Extracting relevant features to explain electricity price variations

Frédéric SUARD
CEA, LIST,
Information Models and Machine Learning Laboratory
F-91191 Gif sur Yvette, FRANCE.
Email: frederic.suard@cea.fr

Sabine GOUTIER
EDF R&D
1 avenue du Général de Gaulle
F-92141 Clamart, FRANCE.
Email : sabine.goutier@edf.fr

David MERCIER
CEA, LIST,
Information Models and Machine Learning Laboratory
F-91191 Gif sur Yvette, FRANCE.
Email: david.mercier@cea.fr

Abstract—This paper proposes to explain the variations of energy price, namely the electricity on the German market. Such price variations are described by a set of characteristics which are not totally relevant to explain the variations. We first propose to find explanations by using visual tools in order to draw some preliminary conclusions. Analysing such kind of data is usually done thanks to visual comparison by plotting the curves chronologically. In a second time, we propose to build a statistical model from data. The aim of such approach is to detail the characteristic that get involved in the solution, so that we can automatically extract the most pertinent characteristics. We apply this approach on a set of historical data (2007-2010). Obtained results show that methodology is very interesting, since the conclusion from the statistical modelling enforce the visual analysis and also add details about the explanation.

I. INTRODUCTION

Due to recent evolutions on energy markets, a lot of studies have been achieved in order to predict or analyse energy prices [5], [3], [4], [8]. This task can be decomposed into different steps : first, we have to describe the price by a set of features, then we extract the most relevant ones in order to model the energy price behaviour. This paper focuses on studying the main causes of the electricity price variation. To reach this objective, we propose two ways to analyse energy prices. At first we propose a visual analysis to extract the main causes of energy price variation in section III. One specificity of our approach is to plot the data in a different way than usual time series in order to highlight the interactions between features.

In a second time, we propose to investigate another way to select the relevant features in section IV, by computing a statistical model using some sparse algorithms. Sparse algorithms define a solution with few but pertinent elements. The aim of such algorithm is to define the energy price variation as a linear weighted combination of features. Analysing the weigth of each characteristic helps us to draw some conclusions about the relevance of features.

We apply this method on a dataset composed with real energy data, which are described by six features : CO_2, Gas, Fuel, Oil, Brent and Coal. The results show that our approach is very interesting since we are able to predict the electricity price variation using a training dataset from previous years.

II. CONTEXT AND PROBLEMATIC

For this work we focus on a dataset describing the electricity price. The dataset has been obtained from the German elec-

tricity market set by the European Energy Exchange (EEX)[1].

We gather the forward electricity price for 4 years, from 2007 to 2010. Each product starts three years ahead, that is to say that the product 2007 started in 2004. Since our study concerns the explanations of electricity price variation, we compute the average value over each week, and then extract the logarithmic variation over two weeks :

$$V(w) = \log(P(w)) - \log(P(w - 1)), \qquad (1)$$

where P is the week average value, w the considered week. We have to note that our approach can be easily extended to other time intervals, like days, months or quarters. Using variation instead of absolute value enables us to get free of temporal drift due to money reevaluation. An example of the transformation is shown on figure 1 for the 2007 product. Each week data is described by six characteristics called features : CO_2, Gas, Fuel, Oil, Brent and Coal price variation. All features are converted in euros to be comparable.
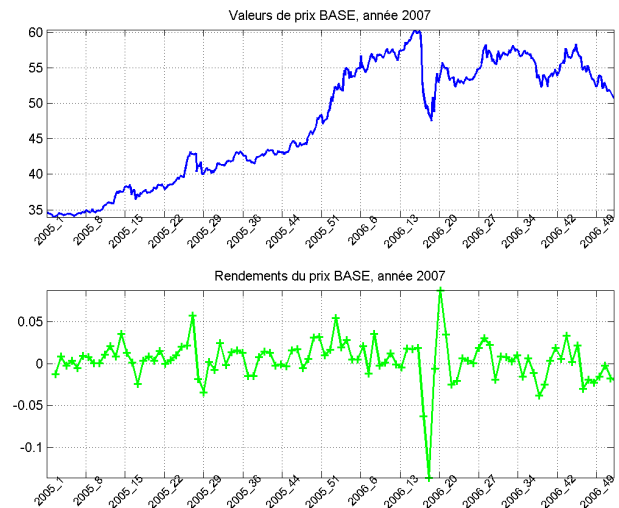

Fig. 1. Example of 2007 electricity price data (top) and computation of variation (bottom).

