Forecasting electricity spot prices using time-series models with a double temporal segmentation

Marie Bessec*, Julien Fouquau* and Sophie Meritet*

*LEDa-CGEMP Université Paris Dauphine, Paris, France; **ESCP Europe, Paris, France

ABSTRACT
The French wholesale market is set to expand in the next few years under European pressure and national decisions. In this article, we assess the forecasting ability of several classes of time-series models for electricity wholesale spot prices at a day-ahead horizon in France. Electricity spot prices display a strong seasonal pattern, particularly in France, given the high share of electric heating in housing during winter time. To deal with this pattern, we implement a double temporal segmentation of the data. For each trading period and season, we use a large number of specifications based on market fundamentals: linear regressions, Markov-switching (MS) models and threshold models with a smooth transition. An extensive evaluation on French data shows that modelling each season independently leads to better results. Among nonlinear models, MS models designed to capture the sudden and fast-reverting spikes in the price dynamics yield more accurate forecasts. Finally, pooling forecasts give more reliable results.

I. Introduction

In Europe, the reorganization of electricity markets is an ongoing project, with increasing market integration and competitive wholesale markets. The central–western European market (covering Benelux, France and Germany) is at the heart of power trading in continental Europe and it provides a blueprint for plans to create a single European electricity market.

The French wholesale market is the third largest in Europe after the German and British ones, and is about to change drastically. In France, it is more than challenging for new entrants to produce electricity, given the very low cost of production of nuclear power plants managed by the historic utility. To open up the French market to more competition, the new national electricity law, passed in December 2010, compels Electricité de France (EDF), the biggest producer of electric power in Europe and the former domestic vertically integrated public monopoly, to sell a proportion of its nuclear output to its rivals. The French government also gives alternative suppliers access to hydroelectric generation and modifies regulated rates for wholesale nuclear power. Moreover, following a European Commission decision, the regulated rates for electricity will discontinue in 2016 for small businesses on the retail market. Therefore, the number of electricity contracts offered by suppliers with market prices will increase. That is why the French wholesale market should become more liquid with more participants in the near future.

It is highly relevant to study the wholesale electricity prices in France at this stage. This article deals more specifically with forecasting French spot prices at a daily horizon. This is a particularly challenging issue due to the specific properties of electricity prices. First, since electricity is not storable, it has to be delivered as soon as it is generated. Excess demand at a given time cannot be compensated by excess supply a few hours before. This characteristic is responsible for price spikes and strong volatility; it can even lead to negative prices. Second, spot prices display a strong seasonal pattern at daily, weekly and yearly levels. This pattern is particularly important in France due to the high thermal sensitivity of electricity consumption.

In France, electricity consumption is highly dependent on temperature. This dependence has increased significantly over the past decade. It is now estimated to be 2300 MW/°C at 7 pm, the peak hour of daily
consumption in winter. The thermal sensitivity is much higher in France than in other European countries. For instance, the German electricity consumption increases only by 500 MW/C at 7 pm in winter (RTE 2012). This phenomenon is partly explained by the high rate of electric heating equipment in French housing. Electric heating has widely developed in France since the 1970s. Since 2005, more than 50% of new houses have been equipped with electric heating due to the low off-peak tariffs offered, and in 2009, this rate even reached 70%. In 2010, 35% of the total stock of housing (new and old) was heated with electricity, i.e. about 9.5 million homes (RTE 2012).

In the literature, there has been considerable work on modelling and forecasting spot prices. Basically, three groups of methods are used for this purpose: equilibrium or game theory approaches, simulation methods and times-series forecasting methods. The latter approach uses the past dynamics of prices and sometimes other exogenous variables to forecast electricity prices. For this purpose, linear and nonlinear models (Markov-switching (MS), threshold, time-varying parameter regressions) are used. Many authors rely only on the past dynamics of prices, while others also include fundamental variables in the model such as climatic variables, forecasted demand and capacity margin. In addition, a strand of this literature allows a time-varying conditional volatility using the class of GARCH models. See Liu and Shi (2013) for a recent survey.

In this article, we further explore in an extensive way the relevance of time-series models for day-ahead forecasting of power prices in France. In particular, we address several issues which are still open when modelling and forecasting electricity spot prices.

A first point concerns the treatment of seasonality in the data. This question is particularly important for the French market given the high sensitivity of electricity consumption to temperature, as explained above. In this article, we implement a double temporal segmentation of the data, to account for the intra- and inter-day seasonality. Prices display a distinct pattern depending on the hour of the day. Moreover, their level and volatility are much higher in fall and in winter. Therefore, the empirical analysis is conducted separately for each trading hour and for each season. The hourly segmentation is now common in the literature (e.g. Crespo Cuaresma et al. (2004), Misiorek, Trueck, and Weron (2006), Karakatsani and Bunn (2008), Bordignon et al. (2013) in the context of power price forecasting). However, to our knowledge, seasonal segmentation has not been applied for power price modelling and forecasting. Other more common approaches to deal with the long-run seasonality are described in Nowotarski, Tomczyk, and Weron (2013) and include the use of dummy variables, sinusoidal functions and wavelets.

Moreover, we investigate whether nonlinear models perform better than the linear ones for price forecasting, and if so, whether latent variable methods are better than threshold methods. We use regime switching models which are more likely to capture the sudden and fast reverting spikes in the price dynamics. We consider two widely used classes of nonlinear models: threshold regressive models with a smooth transition and MS models with fixed and time-varying transition probabilities (TVTP). A few related papers provide evidence in favour of regime-switching models for power price forecasting. Karakatsani and Bunn (2008) found that MS models and models with time-varying coefficients have a better predictive performance in the British market, while the threshold class of models outperforms the other specifications for the forecasting of Californian spot prices in Misiorek, Trueck, and Weron (2006). Kosater and Mosler (2006) also obtained better long-run forecasts of German prices using MS models.

Finally, we explore the benefit of combining forecasts over individual predictors. We compare the performance of individual models, as well as combinations of forecasts with various weighting schemes. Despite a long tradition in the use of forecast combination (Bates and Granger (1969) and Newbold and Granger (1974)) and renewed interest in the forecasting literature in the last decade, the use of combination of forecasts is recent in the context of electricity prices and gives positive results in this area too. The existing papers provide evidence in favour of combinations for several power markets: the UK power exchange (Bordignon et al. 2013), Nord Pool (Raviv, Bouwman, and Van Dijk 2015), the European Energy Exchange (EEX) and the Pennsylvania-New Jersey-Maryland (PJM) inter-connection (Nowotarski et al. 2014). In this article,
we provide additional evidence for various weighting schemes and for a new geographical area.

We conduct an out-of-sample evaluation of the forecasting performance of the various models on a day-ahead basis. We consider forecasts of French wholesale prices over 2012 and assess the performance of the models or their combinations for each trading hour and each season. We calculate several forecast evaluation criteria and test for a significant difference between them with several tests. The evaluation is extensive, since we consider 2880 individual models and 2304 model combinations. The results show that the double temporal segmentation improves the results, except in summer. The gain is stronger if we focus on linear models and on specifications without fundamental variables. Among the nonlinear specifications, the three-state MS model leads to more accurate results. However, the performance of individual models remains unstable across hours and seasons. Hence, pooling the results provides more reliable forecasts.

This article proceeds as follows. In Section II, we present the French electricity markets with a focus on the wholesale one. Section III describes the data. Section IV presents the estimation and forecast design. Section V provides an assessment of the forecasting performance of the various specifications.

