Série des Documents de Travail

n° 2020-06

A Vector Autoregressive Model of Forecast Electricity Consumption in France

Stéphane AURAY[1]
Vincent CAPONI[2]

[1] CREST-ENSAI and ULCO; e-mail: stephane.auray@ensai.fr.
[2] CREST-Ensai, IZA. Email: vincenzo.caponi@ensai.fr.

A Vector Autoregressive Model of Forecast Electricity Consumption in France

Stéphane Auray* Vincenzo Caponi†

February 12, 2020

Abstract

This provides a VARX approach for the estimation of electricity demand in metropolitan France. Our methodology takes into account the complex relationship between weather variables and electricity demand, especially in the short and medium run, and the correlation in the longer run, between electricity and macroeconomic variables. We are able to provide a reliable conditional forecasting that, within the VAR framework, takes into account the common dependency of electricity consumption and other variables. While the VAR approach is not novel within this literature, our main contributions lie on the use of flexible functions that capture the role of weather to explain electricity consumption together with macroeconomic trend and cycle variables, and on the use of very detailed and comprehensive data on actual metered consumption of electricity in France. In-sample and out-sample forecasts provide evidence that our method is reliable for predicting future scenarios conditional on exogenous variables.

JEL: Q43; Q47

Keywords: Electricity Forecast;

*CREST-Ensai and ULCO. Email: stephane.auray@ensai.fr

†CREST-Ensai, IZA. Email: vincenzo.caponi@ensai.fr

1 Introduction

Predicting electricity demand is vital in the energy industry. Forecasting models are used in the literature to predict electricity needs, prices and for predisposing the production of electricity to satisfy the amount of electricity demanded at certain times of the year or day or in different locations. It is also important to recognise that short and long-term demands of electricity (hours to short-term weeks and months to long-term years) respond to different variables, therefore a forecasting model should be flexible enough to capture this behavior. In the short term, weather variables are generally the main drivers of electricity consumption, however, even in the short term weather variables can have different effects when interacting with other economic variables, such as macro variables that capture the cyclically of economic activity, or trend variables that capture structural changes in the use of electricity for economic activity. As an example, we could expect that during hot days in summer electricity consumption increases because of the need of providing conditioned climates in many productive or residential places. However, we could also expect that during a downturn of the economy, this effect is reduced because of a reduced economic activity overall. In the longer term, on the one hand, the evolution of technologies that require less and less energy consumption, or that allow firms and families to produce their own, have a trend impact on the demand faced by traditional providers, on the other hand we also see more and more a shift from other means of energy to more electricity, for example the use of heat pumps instead of other traditional hearting systems based on oil, or even the emerging market of electric cars.[1]

There are several time series models that are used in the electricity demand fore-

[1]There exists an extensive literature on the elasticity of the electricity demand with respect of its price that also highlights this peculiar behavior in time. See among others Filippini (2011), Lim, Lim and Yoo (2014), and more recently for France Auray, Caponi and Ravel (Forthcoming 2020).

casting literature that depend on the availability of data and the range (time domain) of the forecast. AR(I)MA models are generally proposed when the data are relatively rich in terms of the number of observations of the endogenous variable, high frequency and long time span, but there are no or only few covariates that can explain and predict the behavior of that variable. These models are very reliable as long as there are not major changes in the underlying data generating process, that is, as long as the agents that interact in the electricity market do not change their behavior and that the main conditions that explain that behavior cyclically repeat themselves. For example, as long as the night is colder and darker than the day, and summer is hotter and dryer than winter. Yet, even about the weather we cannot be sure about its regularly repeating itself, much less we can trust about other fundamental variables such as technological changes, and other economic and demographic changes.

To address these issues and provide a more sound forecasting in the longer term, a small but significant number of studies offer models based on linear regressions that deal with a larger number of covariates. These models are usually used when the time domain of the study is longer and the forecast is also sought for a longer period, usually years. These models often include economic, weather and demographic variables to explain and predict aggregate energy demand. As an example, Bianco, Manca and Nardini (2009) propose a forecast model based on linear regressions that include macroeconomic variables such as GDP, GDP per capita and population between 1970 and 2007 and a forecast until 2020.[2]

The model we propose here is a natural extension of the two models above and, as such, gives us the opportunity on the one hand to model the behavior of the endogenous series taking into account its regularities in time, on the other hand we extend the model

[2]For a more exhaustive review of different models of energy demand forecasting see Suganthi and Samuel (2012).

to include a series of covariates that the literature has so far suggested be important in predicting the consumption of electricity. Moreover, the VAR model is flexible enough to allow as not to impose causal relationships where it is not clear how the causality works. In our context this is an important issue as we would expect electricity to affect and be affected at the same time other macroeconomic variables. Our model will therefore predict all the variables that enter as endogenous based on their past history and the correlations among them. Yet, some variables are clearly exogenous to our model, in particular weather variables, hence we will use a VARX model, that is a VAR with exogenous regressors.[3]

The use of VAR models for the forecast of electricity consumption is not new[4], however, this approach is very little exploited compared to other approaches that rely more on past information on the endogenous variable only or that limit the extent of covariates to weather variables. Our paper brings several contributions to the literature. First of all, our data on electricity consumption are very accurate and based on actual meter readings of all meters in France. As such, we do not rely on approximations based on the production of electricity, and we have the actual total consumption, not an estimate based on surveys. This is a very important issue, using the production of electricity as a proxy for consumption is very problematic because that production is itself based on the expectation of what the demand would be. We also have electricity consumption for several different types of consumers, which allows us to estimate the behavior of different type of agents before aggregating for the total demand. On the modelling side, other than proposing a VARX, we also model the response of electricity

[3]See Kaytez, Taplamacioglu, Cam and Hardalac (2015) for models that extend the multivariate linear regression approach to the use of artificial neural network. These models contrast to more traditional VARIMAX models in that they are capable to treat non-linearities more efficiently, but at the expense of economic interpretation, which makes them very sensitive to changes in the structure of the economic system in which the forecast needs to be used.