As a result, each product is described thanks to 150 data of dimension 6. The aim of our work is to explain the price variation by considering all inputs. In other words, we would
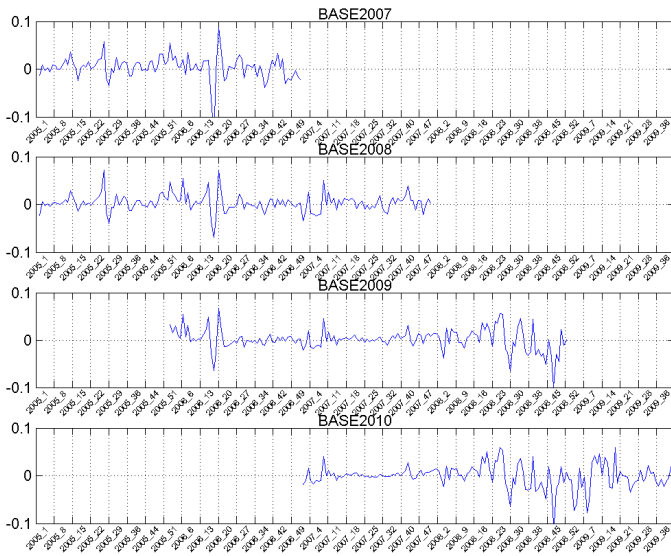
---

[1]http://www.eex.com

Fig. 2. Visualisation of data.

Fig. 4. Visualisation tool to plot all data from a given product (2007).

have to analyse all characteristics in order to extract the most pertinent to explain the energy price variations. The complete dataset is made of 4 products from 2007 to 2010 (figure 2).

III. VISUALISATION TOOLS

The first step is to analyse the data thanks to visualisation tools. The main idea is to reveal some underlying schemes to explain the price variation.

Usually, all data are plotted as time-series, as shown on figure 3. This visualisation can be helpful to explain the causes, if only a unique effect is involved. However, when the explanation comes from a mixture of features, this kind of visualisation fails.
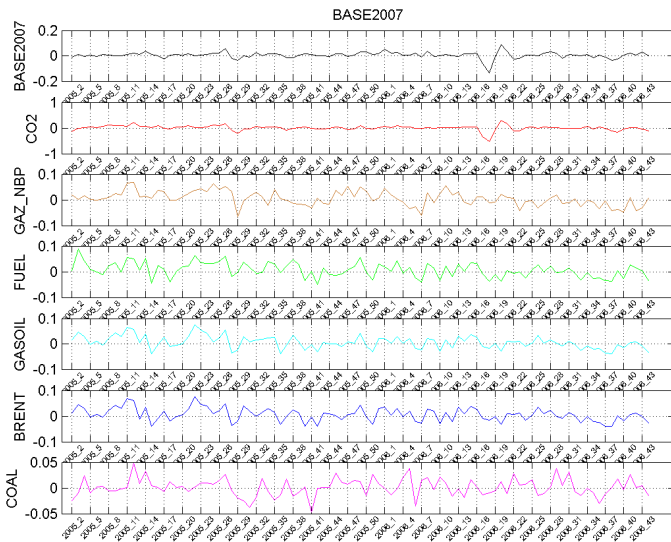

Fig. 3. Example of variation data and its characteristics.