II. The market design

Electricity is traded either on the wholesale market or on the retail market. Retail transactions involve the final sale of power to end-use customers. On the wholesale market, electricity is traded (bought and sold) before being delivered to the network to end users (individuals or companies). The empirical analysis will deal with the wholesale exchanges made on the European Power Exchange (EPEX) spot.

The participants are the electricity suppliers most of whom (but not all) own generating plants, brokers and traders. In 2012, 565 Terawatt hours (TWh) were injected into grids: 449 TWh from generating units, 31 TWh from virtual power plant (VPP) generation, 56 TWh from Regulated Access to Incumbent Nuclear Electricity (ARENH) and 28 TWh from imports.

• EDF offers access to 5400 MW of its domestic capacity via quarterly auctions (VPPs). This capacity can be obtained by producers, suppliers and traders already active in the domestic market or wishing to enter through the quarterly auction mechanism. EDF sells this production capacity in the form of power purchase contracts specifying both capacity and prices.
• The law ‘Nouvelle Organisation du Marché de l’Electricité’ (or NOMEx law) established the ARENH system. This system entitles suppliers to purchase electricity from EDF at a regulated price, in volumes determined by the French energy regulator. One issue with the deregulation of the electricity market in France lies in the fact that the incumbent supplier EDF could offer tariffs related to its fleet of inexpensive generation (dependent on 90% of the cost of nuclear and hydro). It is not the case for the alternative suppliers who make offers based on the European wholesale prices. To reduce this difference, the solution chosen is to allow alternative suppliers to repurchase a share of the EDF nuclear generation at the ARENH rate.
• France also imports electricity when the cost of electricity is more attractive at a given time. The electricity interconnections enable importing less expensive electricity at certain times of the day than that produced by the national production units. This is the case in France at peak times (especially in the evening in winter), when thermal power plants are used. It is usually more profitable to import electricity at these times and to export when domestic demand decreases.

These 565 TWh injected on the grids in 2012 were withdrawn from grids through: 463 TWh for end user consumption, 31 TWh grid losses purchased on markets and 62 TWh for exports.

An important part of the electricity wholesale market activity takes place on power exchange markets: France EPEX spot for spot products, based in Paris and EEX Power Derivatives for future products in Leipzig.\(^1\) EPEX spot was created in 2008 by merging the power spot activities of the French energy exchanges Powernext and the German one EEX. The

---

\(^{1}\)The rest of exchanges takes place over-the-counter (OTC), through direct transactions or through intermediaries (brokers and trading platforms). The total volume of OTC transactions is not public.
EPEX spot trades contracts for the physical delivery of electricity in Austria, Germany, France and Switzerland.

The reference price for the spot trade is the price of the day-ahead product on EPEX spot. EPEX spot prices are negotiated the day before the delivery by an auction mechanism. In the morning, buyers and sellers (consumers and power generators) submit their orders (price/quantity combination) for each hour (or block of hours) of the forthcoming day. The market is closed at 12:00 noon for France and EPEX spot computes the demand and supply curves and the equilibrium price and volume for each hour of the forthcoming day. The results are published as soon as they are available from 12:40 pm for France.

III. Data

We focus on electricity spot prices in euros per megawatt hour (€/MWh) for France from the EPEX market.

We consider several price drivers as explanatory variables. The variables are published at a daily or hourly frequency and must be available before the release of spot prices for day $t$. Widely accepted market fundamentals include the following indicators of supply, demand and risk:

- Demand forecast (in MW): the 24 hourly day-ahead forecasts for continental France are released by the French transmission grid operator, Réseau de Transport d’Electricité (RTE), at 0:00 in $t-1$.

- Capacity margin (in MW): the forecasted margin is published by RTE at 8 pm in $t-1$. It represents the volume of capacity available for RTE over and above the operating schedule capable of being used to cope with generation or consumption contingencies. We take the margin of the morning peak available all over the year.

- Volatility of spot prices (in €/MWh): we use the coefficient of variation (standard deviation/mean) of the hourly prices over the last five days (weekly rolling window).

- Past values of spot prices (in €/MWh): the price of the day before (lag-1) and of the same day the week before (lag-5) and the weighted average price over the previous day are particularly relevant.

- Gas price (in €/MWh): we consider the Dutch day-ahead price of natural gas (Title Transfer Facility (TTF)) which is a reference price in Europe.\(^2\) To avoid an endogeneity problem, we use lag-1 price.

- Forecasted balance of exchange programs with Germany (in MW): RTE provides figures for the total volume of exports from France and imports into France for each hour as it is known each day at the end of the afternoon, for the following day.

The power prices, demand and volumes of exchange have an hourly frequency, while the capacity margin and the gas price are available at a daily frequency. With the exception of cross-border trade, these variables have already been considered as price drivers in the literature (see for instance Karakatsani and Bunn (2008) and Bordignon et al. (2013)).\(^3\) In this article, we also take into account the cross-border exchanges in electricity. As described in the previous section, the French electricity transmission network is interconnected with neighbouring countries. The growth of cross-border trade leads us to account for these exchanges in our models.\(^4\) We focus on the exchanges with the historical partner Germany. We omit temperature and other climatic variables, already taken into account in the demand forecast by RTE (see supplemental data).

In contrast to many papers, we take into account the particular timing of the release of the spot prices, as depicted in Fig. 1. The market is closed at 12:00 noon and the spot prices for the 24 hours of the next day are published from 12:40 pm. Following the market practice, we only use information available up to noon before the market clearing. This means that all variables involved in the forecasting must be available before noon one day before the reference day. This is the case of the

\(^2\)The TTF is a virtual trading point for natural gas in the Netherlands, set up in 2003. It became the biggest gas trading platform in Europe.

\(^3\)Many authors only include the (forecasted) demand in addition to lagged prices in the forecasting equation. See for instance Nogales et al. (2002), Conejo et al. (2005), Misiorek, Trueck, and Weron (2006), Weron and Misiorek (2008).

\(^4\)In November 2010, the Central Western Europe Market Coupling (CWE) was launched. CWE is a cooperation of Transmission System Operators and power exchanges covering the Belgian, Dutch, French and German electricity markets.
forecasted demand released at 0:00 the day before the reference day and the lagged gas price. However, the forecasted capacity margin and exchange volumes with Germany for the next day are published by RTE only at the end of the afternoon. For this reason, we use the lagged values of these two variables in our models. By doing so, we ensure that our results are relevant for practitioners.\footnote{Given the delay in the publication of the explanatory variables and the early publication of spot prices, the forecasts performed here are useful over a short window, from 0:00 to noon. Market participants use this window to optimally elaborate their bidding strategies.}

The whole sample spans from 2 January 2008 to 31 December 2012, yielding 1263 daily observations for each trading hour. We consider several ranges of data to forecast some prices in each season of 2012 (see below). Note that it is particularly challenging to forecast prices in 2012. A historical peak was reached in February 2012 during an exceptionally cold wave, while the lowest consumption level observed for five years occurred in August 2012. The prices, demand and margin are taken in logarithm given their high volatility. Like Karakatsani and Bunn (2008) and Bordignon et al. (2013), we remove weekends and holidays from the data and consider working days only. We also remove two days in 2012 (26 December 2012 and 2 January 2012), where prices are negative at 4 and 5 am.\footnote{Negative price can occur in situations of overcapacity. We discarded these observations to be able to use the logarithmic transformation.} We apply several unit root tests to each variable in logarithm.\footnote{To check the stationarity of the variables, we have implemented the Augmented Dickey-Fuller (ADF), Phillips–Perron (PP) and Kwiatkowski–Phillips–Schmidt–Shin tests. The tests were applied to the series in levels and in first differences when necessary. The results are available upon request.} All series are found to be stationary, except the forecasted demand and the gas price. In the following, we will consider the latter ones in differences.