[4]See Ohtsuka and Kakamu (2013) for an example applied to Japan.

demand on weather variables with more flexibility than commonly done and with a higher degree of refinement in terms of the variables used. That is, it is common to use heating and cooling degree days (HDD and CDD), based on the deviation from "desirable" temperatures, as predictors of electricity demand. Those methods, however, assume a V effect of temperatures on consumption. To avoid imposing this structure we proceed by following a semiparalytic approach to estimate the intra-month density of temperature for each month and approximate this density by a fourier transformation, we the use these approximations as variables for our Temperature Response Function (TRF).[5] This gives more flexibility and allows us to estimate rather than assume the threshold points of "desirability". Moreover, we also use all meteo station variables available instead than some weighted average, but because of the strong correlation among them we proceed to making a Principal Component Analysis (PCA) to reduce the variables to a reasonable subset.

For the rest of the paper, next section briefly describes the standard VAR model we use; section 3 the data and the pre-treatment before inputting the into the VARX model; section 4 presents the results of the VARX analysis; Section 5 concludes.

1.1 The VAR Model

The vector autoregressive model is a extensively used tool in macroeconomics and, more in general, in time series analysis. The econometric model is very agnostic about the economic theory and very flexible at the same time, which makes it on the one hand no more than an explorative tool for economic modelling unless we impose more structure on it, on the other hand a very powerful forecasting tool. In the case of VARX models, forecasting is all more interesting as future values of the variable of interest can be

[5] Chang, Kim, Miller, Park and Park (2016)

obtained conditional on the realization of values of the exogenous variables. In our study we can deliver predictions of electricity consumption under alternative scenarios of weather forecast and other exogenous variables.

The VAR is the natural extension of the AR(p) model to a model in which we allow variables to be related one another. Let's take for example two variables $y_{1,t}$ and $y_{2,t}$ that are, for any possible reason, correlated and are also AR processes. In this case we have that,

$$y_{1,t} = \alpha_{10}^1 + \alpha_{11}^1 y_{1,t-1} + ... + \alpha_{1p}^1 y_{1,t-p} + \alpha_{20}^1 y_{2,t} + \alpha_{21}^1 y_{2,t-1} + \alpha_{22}^1 y_{2,t-2} + ... + \alpha_{2p}^1 y_{2,t-p} + \epsilon_{1,t}$$

and

$$y_{2,t} = \alpha_{20}^2 + \alpha_{21}^2 y_{2,t-1} + ... + \alpha_{2p}^2 y_{2,t-p} + \alpha_{10}^2 y_{1,t} + \alpha_{11}^2 y_{1,t-1} + \alpha_{12}^2 y_{1,t-2} + ... + \alpha_{1p}^2 y_{1,t-p} + \epsilon_{2,t}$$

The above two equations represent a structural VAR model, that is, a model in which it is specified the relationship between the two variables of interest and their own lags. This model, however, cannot be directly estimated as the endogenous variables y_1 and y_2 are both in the right and left side of the equations, i.e., the system is not identified. If the structural model is that of interest, as in many economic studies, then we need to impose further assumptions in order to be able to identify it. However, if predicting future values of our the endogenous variables is the purpose of the model, then we do not need to identify the parameters as they are in the above equations. To see why, let's look at the simpler case in which $p = 1$, then the structural VAR becomes,

$$
\begin{aligned}
y_{1,t} &= \alpha_{10}^1 + \alpha_{11}^1 y_{1,t-1} + \alpha_{20}^1 y_{2,t} + \alpha_{21}^1 y_{2,t-1} + \epsilon_{1,t} \\
y_{2,t} &= \alpha_{20}^2 + \alpha_{21}^2 y_{2,t-1} + \alpha_{10}^2 y_{1,t} + \alpha_{11}^2 y_{1,t-1} + \epsilon_{2,t}
\end{aligned} \tag{1}
$$

We can rewrite the system as follows,

$$
\begin{bmatrix} 1 & \alpha_{20}^1 \\ \alpha_{10}^2 & 1 \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \alpha_{10}^1 \\ \alpha_{20}^2 \end{bmatrix} + \begin{bmatrix} \alpha_{11}^1 & \alpha_{21}^1 \\ \alpha_{21}^2 & \alpha_{11}^2 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}.
$$

In matrix form we can write,

$$
BY_t = \Gamma_0 + \Gamma_1 Y_{t-1} + \epsilon_t
$$

where B Γ_0 and Γ_1 are the matrices that collect the coefficients, Y_t is the vector of variables $y_{1,t}, y_{2,t}$ and Y_{t-1} the vector with the same variable and lagged values. This model can be transformed by inverting the B matrix into,

$$
Y_t = B^{-1}\Gamma_0 + B^{-1}\Gamma_1 Y_{t-1} + B^{-1}\epsilon_t
$$

$$
Y_t = \Pi_0 + \Pi_1 Y_{t-1} + \eta_t \tag{2}
$$

The model in equation (2) is called the reduced form of the VAR and can be readily estimated by OLS equation by equation. All the coefficients in the Π matrices are in fact identified, although, without additional assumptions, it is not possible to derive the original structure of the model from the reduced estimation. For forecasting, however, this is not a problem as all we need is in fact an estimate of the Π matrices. Then, forecasting is done in a very similar manner as in the care of an AR() model, that is,

$$Y_{t+1}|t = \hat{\Pi}_0 + \hat{\Pi}_1 Y_t. \tag{3}$$

Notice here that the forecast is done for the whole vector of endogenous variables, which, in turns, increases the predicting power of the model. This is particularly true when we are interested in forecasting variables that are less predictable but which we know are correlated with others that are more predictable.