Instead of plotting chronologically all data, we plot them with regards to the features. Since our objective is to explain
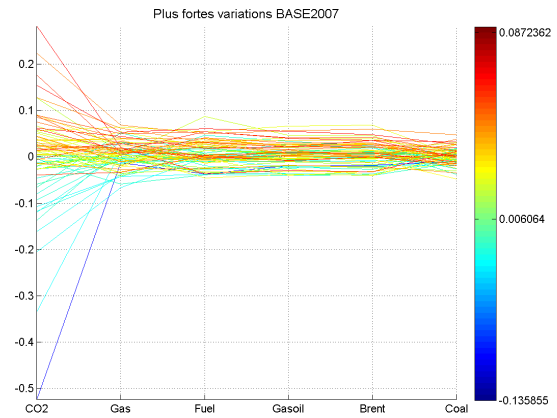
the largest variations, that is to say significant price increase or decrease, we plot data in a different manner, so that large variations would be enlighted.

To reach this goal, we plot each week variation data as a curve. The curve coordinates only depends on the value of the characteristics. As shown on figure 4, the curve is a set of 6 points, which corresponds to the number of characteristics plotted on abscissa. Each point ordinate corresponds to the value of the feature variation.

To add the information of electricity price variation, we colour the curve considering the value of the variation. A dark red curve represents an important electricity price increase, whereas a dark blue curve is significant of a large electricity price decrease. All curves in light green or yellow mean of a medium variation.

Using such representation helps us to conclude about the relation between electricity and characteristics price variation by extracting some set of typical profiles.

Moreover, setting the time period for variation computation can help to extract the main curves, since the number of curves can be easily reduced. On top of that, if the objective is to analyse the highest variations, one can imagine to set a threshold on the electricity price variation to plot only the curves that fit with this constraint.

For example, we can observe a set of ten curves with the same behaviour in 2007, corresponding to a large electricity price decrease (curves in blue), with a high CO_2 decrease and no other variation from the rest of features. This means that when CO_2 is decreasing, the electricity is also decreasing. Moreover, considering the colour of the curves, we can notice that the relation between CO_2 decrease and electricity price decrease is linear. On the contrary when CO_2 is increasing, the relation with the electricity is not linear. Gas has a more linear relation for electricity price increase, as we can notice. In 2007, the main explanations, are based on CO_2 and Gas, with a complementary effect.

In 2008, these explanations are the same, since both Gas and CO_2 can explained the variation. Other characteristics like oil or coal do not seem to have a major impact on electricity

variation. Indeed, looking precisely at these features, no linear relation is visible.

We could explain the impact of CO_2 because of the application of the carbon tax which has been started from 2005 to 2007 [2]. This period corresponds to the evaluation period of product 2007 and 2008. After 2008, CO_2 becomes less influent to explain the electricity variation, since the relation is less linear compared to 2007 and 2008 products. The explanation could be that in 2008 (that corresponds to products 2009 and 2010), quotas on CO_2 have been reevaluated [3].

We can also observe that Coal becomes more pertinent to explain the electricity variation. Particularly for the 2010 product, Coal seems to be an explanation for both increase and decrease, since the relation between Coal and electricity is linear.

To confirm our first conclusions, we propose to compute the correlation values. Generally, in the field of feature selection, the first way to establish a relation between a set of features and a value consists in computing the correlation value. This value is informative about the linear relation that can exist and can measure precisely the intensity of the linear relation.

On table I, we report all values of correlation between the six features and all products 2007-2010. The results we obtain are in agreement with the previous observations from the visual analysis. We can observe that the CO_2 has a major influence for the 2007 product, but its influence decreases over years. However, the Gas relevance increases and seems to be more pertinent than CO_2 in 2010. In opposition to the CO_2, the Coal characteristic becomes more important.

BASE	2007	2008	2009	2010
CO_2	0.8479	0.7631	0.6634	0.5389
GAS	0.3166	0.4592	0.5929	0.6607
FUEL	0.2721	0.2433	0.5566	0.5464
GASOIL	0.2350	0.2035	0.5058	0.5310
BRENT	0.2295	0.1950	0.5077	0.5139
COAL	0.1798	0.2099	0.5845	0.7196

TABLE I
CORRELATION VALUE BETWEEN CHARACTERISTICS AND ENERGY VARIATION.