### IV. Modelling and forecasting procedures

#### Temporal segmentation

Due to the strong seasonality present both in the intra-day and inter-day dynamics of electricity prices, especially in France, given the high share of electricity in residential heating, we implement a double temporal segmentation of the data.

As already mentioned in the literature, prices display a distinct pattern depending on the hour of the day. This variation is mainly a result of the evolution of demand during the day. According to RTE, the intra-day profile of French consumption is characterized by four phases: the night trough, which is the minimum consumption of the day, the morning peak, the afternoon trough and the evening peak. The maximum consumption is reached at the morning peak at 13:00 in summer and at the evening peak around 19:00 in winter (heavy use of electrical appliances, public transport, lighting, heating . . .).

As done in the work of García-Martos, Rodríguez, and Sánchez (2011), Fig. 2(a) shows the box plot of prices between 2008 and 2012 at each trading hour.\footnote{293 extreme values out of the 24 792 observed prices have been removed for the readability of the graph.} In line with the demand profile, peaks in the level occur around 1 and 7 pm, while prices are lower from 2 to 5 am and 3 to 6 pm. Prices display a larger volatility between 3 and 5 am and between 6 and 8 pm. To allow a variation of the coefficients of the forecasting model during the day, we model

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.pdf}
\caption{Daily updating scheme of data releases in EPEX spot.}
\end{figure}
separately each hour of the day, as often done in the literature (see, for examples in forecasting, Crespo Cuaresma et al. (2004), Misiorek, Trueck, and Weron (2006), Weron and Misiorek (2008), Karakatsani and Bunn (2008), Bordignon et al. (2013)).

Spot prices also display a strong seasonal pattern during the year. Their level and volatility are much higher in fall and winter, as shown by the box plot of power prices in each season (Fig. 2(b)). Indeed, consumption is seasonal and the price per MWh varies seasonally: summer demand is lower and the MWh is cheaper. It is produced from nuclear and hydropower. In winter, the demand is higher and the MWh is more expensive. Therefore, the last power plant ‘called’ has to be able to quickly meet demand. These plants can meet this constraint by using fossil fuels at a higher cost. For example in France, about 30 GW are needed in August during the day while up to 102 GW have been called in February at 7.00 pm in a cold winter. To also allow a variation of the coefficients during the year, we model each season independently.

As mentioned above, the hourly segmentation is now common in the literature. However, to the best of our knowledge, seasonal segmentation has not been applied for power price modelling. Other more common approaches to deal with the long-run seasonality are described in Nowotarski, Tomczyk, and Weron (2013): dummies, sinusoidal functions and wavelets. The usual approach employing monthly dummy variables (allowing generally a variation of the intercept only) or sinusoidal functions is less flexible and/or less parsimonious. The use of wavelet decomposition for forecasting requires arbitrary choices about the wavelet family, the vanishing moment in many of them, the number of octaves. Moreover, the use of fundamental variables is less straightforward with the wavelet decomposition.

The variation of the coefficients of a linear regression of the price on the drivers depicted in Fig. 3 across the two dimensions (trading hours and seasons) supports this double segmentation. Figure 3 shows the estimated coefficients of a linear model relating the spot prices to the aforementioned fundamental variables. This regression is conducted either on 2012 data or on seasonal data (i.e. on datasets consisting only of observations of the same season). This will be explained further in the next section. We note a significant variation of the coefficients across hours (variation of the coefficients along

9 Alternative approaches include the use of dummies for the trading periods in a unique equation (Popova (2004)). However, this method is far less flexible or parsimonious. Alternatively, multivariate approaches are employed by Huisman, Huurman, and Mahieu (2007) with a panel framework or by Panagiotelis and Smith (2008) taking into account the hourly heterogeneity in a vector autoregression model.

10 To remove the weekly seasonality, Crone and Kourentzes (2011) split the hourly data of UK electricity load in seven subclasses of subseries, one for each day of the week. In this article, the weekly seasonality is less of an issue since we focus on working days.
each curve) and seasons (variation between the curves).

**Linear versus nonlinear models**

We consider linear and nonlinear specifications of the relationship between spot prices and market fundamentals. A summary of all specifications with their code is given in Table 1. In the following, \( p_{t}^{(h)} \) denotes the log-price and \( x_{t}^{(h)} \) the fundamentals for hour \( h, h = 1, \ldots, 24 \) of day \( t \).

The linear specifications include a basic autoregressive model, an exponential model (EXPO) and autoregressive regressions with exogenous variables. The first two models do not include fundamental regressors and will be useful to assess the contribution of exogenous day-ahead information to forecasting. The autoregressive (AR) model is simply given by:

\[
p_{t}^{(h)} = \alpha^{(h)} + \sum_{i=1}^{p} \phi_{i}^{(h)} p_{t-i}^{(h)} + \epsilon_{t}^{(h)},
\]

\( t = 1, \ldots, T \)  \hspace{1cm} (1)

where \( \alpha^{(h)}, \phi_{i}^{(h)}, i = 1, \ldots, p \) and \( h = 1, \ldots, 24 \) are constant coefficients estimated by least squares and \( \epsilon_{t}^{(h)} \) is the error term. The order \( p \) is selected with the Akaike information criterion (AIC) for a maximum lag length fixed to 5. In this framework, the optimal predictor \( \hat{p}(1) = E(p_{T+1}|p_{T}, p_{T-1}, \ldots) \) of the price

---

**Figure 3. Coefficients of the drivers in linear models.**

**Table 1. Individual models.**

| Model                                                                 | Code  |
|-----------------------------------------------------------------------|-------|
| Autoregressive model                                                  | AR    |
| AR model augmented with forecasted demand                             | AR-X  |
| Linear regression with regressors selected with a stepwise procedure  | LSTEP |
| MS model, 2 regimes, FTP, variation of intercept and variance          | MS_1  |
| MS model, 2 regimes, FTP, variation of intercept, AR coefficients and variance | MS_2  |
| MS model, 2 regimes, FTP, variation of all coefficients                | MS_3  |
| MS model, 2 regimes, TVTP, variation of intercept and variance         | MSV_1 |
| MS model, 2 regimes, TVTP, variation of intercept, AR coefficients and variance | MSV_2 |
| MS model, 2 regimes, TVTP, variation of all coefficients               | MSV_3 |
| MS model, 3 regimes, FTP, variation of intercept and variance          | MS3_1 |
| MS model, 3 regimes, FTP, variation of intercept, AR coefficients and variance | MS3_2 |
| STR model, variation of intercept                                     | STR_1 |
| STR model, variation of intercept and AR coefficients                 | STR_2 |
| STR model, variation of all coefficients                               | STR_3 |

---

11 We have assessed the existence of a break in the forecasting equations due to the implementation of market coupling in November 2010 (see footnote 4). Rolling regressions do not show any major change in the relationship between spot prices and the price drivers. Therefore, we do not incorporate a structural change in the forecasting equations in November 2010.
in \( T + 1 \) based on the information available up to time \( T \) is given by:

\[
\hat{p}_T^{(h)}(1) = \alpha^{(h)} + \sum_{i=1}^{p} q_i^{(h)} \hat{p}_{T+1-i}^{(h)}.
\]  

(2)

Forecasts from Equation (1) perform well as documented in the literature.