Finally, an important extension of the VAR model, especially for our study, is to add exogenous variables that can explain our endogenous ones. The model is then very slightly more complicated by adding to the regressions a series of X_t variables with or without lags. The general VARX model becomes then,

$$Y_t = \Pi_0 + \Pi_1 Y_{t-1} + ... + \Pi_p Y_{t-p} + \Theta_0 X_t + \Theta_1 X_{t-1} + ... + \Theta_q X_{t-q} + \eta_t \tag{4}$$

The VARX model can be very powerful and useful for forecasting as it uses more information to predict the future values of the variables of interest. In particular, values of the exogenous variables could be known in advance compared to our variable of interest, which would make the prediction more accurate. Exogenous variables can also be associated to policies, for example a change in tariffs etc..., in which case the VARX will produce reliable predictions on the effect of these changes.

The next section describes the data we use to implement the VARX model and, most importantly, the preliminary steps in treating those data.

2 Data Used and Pre-Treatment

2.1 Electricity Data

Our objective is to predict future values of total consumption of electricity in France. For this purpose we have available the series of realized consumption from Jan. 1st 2010 to Feb. 1st 2018. This data was provided by Enedis. Together with this data Enedis also provided power subscription and number of sites of energy delivery, number of days within a month in which tariffs for TEMPO or EJP customers are more onerous (effacement), plus a series of calendar data such as the number of holidays and working days in a month. In addition to this data we collected data on economic activity in France, namely GDP, total consumption, exports, investments, as well as employment and unemployment, salaries etc...[6] We also dispose of detailed temperatures observations for 32 meteo stations in metropolitan France, at the frequency of every half hour, as well as a calculated weighted average for the whole France (realized temperatures) and normal temperatures for one whole year to be used in repetition for the same day and half-hour every year. We also have a measure of nebulosity, realized and normalized for France at the same frequency as the temperatures.[7]

2.2 Building on Existing Models

In order to better understand the ground basis of our modelling strategy and therefore our specific contribution in terms of accuracy of predictions, we start with a simple model that is the starting point of many forecasting models of electricity consumption. This model is exclusively based on weather variables and in particular its core is the deviation

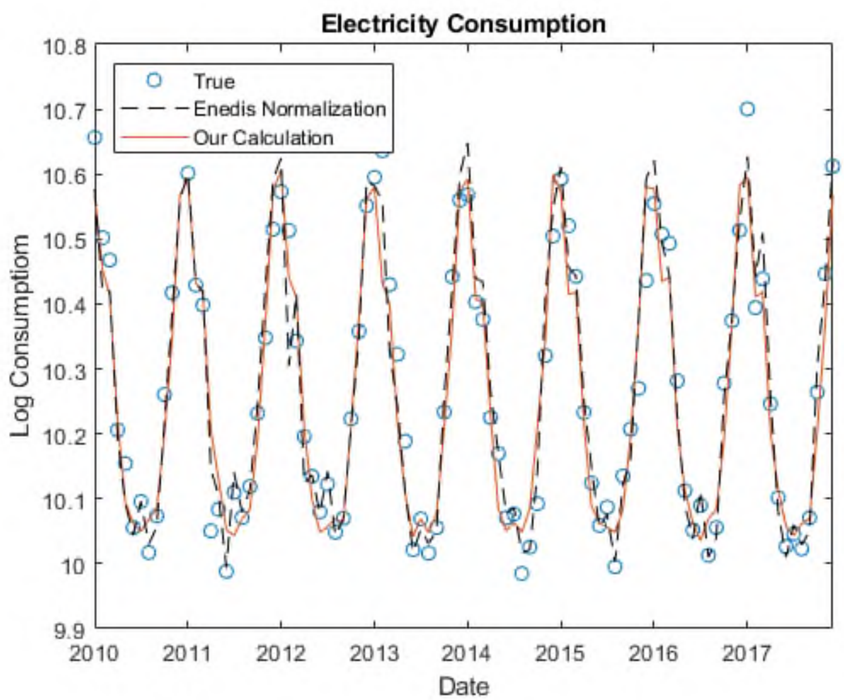
[6] See Appendix 1 for a detailed list of sources.

[7] "Normal" is defined as the average for that day and half-hour in the closest thirty year period, in this case Jan. 1st 1980- De. 31st 2010. The weights to reconstruct the national average from the 32 stations are provided by Enedis and are calculated taking into account electricity consumption.

from "normal" consumption as defined by what it would have been the consumption of electricity if the weather variables considered where at their normal levels as defined in the footnote above. Common models that explain electricity consumption in terms of weather variables (particularly temperatures), are based on the concept of heating and cooling degree days (HDD_R and CDD_R), obtained taking the sum of all the positive intra-day differences between a heating threshold and the realized temperatures (or the realized temperatures and the cooling threshold) within a month. The variables created in this way are then be used as regressors to explain electricity consumption. In our first model, which we call basic, we use these definitions applied to the average national temperature for France, with thresholds 22 and 15 for cooling and heating respectively, together with TEMPO and EJP days, calendar variables (Eff and Days) and a dummy for July 14th (Bastille). All the variables in such a model are exogenous, some are also deterministic (calendar days), while the weather variables are clearly uncertain for forecasting purposes. The logic of the model is therefore to look at what the consumption would have been if the temperatures were "norma". To do so, the coefficients on the HDD and CDD variables are applied to the normalized HDD and CDD, i.e. the variables as they would result from normal temperatures. This exercise gives both, the deviations from "normal" consumption and also the expectation of future consumption assuming temperature will be "normal". Clearly, alternative scenarios can also be constructed for abnormal temperatures as well.