Since a single feature could not explain completely electricity price variation, we then plot the relation between a couple of features. On figure 5, the scatter matrix displays for each couple of features the electricity price variation. For each graph, the point coordinates corresponds to the value of each feature, and we colour the point so that the colour informs about the electricity variation.

Some particularities could be pointed out. First, the couple (CO_2, GAS) seems to be a good explanation until 2009 (figure 6) since the relation between the pair and the electricity price variation is quite linear. As we could observed previously, the
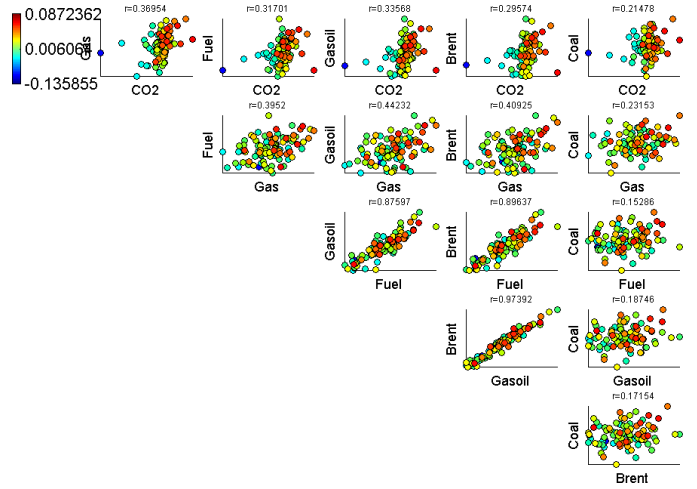
Fig. 5.   Scatter matrix of 2007 data.

couple (GAS, COAL) is a major factor to explain the variation of the 2010 product (figure 7).
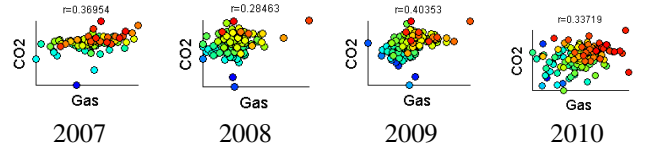

Fig. 6.   Zoom in couple of features (CO_2,Gas) from the scatter matrix.
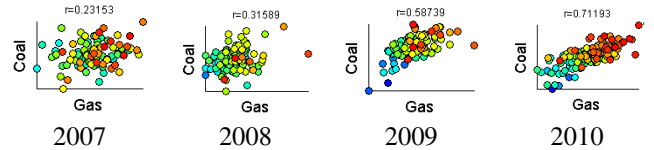

Fig. 7.   Zoom in couple of features (Gas, Coal) from the scatter matrix.

By the way, there is not a unique factor that can explain all electricity price variations. During the observed period, we can observed that the explanations become more complex over years. In fact, a cause is valid over years, but its influence decreases as other factors appears over next years. This first approach to explain price variation proposes some drawings, but it seems that we have to enforce the explanation by updating regularly the observations and the conclusions.

IV. USING STATISTICAL MODELLING

In a second step, we apply some statistical models. We consider a set of n data (in this paper the number of weeks) : $\{\mathbf{x}_i, y_i\}_{i=1...n}$, with $\mathbf{x}_i$ a vector of dimension d (here $d = 6$) and y_i the label associated with data $\mathbf{x}_i$. The label y can be considered here as the electricity variation price, whereas $\mathbf{x}$ contains the characteristics variations values.

The aim is to define a set β of coefficients so that we can define y as a linear combination of features :

$$y_i = \sum_{j=1}^{d} \beta_j x_{ij}. \tag{2}$$

The objective is to propose a feature selection scheme, since we propose to explain the electricity variation by analysing the models generated by statistical modelling. Indeed, we learn the model for each product, and then analyse the value of each coefficient β_j, since the value of β_j is directly linked with the relevance of the feature j.