To forecast stationary data, another usual solution relies on exponential smoothing technique. This approach consists of assigning exponentially decreasing weights over time to the past observations:

\[
s_t^{(h)} = \alpha^{(h)} p_t^{(h)} + \left(1 - \alpha^{(h)}\right) s_{t-1}^{(h)} \quad \text{for} \quad t > 0
\]  

(3)

with \( s_0^{(h)} = p_0^{(h)} \),

where \( s_t^{(h)} \) is the smoothed value and \( \alpha^{(h)} \in [0,1] \) is the smoothing factor. The parameter \( \alpha^{(h)} \) is estimated by minimizing the mean square error (with \( e_t^{(h)} = p_t^{(h)} - s_{t-1}^{(h)} \)). Then, the optimal predictor \( \hat{p}_T^{(1)} \) can be written as follows:

\[
\hat{p}_T^{(h)}(1) = s_T^{(h)}.
\]  

(4)

Third, the autoregressive model with exogenous variables (AR-X) is specified as:

\[
p_t^{(h)} = \alpha^{(h)} + \beta^{(h)} x_t^{(h)} + \varepsilon_t^{(h)},
\]  

(5)

where \( \beta^{(h)} \) is a \( k \times 1 \) vector of constant parameters, \( x_t^{(h)} = (p_{t-1}^{(h)}, \ldots, p_{t-p}^{(h)}, z_t^{(h)}) \) is the \( k \times 1 \) vector of regressors consisting of autoregressive terms and fundamentals \( z_t^{(h)} \) in day \( t \) and \( \varepsilon_t^{(h)} \) is the error term. The model is estimated by least squares. In this case, the optimal predictor \( \hat{p}_T^{(1)} \) is given by:

\[
\hat{p}_T^{(h)}(1) = \alpha^{(h)} + \beta^{(h)} x_{T+1}^{(h)}.
\]  

(6)

We also consider nonlinear models in order to capture infrequent and extreme observations typical of electricity prices (spikes and drops) and the possible change in the relationship between spot prices and their fundamentals during these abnormal periods.

First, we consider MS models popularized by Hamilton (1989). In these models, the parameters of Equation (5) depend on an unobservable discrete variable \( S_t \):

\[
p_t^{(h)} = \alpha^{(h)} (S_t^{(h)}) + \beta^{(h)} (S_t^{(h)}) x_t^{(h)} + \varepsilon_t^{(h)} (S_t^{(h)})
\]  

(7)

where \( \varepsilon_t^{(h)} (S_t^{(h)}) \rightarrow \text{NID}(0, \sigma^2(S_t^{(h)})) \). The variable \( S_t^{(h)} = 1, 2, \ldots, M \) represents the state that the process is in at time \( t \). This variable is assumed to follow a first-order Markov chain defined by the following transition probabilities:

\[
q_{ij}^{(h)} = P(S_t^{(h)} = j | S_{t-1}^{(h)} = i) \quad \forall t,
\]  

(8)

where \( \sum_{j=1}^{M} q_{ij}^{(h)} = 1, \forall i = 1, 2, \ldots, M \). The forecast of the chain \( S_t^{(h)} \) at horizon \( 1 \) is given by

\[
\hat{\xi}_T^{(h)} = P_{T+1}^{(h)} \xi_T^{(h)}
\]  

where \( P_T^{(h)} \) is the transition matrix of the Markov chain and \( \xi_T^{(h)} \) the vector of smoothed probabilities in \( T \). The optimal predictor \( \hat{p}_T^{(1)} \) is derived by weighting each estimated regime with the predicted probabilities of the states in \( T + 1 \), \( \hat{\xi}_T^{(h)} \).

Following Diebold et al. (1994), we extend this specification with TVTP. In this case, the transition probabilities are not time invariant. Instead, they depend on an exogenous variable \( q_t^{(h)} \):

\[
P(S_{t}^{(h)} = i | S_{t-1}^{(h)} = i, q_t^{(h)}) = \Phi(\alpha_i^{(h)} + b_i^{(h)} q_t^{(h)})
\]  

(9)

where \( \Phi \) is the cumulative distribution function of the standard normal distribution, \( \alpha_i^{(h)} \) and \( b_i^{(h)} \) are unknown parameters for regime \( S_t = i \) and \( q_t^{(h)} \) is an exogenous variable (possibly included in \( x_t^{(h)} \)). The forecast of the chain \( S_t^{(h)} \) at horizon \( 1 \) is given by

\[
\hat{\xi}_T^{(h)} = P_{T+1}^{(h)} \hat{\xi}_T^{(h)}
\]  

where \( P_{T+1}^{(h)} \) is the transition matrix of the Markov chain in \( T + 1 \).

The MS models are estimated via a numerical maximization of the maximum likelihood. In the following, we only consider the case of two and three regimes (\( M = 2, 3 \)). For \( M = 2 \), we consider three versions of the MS model with fixed transition probabilities (FTP) or TVTP. The first ones allow a variation of the intercept and the variance only, while the second ones also allow a variation of the autoregressive coefficients. In the third ones, all the parameters depend on \( S_t \). For \( M = 3 \), we only consider the first two constrained specifications with
FTP and do not consider the general form, given the proliferation of parameters when the number of regimes increases.

A second class of nonlinear models considered in this article is the smooth threshold autoregressive models (Granger and Terasvirta 1993) augmented with exogenous variables. The variable $p_t^{(h)}$ follows a smooth transition regression (STR) model if:

$$p_t^{(h)} = \alpha_0^{(h)} + \beta_0^{(h)} x_t^{(h)} + \left(\alpha_1^{(h)} + \beta_1^{(h)} x_t^{(h)}\right) G\left(q_t^{(h)}; y^{(h)}, c^{(h)}\right) + \varepsilon_t^{(h)}$$

for $t = 1, \ldots, T$. The residual $\varepsilon_t^{(h)}$ is assumed to be i.i.d. $(0, \sigma_t^2)$. In our case, the mechanism of transition is defined as a logistic function of order 1:

$$G\left(q_t^{(h)}; y^{(h)}, c^{(h)}\right) = \left[1 + \exp\left(-y^{(h)}\left(q_t^{(h)} - c^{(h)}\right)\right)\right]^{-1}, y^{(h)} > 0$$

where $c^{(h)}$ denotes a location parameter and $y^{(h)}$ determines the slope of the transition function i.e. the speed of the transition from one regime to the other one. The logistic function is a continuous function with S shape bounded between 0 and 1. It represents a smooth transition from an inferior regime (coefficient $\beta_0^{(h)}$) to a superior one (coefficient $\beta_0^{(h)} + \beta_1^{(h)}$) with an infinite number of intermediate regimes in between. The estimation of the parameters in a STR model consists of applying nonlinear least squares. The optimal predictor $\hat{p}_T(1)$ is given by:

$$\hat{p}_T(1) = \alpha_0^{(h)} + \beta_0^{(h)} x_{T+1}^{(h)} + \left(\alpha_1^{(h)} + \beta_1^{(h)} x_{T+1}^{(h)}\right) G\left(q_{T+1}^{(h)}; y^{(h)}, c^{(h)}\right).$$

Again, we consider three versions of this model: the first one with a variation of the intercept only, the second one with a variation of the intercept and the AR coefficients and the last one with a variation of the coefficients of all regressors.