To evaluate this model, other than the usual statistics on R-square and significance of the regressors, we also look at its power to predict out of sample consumption. We therefore run the regression with the realized temperatures on a shorter sample, in particular eight months shorter, and then we look at the sum of the squares of the distances between the forecasted values, which assume normal temperatures, and the actual values. Our basic model tells us that we have an overall variance equal to 0.0138. Figure (1) shows how well the model performs overall.

Figure (1) Forecasting Using the Basic Model



2.3 Adding Flexibility

There are two assumptions in the above method that bring some limitation to our analysis that can be avoided. One first assumption is that the aggregate effect of the climate of each station is a constant proportion of the average climate effect. This is implicit in considering only the weighted climate variable rather than all the single stations. The other assumption is that temperatures have a V effect on consumption (although we actually use a quadratic function, a little more flexible), this assumption is implicit in the construction of HDDs and CDDs. To avoid imposing this structure to our predictive model we proceed by following **?** and semi-parametrically estimate the intra-month density of temperature for each month and approximate this density by a fourier transformation, we then use these approximations as variables for our Temperature Response Function (TRF). This gives more flexibility and allows us to estimate rather than assume the threshold points. Moreover, we also use all meteo station variables instead than a weighted average. However, climate variables of stations in the same country are likely to be very correlated and,therefore having a large number of them will most likely not add much information, for this reason by a Principal Component Analysis (PCA) we reduce as much as we can the dimensionality of our set of variables.

Before proceeding to the use of all meteo stations and the PCA analysis, in order to better appreciate the advantage of using the TRF method, we first apply it to the realized mean temperature of France, so that we can directly compare this method with our earlier method. Figure (2) shows the results from the TRF model compared to the basic model. From the figure is very easy to see that the fit of the TRF model is much closer to that of the basic model, suggesting that the added flexibility dramatically improves the forecasting power of the model. The sum of the squares of the distances between the forecasted values, and the actual values goes from 0.0138 to 0.0089, a change that certifies the dramatic improvement.

The TRF model can also be used to compute the response function to different temperatures. Figure (3) shows the estimated response function for our specification. We can see from the graph that consumption decreases with the increase of temperatures up to a certain point, around 20 degrees, because of less need of heating and then increases again because of the need of electricity for cooling. The increase of cooling is steeper than for heating, this may be sue to the fact that while heating energy can be provided by many different sources, cooling is usually only provided with electrical power.

Figure (2) Forecasting Using the TFR Model

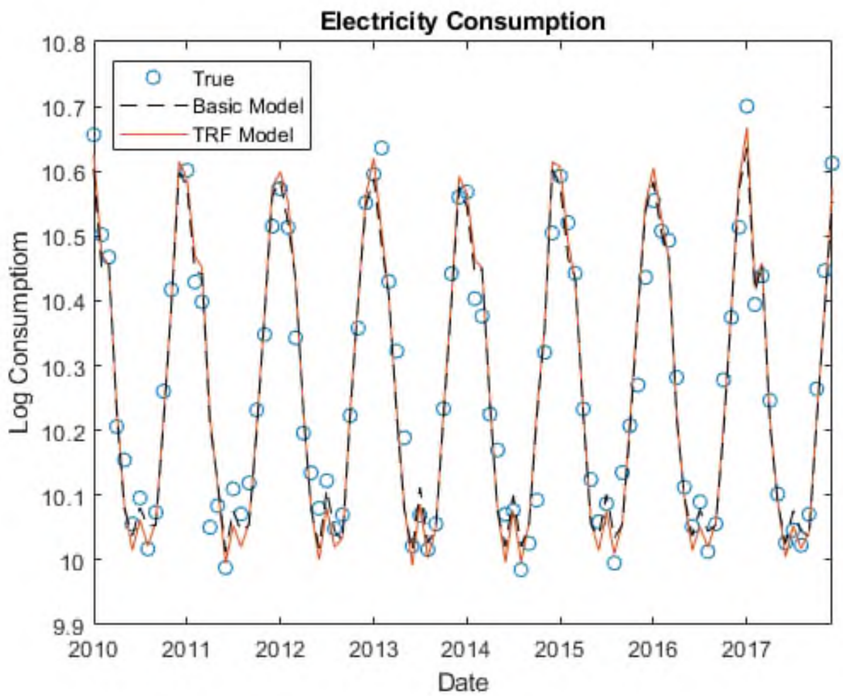
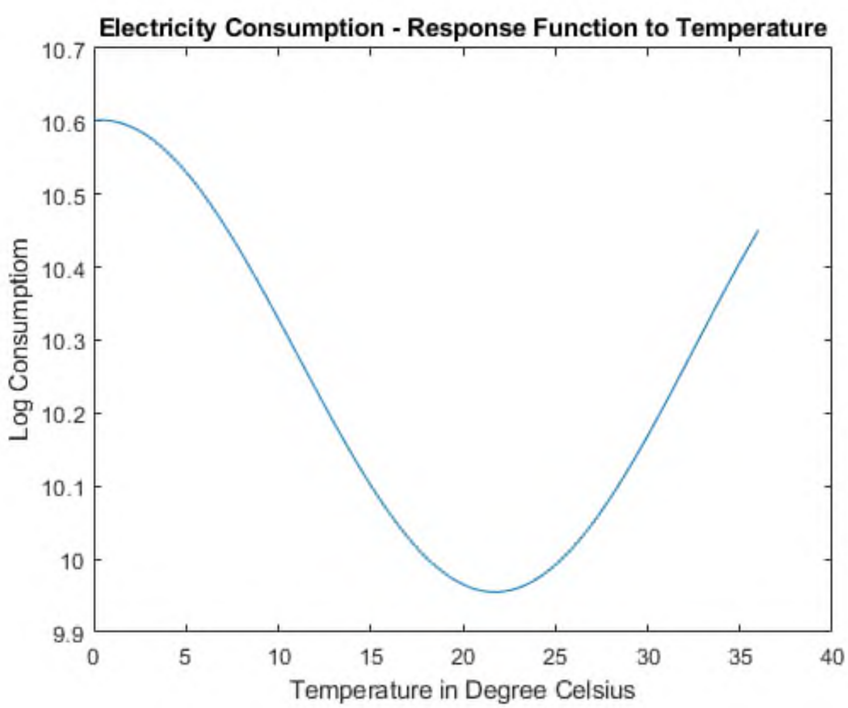