A. Definition of algorithms

We propose to apply three differents algorithms to estimate the β_j values. They minimize the usual sum of squared errors, with a bound on the sum of the absolute values of the coefficients. Though the effects of correlated data on the parameter estimate could be hidden by classical linear regression. That's why the undesirable symptoms of correlated attributes can be reduced by restraining the size of parameter estimates, or prefering small parameter estimates.

a) Ridge Regression: The first used algorithm is the Ridge Regression [2]. Ridge regression is a linear regression technique which modifies the residual sum of squares computation to include a penalty for large parameter estimates. The overall problem is stated as follows :

$$\min_{\beta} \left\{ \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{d} \beta_j \cdot \mathbf{x}_{ij} \right)^2 + \lambda \sum_{j=1}^{d} \beta_j^2 \right\}, \quad (3)$$

where λ is the regularization parameter, that is to say the level of penalty to control the complexity of the solution. The estimation of β is simply obtained by computing a linear system equation with the given value of λ. The parameter λ has to be set. The optimal value can be fitted by using a validation dataset, or by choosing the optimal solution that is more sparse (see section IV-C).

b) Least Absolute Shrinkage and Selection Operator (LASSO): The second algorithm is the Least Absolute Shrinkage and Selection Operator [7]. This algorithm differs from the Ridge regression by using the L_1 norm instead of L_2. This regularisation is usually known to be more sparse. The optimisation problem is defined by :

$$\min_{\beta} \left\{ \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{d} \beta_j \cdot \mathbf{x}_{ij} \right)^2 + \lambda \sum_{j=1}^{d} |\beta_j| \right\}, \quad (4)$$

where λ is the regularization parameter. In [6], various resolution algorithms are proposed taking into account the way to introduce the regularisation penalty. As the Ridge regression, the value of λ has to be tuned to look for an optimal solution.

c) Least Angle Regression Stepwise (LARS): The last algorithm we propose to study is the Least Angle Regression Stepwise [1]. The idea is to propose to compute the complete regularisation path to tackle the issue of setting λ. The initial problem is posed as :

$$\begin{cases} \min_{\beta} & \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{d} \beta_j \cdot \mathbf{x}_{ij} \right)^2 \\ \text{s.c.} & \sum_{j=1}^{d} |\beta_j| < t \end{cases} \quad (5)$$

with t the regularisation term.

The solution is computed iteratively, by adding into an active set all features that does not fit a constraint. The initial set is empty, for each step, features that are more correlated with residuals are added. The value of each coefficient is determined by estimating the step to do before a novel constraint is not respected.

B. Applying Statistical modelling

We apply these three algorithms to our dataset. For each product, we computed the model. In other words, the model 2007 is computed from the 2007 product.

On figure 8 we plot the curves corresponding to β value for each product year. We also report on table II the exact values for the optimal solution.

The conclusions are quite similar with the previous obtained from visualisation. CO_2 is initially an important factor to build the solution, but gradually its coefficient value decreases. On the contrary, Gas and Coal are growing up.

One limitation of this algorithm resides in the lack of sparsity. Looking at the values, we can notice that the oil characteristics have a null value, so they filter out themselves. This is due to the fact that the Ridge Regression is L_2 regularisation and is not able to eliminate completely a feature from the solution, but shares the weight to average all unnecessary data.
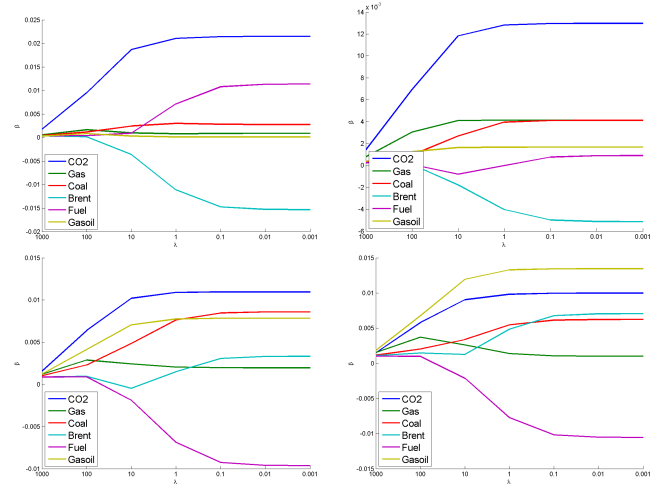

Fig. 8. Results of Ridge Regression Solution.