With respect to the regressors’ selection, we consider a linear model with explanatory variables selected using a stepwise procedure among the set of market fundamentals described in the previous section. To get a fair comparison and make the convergence of the estimation algorithms easier for the nonlinear models, we also consider all the specifications with the same reduced structure: two autoregressive terms – the weighted average price of the preceding day$^{12}$ and the price at the same hour two days before – and demand in first-difference. At this level, we choose the most significant regressors over all data ranges and trading hours. The forecasted demand is found to be more influential in our models since it refers to the day of interest, while the other regressors are lagged because of their delay of publication.

As suggested by Chen and Bunn (2010), the transitional variable can vary according to the season and the trading hour. Therefore, we select the transitional variable of the threshold and MS models with time-varying probabilities for each hour, each season and each treatment of seasonality. The variable is chosen among the whole set of explanatory variables (lagged price, lagged margin, first-difference forecast demand, first-difference of lagged gas price, forecasted power exchange with Germany, past volatility of spot prices). For each regime-switching model with a time-varying transition, we conduct a linearity test with each potential variable and select the one which provides the strongest rejection of the null of linearity.

**Individual versus pooled forecasts**

In the previous subsection, we have proposed several linear or nonlinear candidate methods for forecasting electricity prices. However, an alternative solution used by Bordignon et al. (2013) and Nowotarski et al. (2014) for price forecasting consists of combining forecasts. The objective highlighted by Bordignon et al. (2013) is to obtain accurate predictions when a wide variety of models exists and no predominant one can be selected.

We consider several methodologies for combining forecasts. The easiest way to obtain the combined prediction $\hat{p}_t$ is to consider the mean of the individual predictions $\hat{p}_i^{(k)}$ of the $k$ individual models considered (for $k = 1, \ldots, K$):

$^{12}$The lagged weighted average price is more significant than the price at the same hour the day before. The weighted average price is also smoother and it makes the estimation of nonlinear models easier.
\[ \hat{p}_t^c = \frac{1}{K} \sum_{k=1}^{K} \hat{p}_t^{(k)} . \]  

An alternative approach is to weigh the predictions based on a selected criterion:

\[ \hat{p}_t = \sum_{k=1}^{K} \omega_{t,k} \hat{p}_t^{(k)} \]  

with \( \omega_{t,k} \), the weight given at time \( t \) to the predictor \( k \), \( k = 1, \ldots, K \). Like Bordignon et al. (2013), we adopt the usual solution proposed by Bates and Granger (1969) which consists of assigning weights depending on the inverse of means square prediction errors of the \( l \) most recent observations:

\[ \omega_{t,k} = \left( \frac{\sum_{i=t-l}^{t-1} e_{r,i}^2}{\sum_{j=1}^{K} \left( \sum_{i=t-l}^{t-1} e_{r,i}^2 \right)} \right)^{-1} , \]  

where \( e_{r,i} = p_t - \hat{p}_t^{(k)} \), \( \omega_{1,k} = 1/K \). Accordingly, \( 0 \leq \omega_{t,k} \leq 1 \), \( \sum_{k=1}^{K} \omega_{t,k} = 1 \). The adaptive weights are larger if the individual models yield more accurate forecasts. The third solution is the aggregated forecast through exponential re-weighting (AFTER) (Yang 2004; Zou and Yang 2004). The weights are obtained recursively according to:

\[ \omega_{t,k} = \frac{\hat{v}_{t-1,k}^{-1/2} \exp \left[ -\frac{e_{r,t-1,k}^2}{2\hat{v}_{t-1,k}} \right]}{\sum_{i=1}^{K} \hat{v}_{t-1,i}^{-1/2} \exp \left[ -\frac{e_{r,t-1,i}^2}{2\hat{v}_{t-1,i}} \right] \omega_{t-1,i}} , \]  

where \( e_{r,t-1,k} = p_{t-1} - \hat{p}_t^{(k)} \), \( \omega_{1,k} = 1/K \). Accordingly, \( 0 \leq \omega_{t,k} \leq 1 \), \( \sum_{k=1}^{K} \omega_{t,k} = 1 \) and \( \hat{v}_{t-1,k}^{-1} = \frac{1}{\sum_{r=1}^{t-1} e_{r,t}^2} \) is the estimation of the prediction variance.

A last solution adopted in this article relies on the predictive likelihood model average (PLMA) proposed by Kapetanios, Labhard, and Price (2006). The weights are assigned depending on AIC information criteria built on a sample of the \( l \) previous forecast errors:

\[ \omega_{t,k} = \frac{\exp \left( -1/2 \Psi_{t,k} \right)}{\sum_{j=1}^{K} \exp \left( -1/2 \Psi_{t,j} \right)} \]  

for \( k = 1, \ldots, K \), where \( \Psi_{t,k} = \text{AIC}_{t,k} - \min_j \text{AIC}_{t,j} \), \( 0 \leq \omega_{t,k} \leq 1 \) and \( \sum_{k=1}^{K} \omega_{t,k} = 1 \). The AIC criterion is not calculated on the basis of the estimated log likelihood (which is derived in-sample from the observed data). Instead, this criterion is based on a predictive likelihood derived from out-of-sample forecast errors. This construction has the advantage to combine a measure of forecast accuracy and a penalty term for model complexity. A summary of all possible combinations with their code is given in Table 2.

### Table 2. Combination of models.

| Model combinations | Code   |
|--------------------|--------|
| Combination by average, all models | C1     |
| Combination by average, linear models | C1_L   |
| Combination by average, nonlinear models | C1_NL  |
| Combination with Bates and Granger weights, all models | C2     |
| Combination with Bates and Granger weights, linear models | C2_L   |
| Combination with Bates and Granger weights, nonlinear models | C2_NL  |
| Combination with AFTER weights, all models | C3     |
| Combination with AFTER weights, linear models | C3_L   |
| Combination with AFTER weights, nonlinear models | C3_NL  |
| Combination with PLMA weights, all models | C4     |
| Combination with PLMA weights, linear models | C4_L   |
| Combination with PLMA weights, nonlinear models | C4_NL  |

### V. Day-ahead forecasting

#### Empirical design

We compare the various approaches described above for modelling French spot prices through their forecasting performance.

The evaluation is conducted under real conditions, i.e., out of the estimation period. We remove the last observations of our sample. We estimate the 15 specifications over a rolling window and run a day-ahead forecast of the variable \( p_t^{(h)} \) for each trading hour \( h = 1, \ldots, 24 \). We repeat these calculations up to the last observation of the out-of-sample period. The out-of-sample period consists of the last 35 observations of each season in 2012. The estimation set includes the preceding weekdays either of the entire past year, or of the same season (hereafter referred to as nonseasonal and seasonal ranges, respectively). To have a comparable number of observations (around 255 weekdays) across simulations, the seasonal range consists of four seasons. The alternative datasets used to forecast the last observations of each season are described in Fig. 4.

The choice of the specifications is done in real-time. In the case of the AR process, the lag length is chosen with the AIC criterion at each iteration of the out-of-sample exercise. Similarly, the choice of the regressors in the linear model with the stepwise
procedure is done at each iteration on the basis of the available information at the time of the forecast. The threshold variable in the STR models and the variable entering the transition probabilities in the MS models are selected on the sample excluding the forecasting period at the beginning of the iterations.\(^\text{13}\) As suggested by Chen and Bunn (2010), the optimal transitional variables vary a lot depending on the time period, the season as well as the specification. The forecasted balance of power exchange, the demand forecast and the past volatility of spot prices are chosen more often.\(^\text{14}\) At each period of the evaluation period, we use a large set of initial values to avoid any dependency of the results on the initial values for the estimation of the MS and STR models.