Figure (3) The Temperature Response Function



2.4 Space-Differentiated Weather

Our last step in preparing the data, is the use of information coming from 32 meteo stations rather and, rather than taking a weighted average of them to construct a national measure, we take the first five component of the PCA analysis. Figure (4) shows the fitted models as we cannot, in this case, use the aggregate measure of normal temperatures, and compares the model with the PCA of 32 stations with the one with only one variable. The fit in this case increases as we obtain a R-square of 0.9753, which gives more confidence for our forecast model.

2.5 Harmonizing the Series

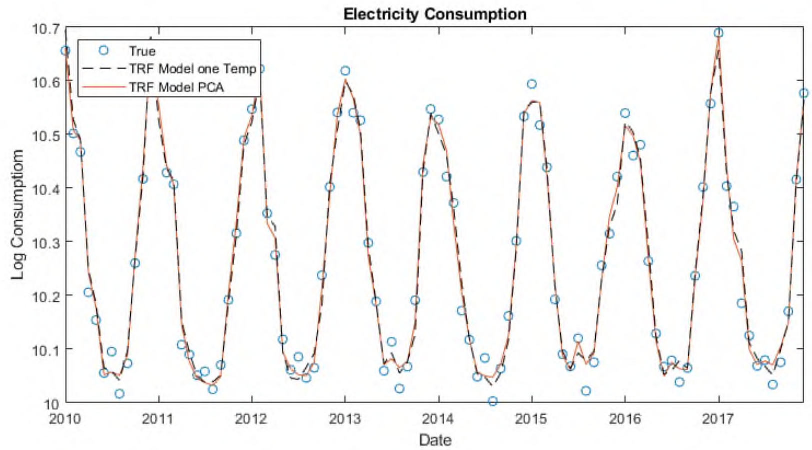
The first step of our analysis was to harmonize the series, that is to make them all at the same frequency. Many economic series such as GDP can be found only at quarterly frequencies, in this case we used the method of cubic spline interpolation, widely used in these cases, to make monthly observations. The caveat in this case is that we do not have a strong monthly seasonality, as with quarterly data that is not possible to be reconstructed. The rest of the series were collected in monthly frequencies. All economic variables were also deflated to give represent real values at 1999 prices.

3 VARX Analysis

In this section we document our VARX forecast analysis. In particular we describe the variables we used as endogenous and as exogenous, we describe some details on the choice of lags and inferencing and discuss the forecast results.

We introduced three types of variables in our analysis: economic variables, weather variables and "technical" variables. Economic variables we believe are important as