Year	CO_2	GASNBP	FUEL	GASOIL	BRENT	COAL
2007	0.0211	0.0008	0.0030	-0.0111	0.0072	0.0001
2008	0.0128	0.0041	0.0039	-0.0040	0.0000	0.0017
2009	0.0109	0.0020	0.0076	0.0015	-0.0069	0.0078
2010	0.0098	0.0014	0.0055	0.0048	-0.0077	0.0133

TABLE II
VALUES OF β COEFFICIENT FOR THE RIDGE ALGORITHM ($\lambda = 1$).

In the same way, we plot the value of LASSO coefficients on figure 9 and reported the optimal values on table III. We could observe some similarities with the Ridge Regression and
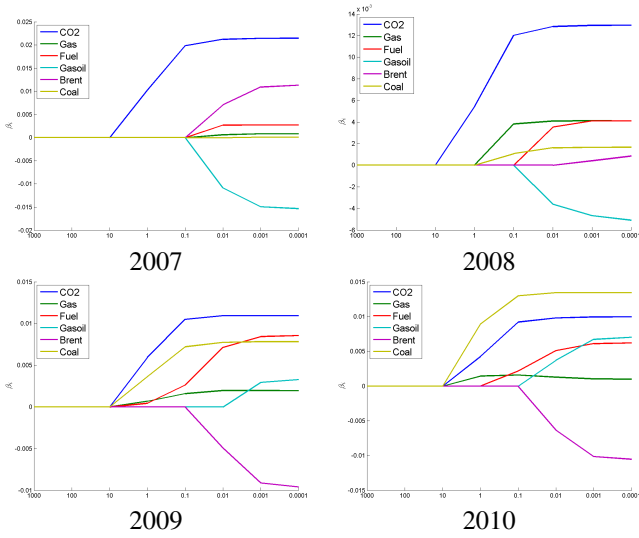
Fig. 9.   Results of LASSO Solution.

Year	CO_2	GAS	FUEL	GASOIL	BRENT	COAL
2007	0.0213	0.0007	0.0027	-0.0108	0.0071	0
2008	0.0128	0.0041	0.0035	-0.0036	0	0.0016
2009	0.0110	0.0022	0.0072	0	-0.0049	0.0078
2010	0.0098	0.0013	0.0051	0.0037	-0.0063	0.0135

TABLE III

VALUES OF β COEFFICIENT FOR THE LASSO ALGORITHM ($\lambda = 0.01$).
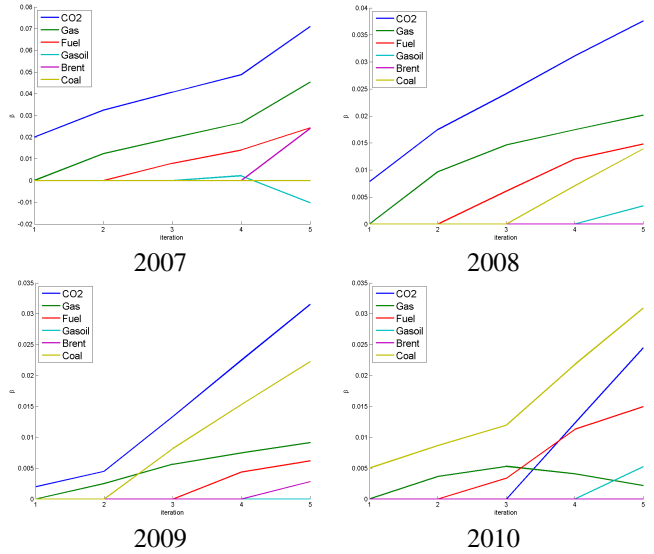
Fig. 10.   Results of LARS Solution.