To mimic the practice of forecasters, we use an automatic insanity filter (Stock and Watson 1999). This filter removes unrealistic forecasts and replaces them with some reasonable values. Given the existence of spikes in the price series, this filter must be used with parsimony. We only discard values found outside a wide interval given by the historical range of prices (over all hours) enlarged by plus/minus the average standard deviation of hourly prices. This rule requires discarding 0.45% of forecasts only. These insane forecasts are replaced by the last observed price at the same hour (naive forecast).

The out-of-sample evaluation is conducted for each model or combination of models in each season and trading hour. We consider a total of 2880 individual models and obtain a set of 35 forecasts with each of them. Pooled forecasts are also obtained by combining these individual forecasts with the four combination methods described in the previous section. In the adaptive rules, the weights are computed from the last \(l = 10\) more recent observations. In the case of the last weighting scheme, the AIC criteria are computed with the number of parameters estimated recursively for each hour and each season.\(^\text{15}\) At this level, we also distinguish combinations from linear and nonlinear specifications. A total of 2304 combinations is obtained.

Several comparisons are done with these results: first, we contrast the forecasting accuracy of regime

---

\(^{13}\)The linearity tests are not performed at each recursion because it is too time consuming. However, we exclude the 35 forecasted observations to have an out-of-sample experience.

\(^{14}\)For sake of parsimony, the results are not reported in the paper but they are available upon request.

\(^{15}\)The number of parameters in the AR and LSTEP models varies at each iteration and it is fixed in the other specifications.
switching models (MS and STR) to linear specifications (AR, exponential and AR-X). Second, the benefit of combining forecasts to achieve more accurate predictions is studied by comparing the individual forecasts of the 15 models versus pooled forecasts. Finally, we assess the gain due to the double temporal segmentation of the data. To this end, we compare the forecasts of the models estimated over an entire year or on data belonging to the same season, as described above.

The forecast accuracy of each model or combination of models is measured with the usual root mean square error (RMSE), mean absolute error (MAE) and mean absolute per cent error (MAPE). Using these criteria, we provide a ranking of the models according to their RMSE/MAE and report an average rank over the 24 trading hours for each season and the full year. This evaluation is qualitative since it relies on averages of rankings. To take into account quantitatively the forecast accuracy of each model, we also compute MAPE for each model over all hours for each season and the whole year and rank the models according to the results (see Nowotarski, Tomczyk, and Weron (2013) for a similar approach).

To compare the predictive ability of the competing approaches, we implement several pairwise tests. First, to assess if there is a significant difference between the MSE and MAE criteria, we perform the test of Diebold and Mariano (1995) with the correction of Harvey, Leybourne, and Newbold (1997) in small samples. We also conduct the encompassing test of Harvey, Leybourne, and Newbold (1997). If the null hypothesis of this test is rejected, the competing predictor contains useful information not present in the model.\(^{16}\)

Finally, since we consider many models, we implement a multiple comparison-based test with the model confidence set (MCS) approach developed by Hansen, Lunde, and Nason (2011). This procedure aims at determining the best model(s) from a collection of models with a given level of confidence. The MCS approach puts together all the models and eliminates step-by-step the model with the lowest predictive ability from the pool of models up until the final set where all models are equal in terms of predictive ability. In our paper, this selection relies on the RMSE and MAE criteria. We use the test statistics \(T_R\) and we apply a block bootstrap procedure to estimate the distribution of the statistic under the null hypothesis.\(^{17,18}\)

### Out-of-sample results

The rankings of the models are provided in Tables 3 and 4. The results of the tests appear in Tables 5 and 6.

To give an overall picture of the forecasting performance of our fundamental models relative to a naive benchmark, we compute the RMSE of the forecasting models over the RMSE of a random walk with a drift. The latter is a usual benchmark when forecasting financial variables. A ratio below one indicates a gain relative to this benchmark model. Figure 5 depicts the estimated density of the ratios for each model. Each distribution is estimated from 192 values (24 hours \(\times\) 4 seasons \(\times\) 2 estimation periods) using a nonparametric kernel approach. The estimate is based on a normal kernel function. Ratios inferior to one are more frequent. The median ratio is around 0.8 for the specifications including fundamental variables (except STR_2 and STR_3 for which the median is above 0.9) and around 0.9 for the two models relying on the past dynamics of prices only. Hence, the naive benchmark is outperformed in most cases, especially by the models including market variables.

To compare the performance of individual models, Table 3 gives the number of times over the 24 hours of each season that a given model provides the best forecast in terms of MSE, MAE and MAPE. Table 4 provides the average rank per season for each model and the global rank over the entire year. We provide a qualitative ranking based on the RMSE and MAE criteria and a quantitative ranking relying on the aggregate MAPE of the models. Table 5 reports the results of the tests by a class of models. We provide the percentage of times that

---

\(^{16}\) Using Monte Carlo simulations, Busetti and Marcucci (2013) provided a comparison of the finite sample properties of several tests of equal MSE and forecast encompassing.

\(^{17}\) Like Hansen, Lunde, and Nason (2011), we use a block bootstrap procedure with block length of 12 observations and 1000 resamples for the computation of the \(p\)-values.

\(^{18}\) We thank an anonymous referee for this suggestion.
each approach gives lower MSE or MAE and the percentage of times that these criteria are significantly lower according to the Diebold–Mariano test. The last block of results gives the percentage of rejection of the null hypothesis for the encompassing test of Harvey, Leybourne, and Newbold.
54.9 70.8 89.6 34.1 36.1 80.9 25.2 26.9 27.1 36.4 43.9 38.9 35.2 37.5 41.7 54.9 77.8 77.8 55.6 77.8 69.4 79.9 54.5 95.8 50.0 30.4 33.0 51.0 60.7 56.3 50.4 69.1 24.5 52.1 14.6 13.9 52.1 47.2 27.1 27.1 20.8 52.8 29.3 39.7 43.3 56.3 66.8 59.4 54.4 52.5 16.7 45.8 46.9

Table 5. Comparison of nonlinear, pooled, seasonal models (in %).

Table 6. Model confidence set (MCS) results (in %).

Notes: The first block reports the percentage of times over the 24 hours that (a) yields a lower loss than (b) with MSE or MAE loss functions. For example, nonlinear models beat linear specifications in 72.6% of cases in spring (seasonal range). The second block provides the results of the one-sided Diebold and Mariano test for the null hypothesis that (a) has the same forecast accuracy than (b) with MSE or MAE loss functions. Each entry represents the percentage of times that (a) is significantly better than (b). The last block provides the results of the encompassing test of Harvey, Leybourne and Newbold for the null hypothesis that (a) encompasses (b). Each entry represents the percentage of rejection of the null hypothesis. The results of the test are reported at a 10% significance level.

Notes: This table shows the frequencies at which the optimal set contains each model over the 24 hours of each season. In the first block, the loss function is computed with squared forecast errors. In the second one, it is computed with absolute errors. For example, the AR model belongs to the MCS in 50% of cases in winter (seasonal range). We use the \( T_r \) test statistic and a block bootstrap procedure with blocks of 12 observations and 1000 replications for the computation of the \( p \)-values.

(1997). The results of the tests are given at a 10% significance level.