Figure (4) Forecasting Using the TFR Model and 32 Meteo Stations



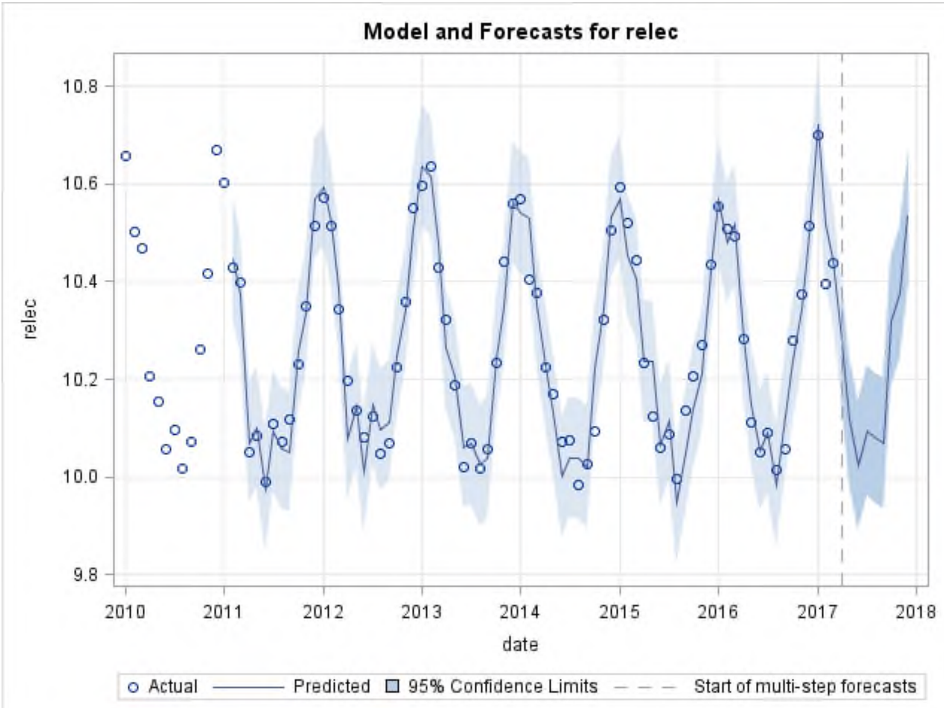
higher economic activity needs more energy, hence we postulate a correlation between electricity consumption and economic activity. Among the variables that describe economic activity we included GDP, total consumption of goods and services and total employment as share of working age population. All variables are used in natural log transformation. For "technical" variables, we indicate all other variables that have a relationship with electricity consumption but are not classified among the first two types. In particular we use the number of days in a month, the number of Saturdays or Sundays, and other variables such as the number of sites that deliver electivity etc... Among all these variables we divide between endogenous and exogenous. Endogenous variables are those that are correlated but may cause or be caused at the same time by electricity consumption, for example GDP we treat is as endogenous. The reason is that for producing more GDP there is the need of higher electricity consumption, therefore there is a causation from GDP to electricity, however, it may be that lower prices of electricity increase its consumption and at the same time GDP, in which case the causation is rather on the other sense. For VAR specification we don't need to take a stand on causation when we are interested only on forecast, it is however important to recognize the possibility of erogeneity. We also include (the log of) wind energy production as an endogenous variable, for the same reason as above. As exogenous variables we include the number of days in a month, the number of Saturdays or Sundays as well as weather variables. In this case we are confident that they are indeed exogenous as, quite obviously, the number of days in a month does not change conditional on energy consumption, the same can be said for weather conditions, although there is increasing evidence that even at micro-climate levels energy consumption might have an effect on temperatures. We tried many different specifications with different sets of variables, while the results change a little, they do not dramatically when the choice is made wisely. We present only a sub set of results.

In order to avoid the problem of unit roots with the VAR system, and in order to treat seasonality, we proceed by taking seasonal differences of the time series, i.e. we use series that represent changes (given the log transformation, percent changes) of one month compared to the same month one year earlier. We loose 12 observation by differentiating, but this procedure is necessary to avoid meaningless results to due spurious correlation arising when integrated series (with unit roots) are used.

Here we present the graphs of all the endogenous variables in which we can see the model fit together with the predicted path for one year ahead. In this first specification we use as exogenous variables the temperature factor variables, as well as number of Sundays, Saturdays, number of days in the month, and tempo days red and white. For all these exogenous variables we have data until December 2017, while for the other variables used in the VARX estimation we have data only until March 2017. Therefore, the month from April to December 2017 are predicted. Indeed, we also have data for actual realization of consumption for those months, which we will compare with our predications.

The following Figure shows the predicted results against the actual ones. In partic-

Figure (5) Forecast of Log Electricity Consumption

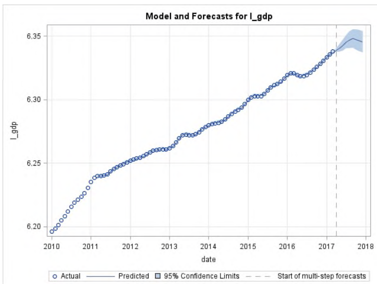


ular, the values on the right of the red vertical line of actual observations were not used to estimate the model, therefore, the difference between the actual and predicted results from April 2017 on can be taken as the forecast errors that were made. We also did not use the number of sites as well as the power subscripted actual values from April 2017, but their forecast as shown in the figures above.

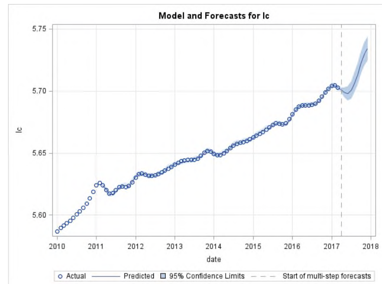
3.1 Diagnostic of the VARX model

The above model estimates a VARX(2,0), that is, two lags were chosen among the endogenous variables and none among the exogenous ones. When it comes to how to specify the model, i.e. the variables to include, which ones endogenous and which exogenous, and the number of lags, there are several factors that help choosing. First of all, economic and technical knowledge can guide in the choice of the variables to

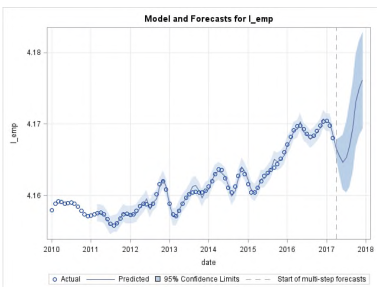
Figure (6)   Forecast of Other Endogenous Variables



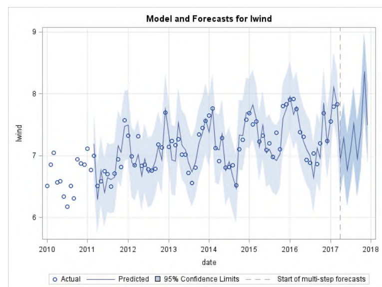
(a) Log of Real GDP



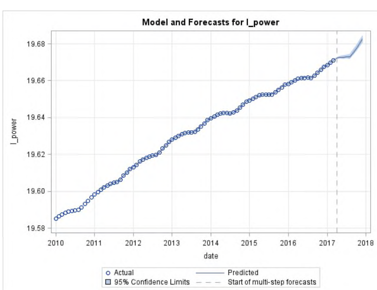
(b) Log of Consumption of Goods and Services