Year	CO_2	GAS	FUEL	GASOIL	BRENT	COAL
2007	0.0709	0.0453	0.0243	-0.0102	0.0240	0
2008	0.0376	0.0201	0.0148	0.0034	0	0.0139
2009	0.0315	0.0092	0.0062	0	0.0028	0.0223
2010	0.0245	0.0022	0.0149	0.0053	0	0.0309

TABLE IV

VALUES OF β COEFFICIENT FOR THE LARS ALGORITHM (ITERATION 5).

2007	CO_2	GAS	FUEL	GASOIL	BRENT
2008	CO_2	GAS	FUEL	COAL	GASOIL
2009	CO_2	GAS	COAL	FUEL	BRENT
2010	COAL	GAS	FUEL	CO_2	GASOIL

TABLE V

RANKING ORDER OF FEATURES FOR LARS SOLUTIONS.

the visualisation analysis. However, since the LASSO has a more sparse effect on the solution, the model is here able to eliminate completely a feature from the solution.

Finally we plot on figure 10 and report on table IV the solutions for the LARS algorithm. In terms of coefficient values, the global behaviour of LARS is similar to Ridge Regression and LASSO. Similarly to the mechanism of LASSO, the LARS solution is sparse so that some features can be suppressed. This capacity is helpful to keep only the most relevant features.

The ability of extracting from the set of solutions a ranking order between all features is one major advantage of LARS. Indeed, during the resolution, the active set is built iteratively, by adding the most correlated features at a time. So the first feature that is added to the solution is the most correlated one. The difference between the correlation computation stays in the ability to add a couple of features which is equicorrelated after having added previously another feature. It means that if a relevant feature is added first, the next features will be added according to their own effect when the first feature has been added. For example, if a feature is a consequence of another, only the cause feature would be added. On table V we then report the evolution of the feature ordering over products. This aspect is an advantage to compare more precisely the relevance of features.

C. Models validation

Generally, a feature selection approach like correlation or Principal Component Analysis is limited to analysis. Sparse statistical models like LASSO or LARS are intended to include the feature selection step during the solution computation. Such selection can then be approved by applying the model on new data as a validation step. This procedure can also be used to set some parameters of the optimisation problem.

This validation is done by learning a solution on a given product, for example 2007 product, and then predict the electricity price variation of a validation product, here products 2008, 2009 and 2010. The result of such validation is shown on figure 11. As we can see, the prediction fits the real value with a good precision, even on unknown data, that is to say for data after 2009. In particular, the 2010 product has no overlap with 2007 product, however the prediction is good. Globally, the solution tends to predict the good sign of the variation, with a good precision on the real value.

Looking at the performance (table VI), we can notice that the error rate, here the absolute value, increases when the training dataset becomes too old compared to the test dataset. However, if we consider a more recent solution, the error tends to decrease. This can be justified by the fact that the explanation becomes more complex over years. For example,
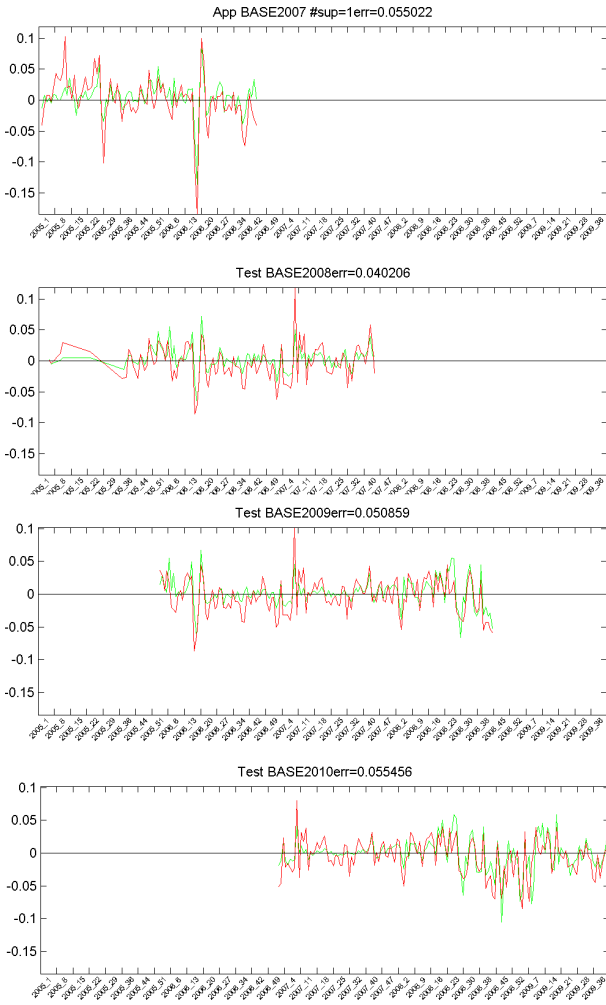
Fig. 11. Application of 2007 LARS solution to 2008, 2009 and 2010 products.