In a final step, we conduct a multiple comparison of the models with the MCS test of Hansen, Lunde, and Nason (2011). Given the high number of specifications considered in our paper, it could be cumbersome to apply the test to all our models. Hence, we focus on successful models in each group (among linear, nonlinear and pooled models) and see which one is selected by the MCS approach. We consider the AR model, the autoregressive model augmented with fundamental variables (AR-X), the MS model with three regimes (MS3_1), the STR model (STR_1) and the Bates and Granger combination (C2), given their good performance according to Tables 3 and 4. Table 6 provides the percentage of times over the 24 hours that each model belongs to the optimal subset of models at the 10% significance level (when the optimal set is smaller than the initial set of models).\(^\text{19}\)

The comparison of the models in Table 3 shows that the inclusion of fundamental variables such as the forecasted demand improves the forecast accuracy of the models: specifications using only the past dynamics of prices display better forecasts in only 14% of cases in terms of MSE (in 19% of cases in terms of MAE or MAPE). This result is consistent with the findings by Misiolek, Trueck, and Weron (2006) for the California power exchange, Karakatsani and Bunn (2008) and Bordignon et al. (2013) for the UK market and Conejo et al. (2005) for the PJM interconnection. The good performance of linear regression with regressors selected with a

\(^{19}\)To compute the model confidence set of Hansen, Lunde, and Nason (2011), we use the Matlab code provided by K. Sheppard.
stepwise procedure (LSTEP) also shows that the restriction of the set of fundamentals to the forecasted demand can be detrimental, since LSTEP outperforms the other individual models in 23% of cases in terms of MSE (and 18–19% of cases in terms of MAE or MAPE). Therefore, it could be interesting to enlarge the comparison of the various specifications with a larger set of regressors.

With respect to the comparison of linear and nonlinear models, Tables 3 and 4 show the good performance of MS models, especially the three-state model MS3_1 allowing a variation of the intercept and the variance. This model often leads to the best MSE, MAE and MAPE in Table 3 and it ranks first among the individual specifications in Table 4 over all seasons with the three criteria. The good performance of MS3_1 seems to be due to its ability to catch the spikes and drops (or less extreme values) of power prices. By contrast, the three threshold specifications rarely outperform the other models in Table 3 and appear among the worse specifications in Table 4.

For this reason, the results of the tests reported in Table 5 are not clear-cut when considering all nonlinear models against the linear ones. The results are more supportive of nonlinear models when we focus on MS models, especially the specification MS3_1 with three regimes. This model outperforms the linear specification in terms of MSE or MAE in 54.9–79.9% of cases (except in winter in the seasonal range). The results of the Diebold–Mariano test are consistent with these findings with a more frequent rejection of the null hypothesis in favour of a better forecast accuracy of MS3_1 than the linear specifications. The null hypothesis that the linear models contain all the information provided by the nonlinear specifications is also more frequently rejected with the encompassing test of Harvey, Leybourne, and Newbold (1997) in the particular case of MS3_1. The MCS test reported in Table 6 also finds MS3_1 superior to the threshold model STR_1 and in most cases to the linear models (AR and AR-X). For instance, for a loss function computed with squared errors, MS3_1 belongs to the optimal set in 66.7% of cases in spring (seasonal range) versus 16.7% for the linear AR model and 45.8% for STR_1.

Despite the good performance of MS3_1, it is clear from the results in Table 3 that no individual model (including MS3_1) exhibits a superior performance for all trading hours and seasons. This instability is favourable to combinations, as shown by the results in Tables 4–6. Table 4 provides the average rank of each model or combination of models. The results are striking. The pooled forecasts always appear among the first best competitors. The combinations over linear models and over all models rank first. As far as the weighting rule is concerned, the Bates and Granger weights and the

---

To get a fair comparison, the specification LSTEP (which contains more regressors than the non-linear models) is excluded in the comparison of the linear and nonlinear models.
simple average are more successful, while more sophisticated rules give larger RMSE. Bordignon et al. (2013) and Nowotarski et al. (2014) obtained similar findings in the context of power price forecasting.

The results of the tests in Tables 5 and 6 also give evidence in favour of combination. The pooled forecasts yield lower RMSE and MAE in 61.9% to 74.6% of cases and the difference is statistically significant in a significant proportion of cases according to the Diebold–Mariano test. The encompassing test rejects the null hypothesis that the individual forecasts contain all the information in the combined forecasts in a large proportion of cases, especially in summer and in fall. The results of the MCS tests in Table 6 are consistent. In each season, the Bates and Granger combination C2 displays the best or second best percentages in the MCS test. For instance, with squared prediction errors, C2 belongs to the optimal set in 79.2% of cases in fall (seasonal range), while the second best (AR-X) appears in the MCS in 66.7% of cases and the less successful model (AR) in only 20.8% of cases.

Finally, the results in Table 5 support estimations on data belonging to the same season rather than to an entire year. Indeed, most seasonal models yield better forecasts than nonseasonal ones, except in summer. More accurate forecasts are obtained using a model or a combination of models estimated in a seasonal range in almost 70 to 80% of cases in winter and in spring and in 55% of cases in fall. The difference is significant in a part of the cases according to the Diebold–Mariano test. In spring and in fall, the encompassing test shows a high rejection of the null hypothesis that the seasonal models do not contain information in addition to the information already included in the nonseasonal models. The results are more disappointing in summer with only 30.4% of cases favourable to the seasonal segmentation and a low rejection rate with the Diebold–Mariano test for a lower or equal forecast accuracy of the nonseasonal models.21 However, the magnitude of forecast errors is much smaller in both cases than in the other seasons.

Interestingly, the gain of the seasonal segmentation is more evident if we restrict the comparison to models without fundamental regressors (the autoregressive and exponential models, AR and EXPO). In this case, the seasonal range leads to lower MSE or MAE than the nonseasonal specifications in 74% of cases over all seasons (versus 58% when we consider all specifications). The improvement is particularly strong in winter and in fall (95.8% versus 69.1% and 70.8% versus 54.5%). Similarly, the gain of the seasonal range is stronger when we focus on the linear specifications (with or without fundamentals). The linear nonseasonal models are beaten by the seasonal ones in 67.7% of cases. Again, the difference is larger in the colder seasons (89.6% in winter and 64.6% in fall). This means that a part of the seasonality is captured by the fundamental variables and by the change of regimes. Overall, our simple treatment of the long-run seasonality present in electricity prices appears to improve forecast accuracy, especially when we use simple specifications for the price forecasting.

VI. Conclusion

The ongoing reorganization of the French electricity market will lead to an increasing volume of trade on the wholesale market. This being the case, developing forecasting tools of electricity spot prices is a key issue for academics and practitioners alike.

In this article, we have investigated the forecasting ability of several classes of time-series models for day-ahead spot prices in France. We have estimated these models with a double temporal segmentation of the data. Our out-of-sample evaluation indicates that this novel approach improves the forecasting ability of the models. Among the nonlinear models, a three-state MS model designed to capture the sudden and fast-reverting spikes in the price dynamics yields better forecasts. However, the forecast accuracy of combinations of models is more stable across all the hours and seasons.

There are a number of potential extensions to this article. In particular, it could be interesting to explore how revised weather forecasts available in the morning before the auction could add extra value beyond the midnight release of demand forecasts. Given the importance of the demand as a

21This result is probably due to a drop in prices in summer 2009 included in the seasonal dataset. This phenomenon reflects the fall in industrial activity during the 2008–2009 recession in France. This is less visible in other seasons where electricity demand is mainly driven by the residential sector (heating effect).
regime driver and the high thermal sensitivity of the French consumption, this appears as a promising way to improve forecasts.