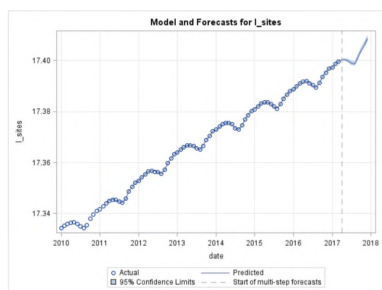
(c) Log of Employment Share



(d)

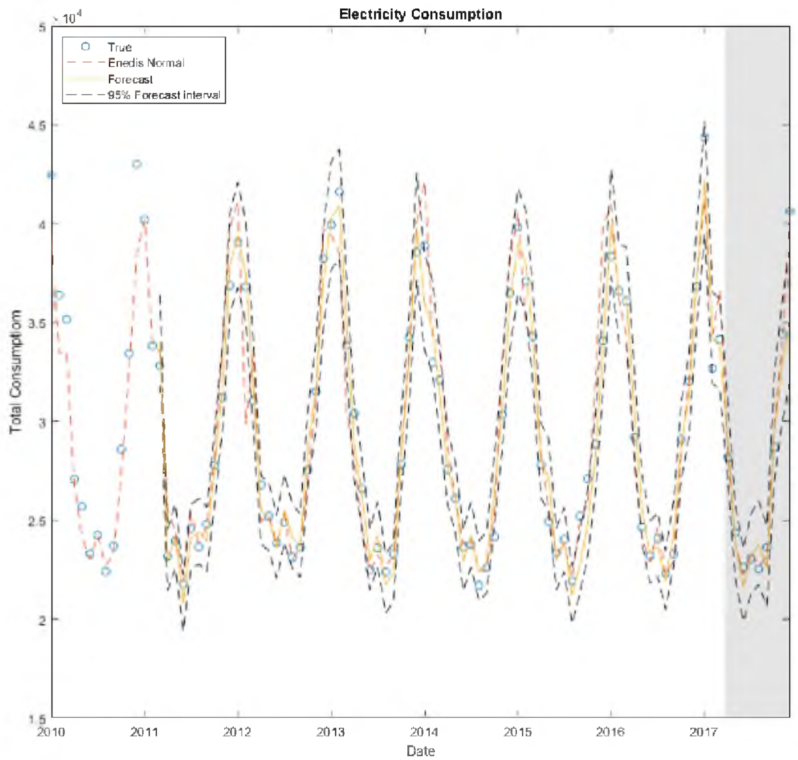


(e) Power Subscription



(f) Number of Sites

Figure (7) Forecast vs Actual Total Electricity Consumption



consider. We discussed this point above. However, in estimating the model we can include or exclude variables based on a series of tests on the significance they have in predicting our variables of interest. The same logic applies also to the choice of lags to include, on the one hand the more lags we include, the more we can explain in the model, on the other hand though, too many variables and lags can make the model difficult to interpret and less efficient in forecasting (i.e. increase the margin of the forecasting error). Table (1) shows the estimates of the autoregressive parameters in the model (for the exogenous parameters there are 7x24, too many to synthesize them here).

The table shows the coefficients table resumes the autoregressive coefficients representing the matrices Π_1 and Π_2 of the VAR system. All coefficients have an associated standard error, but given the number of coefficients it is easier and more informative to conduct tests on groups of coefficients as to test the significance of various parts of the model. In particular we test the significance of all the exogenous variables together, with a Wald test, in their contemporaneous effect and in their lagged effect. Moreover, we also test the effect of each endogenous variable on all the others (excluding on its own). Table (2) resumes the battery of tests we do.

The Table shows interesting results. First of all, the set of exogenous variables are very significant in explaining electricity consumption. To the extent that these variables can be known, or well predicted, in advance (in particular weather variables) this gives more confidence to the forecast power of the model. Among the endogenous variables that seem strongly correlated with electricity consumption are especially consumption of goods and services and the share of employment; GDP, wind electricity production as well as number of sites and power subscribed do not seem very significant. However, when we turn to the whole system, GDP and wind production become also significant while consumption it is not, sites and power remain not significant. This suggests that

Table (1) AR Coefficient Estimates

Lag	Variable	relec	l_gdp	lc	l_emp	lwind	l_sites	l_power
1	relec	0.0689	2.6082	14.8233	0.0467	-26.2543	74.4038	-37.5197
	l_gdp	-0.0002	1.2921	0.0569	0.0004	-0.4484	-0.1974	0.4415
	lc	-0.0052	-0.3106	1.7117	-0.0006	-0.2971	1.609	-1.6856
	l_emp	0.2905	-30.4385	25.9606	0.141	55.0496	-177.9324	269.5842
	lwind	0.0001	0.0667	0.0215	0.0002	1.7158	0.3063	-0.6425
	l_sites	0.0013	0.0319	-0.0078	-0.0001	0.0329	0.7722	0.1085
	l_power	-0.0002	0.0038	0.0096	-0.0002	0.0158	-0.1047	1.0288
2	relec	0.017	-2.9569	-13.7336	0.0657	28.3337	-76.4171	39.8822
	l_gdp	-0.0001	-0.4638	-0.0297	0	0.2855	-0.266	-0.2987
	lc	-0.0059	0.3839	-0.7736	0	0.0765	0.4463	0.4981
	l_emp	1.6399	54.8287	-47.5257	-0.0878	-54.8917	128.4344	-246.2731
	lwind	0.0009	-0.0899	0.0032	0.0006	-0.8948	-0.2911	0.5669
	l_sites	-0.0007	-0.0164	0.0186	-0.0002	-0.017	0.0498	-0.003
	l_power	-0.0015	0.0108	0.0213	-0.0001	-0.0384	0.3376	-0.1401

