters, which is a disadvantage. However, in this work, a method for reducing the dimension is given. The estimated models give rise to a very interesting interpretation of price dynamics. This interpretation is related to the nonobserved factor models, common trends, and co-integration.

The model has been applied to the study of the behavior of three multivariate series: the French market (Powernext); the Scandinavian market (Nord Pool); and the Spanish market (Omel). The obtained results are consistent for all of them. The models allow extensions to add seasonality and deterministic effects. The estimated models reflect price dynamics, and residual analysis bears this out.

![Figure 4. MWE corresponding to 30 weeks (01/06–25/12 2009) for each market calculated with exponential smoothing Model (16) order $r = 10$, Model A, and Model B.](image-url)

Table 2. Root mean square error (RMSE) for the three markets for different models

| Model        | $r = 3$ | $r = 7$ | $r = 10$ | $r = 15$ | $r = 20$ | $r = 24$ | $\hat{y}_s = y_{s-1}$ | $\hat{y}_s = Gm_s$ |
|--------------|---------|---------|---------|---------|---------|---------|---------------------|---------------------|
| OMEL         | 0.306   | 0.261   | 0.256   | 0.259   | 0.259   | 0.258   | 0.299               | 0.748               |
| POWERNEXT    | 0.457   | 0.423   | 0.417   | 0.416   | 0.416   | 0.416   | 0.506               | 1.024               |
| NORD POOL    | 0.198   | 0.183   | 0.182   | 0.180   | 0.180   | 0.180   | 0.205               | 0.799               |
SUPPLEMENTARY MATERIALS

The following supplementary materials accompany the main article:

**Data and Matlab functions:** The set of data and the Matlab functions used in this article, as well as the instructions to use them can be found in “Data-Matlab-functions.zip.” (zip file).

**EM algorithm:** A brief explanation of the algorithm, used to estimate the multivariate exponential smoothing model, is shown in “EM-algorithm.pdf.” (pdf file).

**Comparison between models:** The comparison between methodologies showed in Section 4 is explained in greater detail in “Comparison-between-models.pdf.” (pdf file).

ACKNOWLEDGMENTS

This research was supported by Project MTM 2005-08896 from the Spanish Ministry of Science and Innovation. The authors thank Daniele Vallesi, Juan Temboury, and Francisco Pérez-Thodén for their aid in the preliminary stages of this work.

[Received April 2013. Revised October 2013.]

REFERENCES

Alonso, A., García-Martos, C., Rodríguez, J., and Sánchez, M. J. (2011), “Seasonal Dynamic Factor Analysis and Bootstrap Inference: Application to Electricity Market Forecasting,” Technometrics, 53, 137–151. [495,501]

Bai, J., and Ng, S. (2002), “Determining the Number of Factors in Approximate Factor Models,” Econometrica, 70, 191–221. [500]

Billah, B., Hyndman, R. J., and Koehler, A. B. (2005), “Empirical Information Criteria for Time Series Forecasting Model Selection,” Journal of Statistical Computation and Simulation, 75, 831–840. [500]

Brown, R. G. (1959), Statistical Forecasting and Prediction of Discrete Time Series, Englewood Cliffs, NJ: Prentice Hall. [494]

Cavanaugh, J. E., and Shumway, R. H. (1997), “A Bootstrap Variant of AIC for State-Space Model Selection,” Statistica Sinica, 7, 473–496. [500]

Conejo, A. J., Contreras, J., Espinola, R., and Plazas, M. A. (2005), “Forecasting Electricity Prices for a Day-Ahead Pool-Based Electric Energy Market,” International Journal of Forecasting, 21, 435–462. [495,501,502]

Engle, R. F., and Watson, M. (1981), “A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates,” Journal of the American Statistical Association, 76, 774–781. [495]

Engle, R. F., and Yoo, B. S. (1987), “Forecasting and Testing in Co-Integrated Systems,” Journal of Econometrics, 35, 143–159. [497,498]

Enns, P. G., Machak, J. A., Spivey, W. A., and Wrobleski, W. J. (1982), “Forecasting Applications of an Adaptive Multiple Exponential Smoothing Model,” Management Science, 28, 1035–1044. [498]

Escribano, A., and Peña, D. (1994), “Co-Integration and Common Factors,” Journal of Time Series Analysis, 15, 577–586. [498]

García-Martos, C., Rodríguez, J., and Sánchez, J. M. (2007), “Mixed Models for Short-Run Forecasting of Electricity Prices: Application for the Spanish Market,” IEEE Transactions on Power Systems, 22, 544–552. [495,502]

Härdle, W. K., and Trück, S. (2010), “The Dynamics of Hourly Electricity Prices,” Discussion Paper 2010-013, SFB 649 Economic Risk, Humboldt-University of Berlin, Germany. [495]

Hyndman, R. J. (2006), “Forecasting Functions for Time Series: R Package Version 2.13,” available at http://CRAN.R-project.org/package=forecasting. [496]

Jones, R. H. (1966), “Exponential Smoothing for Multivariate Time Series,” Journal of the Royal Statistical Society, Series B, 28, 241–251. [494,495,496,498]

Koopman, S. J., Ooms, M., and Carnero, M. A. (2007), “Periodic Seasonal Reg-ARFIMA-GARCH Models for Daily Electricity Spot Prices,” Journal of the American Statistical Association, 102, 16–27. [494]

Linhart, H., and Zucchini, W. (1986), Model Selection, New York: Wiley. [500]

Misiorek, A., Trück, S., and Weron, R. (2006), “Point and Interval Forecasting of Spot Electricity Prices: Linear vs. Non-Linear Time Series Models,” Studies in Nonlinear Dynamics and Econometrics, 10. [495,502]

Peña, D., and Box, G. E. P. (1987), “Identifying a Simplifying Structure in Time Series,” Journal of the American Statistical Association, 82, 836–843. [495]

Peña, D., and Poncela, P. (2004), “Forecasting With Nonstationary Dynamic Factor Models,” Journal of Econometrics, 119, 291–321. [495]

--- (2006), “Nonstationary Dynamic Factor Analysis,” Journal of Statistical Planning and Inference, 136, 1237–1257. [495]

Reinsel, G. C. (1993), Elements of Multivariate Time Series Analysis, New York: Wiley. [494,495,496,497,498]

Shumway, R. H. (1988), Applied Statistical Time Series Analysis, Englewood Cliffs, NJ: Prentice-Hall. [500]

Shumway, R. H., and Stoffer, D. S. (2006), Time Series Analysis and its Applications, New York: Springer. [496,498]

Stock, J. H., and Watson, M. V. (1988), “Testing for Common Trends,” Journal of the American Statistical Association, 83, 1097–1107. [495,497]

Weron, R. (2000), Modeling and Forecasting Electricity Load and Prices: A Statistical Approach (The Wiley Finance Series), New York: Wiley. [494,495,501]

--- (2009), “Forecasting Wholesale Electricity Prices: A Review of Time Series Models,” in Financial Markets: Principles of Modelling, Forecasting and Decision-Making (FindEcon Monograph Series), eds. W. Milo and P. Wdowinski, Lodz: WU L. [495]

Winters, P. R. (1960), “Forecasting Sales by Exponentially Weighted Moving Averages,” Management Science, 6, 324–342. [494]