Acknowledgements

We would like to thank Derek W. Bunn, Gorkem Celik, Michel Cruciani, Cathy Dolignon, Mathieu Bouville, Philippe Vassiliopoulos and anonymous referees for their comments and suggestions and Audrey Mahuet for her help in EPEX data. All remaining errors of course are ours.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Bates, J. M., and C. W. J. Granger. 1969. “The Combination of Forecasts.” *Journal of the Operational Research Society* 20: 451–468. doi:10.1057/jors.1969.103.

Bordignon, S., D. W. Bunn, F. Lisi, and F. Nan. 2013. “Combining Day-Ahead Forecasts for British Electricity Prices.” *Energy Economics* 35: 88–103. doi:10.1016/j.eneco.2011.12.001.

Busetti, F., and J. Marucci. 2013. “Comparing Forecast Accuracy: A Monte Carlo Investigation.” *International Journal of Forecasting* 29: 13–27. doi:10.1016/j.ijforecast.2012.04.011.

Chen, D., and D. W. Bunn. 2010. “Analysis of the Nonlinear Response of Electricity Prices to Fundamental and Strategic Factors.” *IEEE Transactions on Power Systems* 25: 595–606. doi:10.1109/TPWRS.2009.2035129.

Conejo, A. J., J. Contreras, R. Espinola, and M. A. Plazas. 2005. “Forecasting Electricity Prices for a Day-Ahead Pool-Based Electric Energy Market.” *International Journal of Forecasting* 21: 435–462. doi:10.1016/j.ijforecast.2004.12.005.

Crespo Cuaresma, J., J. Hlouskova, S. Kossmeier, and M. Obersteiner. 2004. “Forecasting Electricity Spot-Prices Using Linear Univariate Time-Series Models.” *Applied Energy* 77: 87–106. doi:10.1016/S0306-2619(03)00096-5.

Crone, S. F., and N. Kourentzes. 2011. “Segmenting Electrical Load Time Series for Forecasting? An Empirical Evaluation of Daily UK Load Patterns.” Paper presented at the 2011 International Joint Conference on Neural Networks, San Jose, CA, July 31–August 5.

Diebold, F. X., and R. S. Mariano. 1995. “Comparing Predictive Accuracy.” *Journal of Business & Economic Statistics* 13: 253–263.

Diebold, F. X., R. S. Mariano, J. H. Lee, and G. C. Weinbach. 1994. “Regime Switching with Time Varying Transition Probabilities.” In *Nonstationary Time Series Analysis and Cointegration*, edited by C. P. Hargreaves, 283–302. Oxford: Oxford University Press.

García-Martos, C., J. Rodríguez, and M. J. Sánchez. 2011. “Forecasting Electricity Prices and Their Volatilities Using Unobserved Components.” *Energy Economics* 33: 1227–1239. doi:10.1016/j.eneco.2011.07.005.

Granger, C. W. J., and T. Terasvirta. 1993. *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press.

Hamilton, J. 1989. “A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle.” *Econometrica* 57: 357–384. doi:10.2307/1912559.

Hansen, P. R., A. Lunde, and J. M. Nason. 2011. “The Model Confidence Set.” *Econometrica* 79 (2): 453–497. doi:10.3982/ECTA5771.

Harvey, D., S. Leybourne, and P. Newbold. 1997. “Testing the Equality of Prediction Mean Squared Errors.” *International Journal of Forecasting* 13: 281–291. doi:10.1016/S0169-2070(96)00719-4.

Huisman, R., C. Huurman, and R. Mahieu. 2007. “Hourly Electricity Prices in Day-Ahead Markets.” *Energy Economics* 29: 240–248. doi:10.1016/j.eneco.2006.08.005.

Kapetanios, G., V. Labhard, and S. Price. 2006. “Forecasting Using Predictive Likelihood Model Averaging.” *Econometrics Letters* 91: 373–379. doi:10.1016/j.econlet.2005.12.017.

Karakatsani, N. V., and D. W. Bunn. 2008. “Forecasting Electricity Prices: The Impact of Fundamentals and Time-Varying Coefficients.” *International Journal of Forecasting* 24: 764–785. doi:10.1016/j.ijforecast.2008.09.008.

Kosater, P., and K. Mosler. 2006. “Can Markov Regime-Switching Models Improve Power-Price Forecasts? Evidence from German Daily Power Prices.” *Applied Energy* 83: 943–958.

Liu, H., and J. Shi. 2013. “Applying ARMA-GARCH Approaches to Forecasting Short-Term Electricity Prices.” *Energy Economics* 37: 152–166. doi:10.1016/j.eneco.2013.02.006.

Misiorek, A., S. Trueck, and R. Weron. 2006. “Point and Interval Forecasting of Spot Electricity Prices: Linear vs. Non-Linear Time Series Models.” *Studies in Nonlinear Dynamics and Econometrics* 10: 3.

Newbold, P., and C. W. J. Granger. 1974. “Experience with Forecasting Univariate Time Series and the Combination of Forecasts (With Discussion).” *Journal of the Royal Statistical Society. Series A (General)* 137: 131–164. doi:10.2307/2344546.

Nowotarski, J., E. Raviv, S. Trück, and R. Weron. 2014. “An Empirical Comparison of Alternative Schemes for Combining Electricity Spot Price Forecasts.” *Energy Economics* 46: 395–412. doi:10.1016/j.eneco.2014.07.014.

Nowotarski, J., J. Tomczyk, and R. Weron. 2013. “Robust Estimation and Forecasting of the Long-Term Seasonal
Component of Electricity Spot Prices." *Energy Economics* 39: 13–27. doi:10.1016/j.eneco.2013.04.004.

Panagiotelis, A., and M. Smith. 2008. “Bayesian Density Forecasting of Intraday Electricity Prices Using Multivariate Skew t distributions.” *International Journal of Forecasting* 24: 710–727. doi:10.1016/j.ijforecast.2008.08.009.

Popova, J. 2004. *Spatial Pattern in Modelling Electricity Prices: Evidence from the PJM Market*. Working Paper, West Virginia University.

Raviv, E., K. E. Bouwman, and D. Van Dijk. 2015. “Forecasting Day-Ahead Electricity Prices: Utilizing Hourly Prices.” *Energy Economics* 50: 227–239. doi:10.1016/j.eneco.2015.05.014.

RTE. 2012. “La Vague De Froid De Février 2012.” Document De La Direction Economie Prospective Et Transparence. http://www.rte-france.com/sites/default/files/rex_vague_froid-2012.pdf

Stock, J. H., and M. W. Watson. 1999. “A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series.” In *Cointegration, Causality and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, edited by R. F. Engle and H. White. Oxford: Oxford University Press.

Weron, R., and A. Misiorek. 2008. “Forecasting Spot Electricity Prices: A Comparison of Parametric and Semiparametric Time Series Models.” *International Journal of Forecasting* 24: 744–763. doi:10.1016/j.ijforecast.2008.08.004.

Yang, Y. 2004. “Combining Forecasting Procedures: Some Theoretical Results.” *Econometric Theory* 20: 176–222. doi:10.1017/S0266466604201086.

Zou, H., and Y. Yang. 2004. “Combining Time Series Models for Forecasting.” *International Journal of Forecasting* 20: 69–84. doi:10.1016/S0169-2070(03)00004-9.