Learning vs. Test		2008	2009	2010
Ridge	2007	0.0165	0.0356	0.063
	2008		0.0245	0.0394
	2009			0.0266
LASSO	2007	0.01566	0.0351	0.0693
	2008		0.0235	0.0407
	2009			0.0275
LARS	2007	0.01566	0.0351	0.0693
	2008		0.0235	0.0407
	2009			0.0275

TABLE VI
ERROR PERFORMANCE OF RIDGE, LASSO AND LARS SOLUTION.

such algorithm is the ability to define the electricity variation as a linear weighted combination of features. By analysing the solution, that is to say the weight associated with each feature, we are able to define the relevance of each feature.

We compare three algorithms : Ridge Regression, LASSO, and LARS applied on the same data. The sparse capacity of both LASSO and LARS to dismiss completely a set of unrelevant features is a major advantage. On top of that, the LARS solution could propose a rank to order the feature.

In order to validate all solutions of statistical models we apply all solutions to predict the value of a test dataset. The results we obtain are very promising, since the solution is able to predict a good value compared with the original electricity variation. These predictions show that the proposed solution is good, since the explanation of a given year are still valuable over years.

In a very close future, this study will be extended to other markets, like the French electricity market. We also envisage to complete the use of statistical modelling, by using some kernel functions in order to face the non-linear relations. Finally, we would also integrated a unique model that could update its parameters to consider the new relations.

looking at the LARS solution, the optimal solution of 2007 product has a size of 1 feature (CO_2), whereas the 2008 solution contains 2 features (CO_2 and Gas) and the 2009 product contains 3 features (CO_2, Gas and Coal). This observation enforces the previous conclusions of visual analysis, when we stated that more features seems to be related with the energy variation. Applying such algorithm would then be constrained to be updated frequently, in order to be able to consider new factors.

V. CONCLUSION AND PERSPECTIVES

In this paper, we propose to analyse a dataset composed of electricity price variation described by a set of characteristics. The objective is to propose different ways to analyse this dataset in order to extract the main explanations. The visualisation tools are useful to draw some preliminary learnings. We can then observe the evolution over the studied years, where the CO_2 feature initially explain the variation, but became gradually less pertinent compared to Gas and Coal.

In order to enforce our visual analysis, we propose to find a statistical model in order to describe the data. The aim of

REFERENCES

[1] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. pages 407–499, january 2003.
[2] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, August 2001.
[3] H. Higgs and A. Worthington. Stochastic price modeling of high volatility, mean-reverting, spike-prone commodities: The australian wholesale spot electricity market. *Energy Economics*, 30:3172–3185, 2008.
[4] J. Lucia and E. Schwartz. Electricity prices and power derivatives : Evidence from the nordic power exchange. 5:5–50, 2002.
[5] S. Muller and D. Mercier. Forecasting electricity price spikes on the french powernext market. In *International Symposium of Forecasting*, 2008.
[6] M. Schmidt. Least squares optimization with l1-norm regularization. Technical report, 2005.
[7] R. Tibshirani. Regression shrinkage and selection via the lasso. 58:267–288, 1996.
[8] M. Verschuere. Modeling electricity markets. *1st Flemish Actuarial and Financial Mathematics Day*, page 25, 2003.