Table (2) Testing on Parameters Significance

Test	DF	Chi-Square	Pr > ChiSq
All Exogenous Variables on Elec. (Lag 0)	24	37.39	0.04
Log GDP on Elec.	2	0.61	0.7381
Log Cons. on Elec.	2	8.2	0.0166
Log Emp. on Elec.	2	7.94	0.0189
Log Wind on Elec.	2	6.51	0.0386
Log Sites on Elec.	2	3.45	0.1781
Log Power on Elec.	2	1.47	0.4792
Log GDP on System	12	26.69	0.0086
Log Cons. on System	12	20.81	0.0533
Log Emp. on System	12	29.69	0.0031
Log Wind on System	12	29.65	0.0032
Log Sites on System	12	22.29	0.0344
Log Sites on System	12	25.4	0.013

while consumption has an important direct effect on electricity, it is itself explained by other variables, GDP for example. Therefore, in order to provide a good forecast of electricity consumption, we also need to include other variables such as GDP that can help us to provide a good forecast of consumption. This is, indeed, one of the main strengths of the VAR estimation in forecasting as we can easily include variables that can also have an effect through other variables, without necessarily modelling the causal relationship of these effects.

4 Conclusion

We proposed a VARX approach for the estimation and forecasting of the demand of electricity in metropolitan France. We paid special attention to the treatment of weather variables as we know that, in the short to medium run, they are highly correlated to electricity consumption. We showed how relying on a more flexible estimates of the effects of weather variables we can enhance greatly the predicting power of our model. We then turn to the longer run and propose a VARX model, which includes demographic and economic variables. We therefore showed that our VARX model has the property to perform very reasonably not only in the short run, but also in the longer run providing out-of-sample forecasting that are reasonably close to realized data.

References

Auray, Stephane, Vincenzo Caponi, and Benoit Ravel, "Price Elasticity of Electricity Demand in France," *Economics and Statistics*, Forthcoming 2020.

Bianco, Vincenzo, Oronzio Manca, and Sergio Nardini, "Electricity consumption forecasting in Italy using linear regression models," *Energy*, 2009, *34* (9), 1413 – 1421.

Chang, Yoosoon, Chang Kim, James Miller, Joon Park, and Sungkeun Park, "A New Approach to Modeling the Effects of Temperature Fluctuations on Monthly Electricity Demand," *Energy Economics*, 09 2016, *60.*

Filippini, Massimo, "Short- and long-run time-of-use price elasticities in Swiss residential electricity demand," *Energy Policy*, 2011, *39* (10), 5811 – 5817. Sustainability of biofuels.

Kaytez, Fazil, M. Cengiz Taplamacioglu, Ertugrul Cam, and Firat Hardalac, "Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines," *International Journal of Electrical Power & Energy Systems*, 2015, *67*, 431 – 438.

Lim, Kyoung-Min, Seul-Ye Lim, and Seung-Hoon Yoo, "Short- and long-run elasticities of electricity demand in the Korean service sector," *Energy Policy*, 2014, *67*, 517 – 521.

Ohtsuka, Yoshihiro and Kazuhiko Kakamu, "Space-Time Model versus VAR Model: Forecasting Electricity demand in Japan," *Journal of Forecasting*, 01 2013, *32.*

Suganthi, L. and Anand A. Samuel, "Energy models for demand forecasting—A review," *Renewable and Sustainable Energy Reviews*, 2012, *16* (2), 1223 – 1240.

5 Appendix A - Data Sources

As mentioned in the text above, the main source of our data is Enedis, that has provided observations for consumption of electricity, number of holidays, Sundays and Saturdays, within holidays days established as holidays as well (ponts), and the number of days in a month. It also provided the tariffs for Tempo and EJP customers. Temperatures were also taken from Enedis. Other variables we used are total Employment as a fraction of working age population, from INSEE Wind electricity production from "Pègase" of the Ministry of Sustainable Growth and for the aggregate economic series, such as GDP, the IMF. Below Table (3) resumes the sources.

Table (3) Data Used

Source	Name from Source	Name Used
IMF - National Accounts, Current Prices, Non-Seasonally Adjusted		
Gross Domestic Product, Nominal, Domestic Currency	NGDP_XDC	GDP_N
Household Consumption Expenditure, incl. NPISHs, Nominal, Domestic Currency	NCP_XDC	Consumption as share of GDP_N
Exports of Goods and Services, Nominal, Domestic Currency	NX_XDC	Exports as share of GDP_N
Imports of Goods and Services, Nominal, Domestic Currency	NM_XDC	Imports as share of GDP_N
Government Consumption Expenditure, Nominal, Domestic Currency	NCGG_XDC	Government Consumption as share of GDP_N
Gross Fixed Capital Formation, Nominal, Domestic Currency	NFI_XDC	Investments as share of GDP_N
Change in Inventories, Nominal, Domestic Currency	NINV_XDC	Investments (part of) as share of GDP_N
IMF - National Accounts, Constant Prices, Non-Seasonally Adjusted		
Gross Domestic Product, Volume	NGDP_R_K_IX	GDP
Pégase - Èlectricitè, approvisionnement et consommation en France, en GWh - Production brute d'èlectricitè èolienne (en GWh)		Wind
INSEE - Personnes en emploi (taux d'emploi) au sens du BIT - Ensemble des 15 à 64 ans (001688428)		Emp.