Évolution de la demande en électricité à l'horizon 2030

Thomas Da Costa[1,3], Léa Guillaut[2,3]

[1]École Normale Supérieure – Université Paris Sciences et Lettres (ENS - PSL)
[2]Institut d'Études Politiques de Paris (SciencePo Paris)
[3]AgroParisTech – Université Paris-Saclay

Résumé -
À faire :
- Pour des time-series seules : univariate analysis.
  - **Tester quelles variables sont stationnaires autour d'une trend déterministe** $\rightarrow$ ça nous permet de savoir ensuite si on peut faire des ARIMA, etc. Plot Autocorrelogram by using (partial) autocorrelation function. If the PACF of residuals is out of the confidence interval for a given lag k, the process has to be respecified as regards the choice of p or q. Ljung-Box + Shapiro Wilk over residuals.
  - unit root : determine whether a time series variable is non-stationary and possesses a unit root, meaning it exhibits a stochastic trend. If a time series has a unit root, it implies that the series follows a random walk and that shocks to the system have permanent effects, making it non-mean reverting. **Elliott-Rothenberg-Stock (ERS) Test and KPSS Test**.
- Multivariate time series = dynamical modelization of a vector of time series.
  - BIC over ARDL model to choose variables and lags.
  - If non-stationary, go in log then first difference (approximation of the growth rate). If non-stationary, OLS is inconsistent.
  - Reproduire QM1-PS5 en controlant la consommation d'électricité par la taille de la population et en retirant IRC. On choque l'indice des prix à la consommation. Short-run restriction (ordered data) with 10 lags.
  - Structural VAR : ordering of the endogenous variables from the most exogeneous : IPC > Prix de l'électricité > Consommation d'électricité (corrigée de la taille de la population) > PIB.
  - Si structural VAR, GIRF et choc sur le prix de l'électricité // choc sur l'inflation.
  - **Vector AutoRegression** $\rightarrow$ Impulse Response Function, construire des chocs sur le prix de l'électricité $\rightarrow$ voir ce qu'il se passe sur la consommation. See E2-PC4.
  - Regarder les codes R du chap 2 ici.
  - Dire que notre SVAR souffre d'un omitted variable bias en n'ayant pas pris en compte l'IRC. Interesting to add a Markov Switching process to account for IRC (up or down). Or use an Error Correction Model if our variables are non-stationary (Co-integrated VAR).
- Utiliser le package ts pour gérer les time series.
- Account for structural break in 2020 ? see E2-PC1.
- **Shapiro-Wilk test** for normality of residuals (small samples). Jarque-Bera test is better for large samples.
- On peut se contenter d'interpréter le tableau de régression tout prêt.
- Investiguer la méthode LASSO ou RIDGE pour sélectionner les variables.
- Utiliser des BIC plutôt que AIC pour sélectionner les variables.
- Transformer IPC en taux d'inflation : $\pi = 100 \cdot \frac{IPC_t - IPC_{t-1}}{IPC_{t-1}}$?
- Prévision 2030 : si on utilise des AR process, on peut aisément faire des prévisions par régressions successives.
- Intéressant d'avoir $log(c_{elec})_t = \alpha + \boldsymbol{\beta X_t} + \boldsymbol{\gamma X_{t-1}} + \chi log(c_{elec})_{t-1} + \varepsilon$? $\rightarrow$ Breusch–Godfrey test for autocorrelation (The regression models to which the test can be applied include cases where lagged values of the dependent variables are used as independent variables in the model's representation for later observations. BG test requires the assumptions of stronger forms of predeterminedness and conditional homoscedasticity.). If excluding $\chi log(c_{elec})_{t-1}$, the Ljung-Box test can be used (valid assumption of strict exogeneity).

Mots clefs - Économétrie, Régression Linéaire, Prédiction.

1 Notes en vrac

- Synthèse : 5 à 10 pages. Might be something good to write in english.
- Goal : electricity demand modeling, with forecasting for 2030.
- Observer la **relation entre prix de l'électricité et changement dans la structure des moyens de production** ces dernières années → besoins de meilleures variables ? ? Ou juste regarder le PIB, contrôler pour l'IPC, etc.
- **impact des marchés de l'électricité sur la demande d'électricité** → on juge que le prix de l'électricité capture l'effet du "marché" ? ?

1.1 Description

1. Relation économétrique entre les variables.
2. Prévision de la variable dépendante.

The given dataset is composed of 6 time series over electricity consumption, GDP, population, inflation, electricity price, climate index. The data is available from 1990 to 2021, with 32 observations.

1.2 Notes cours C. Doz

1.2.1 Univariate time series

Stationary around a deterministic trend : $X_t = a+bt+Y_t$ where $(Y_t)_t \in Z$ is a stationary process. (Stationarity if esperance and variance does not depend on t).

White noise : variance is constant, no autocorrelation, mean is zero.

Wold theorem : any stationary process can be written as a linear combination of white noise. $X_t = m + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$.

Lag operator : $LX_t = X_{t-1}$. Now, if $(X_t)_t$ is a stationary process :

- $AR(p)$ process : $X_t = \mu + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t$.
- Best Linear Forecast of an $AR(p)$ process : $X^*_{t+1|t} = \mu + \sum_{i=1}^{p} \phi_i X_{t+1-i} + \varepsilon_t$.
- Moving average process $MA(q)$: $X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$.
- $ARMA(p,q)$ process : $X_t = \mu + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$.
- **If $(X_t)_t$ is an $ARMA(p,q)$ process, autocorrelation should tend exponentially to 0 with increasing lags.**
- On these processes, the impact of a shock is transitory.

In our case, electricity consumption might not be ARMA processes, so we might not be able to forecast as such... it is to be tested. We have to test which variables can be considered stationary around a deterministic trend.

Now, if $X_t = \mu + X_{t-1} + \varepsilon_t$ (random walk), we have :

- ARIMA : $(1-L)^d \Phi(L) X_t = \mu + \theta(L)\varepsilon$
- On ARIMA process, the impact of a shock is permanent.
- Autocorrelation of X_t don't exponentially tend to 0 with increasing lags.
- Identication and estimation of an $ARIMA(p,d,q)$:
  1. choice of d : visual inspection of the estimated autocorrelogram + unit root tests (see below).
  2. If $(X_t)_t$ appears to be non-stationary, study $(1-L)X_t$, etc...
  3. Choose the smallest d such that $(1-L)^d X_t$ appears to be stationary.
  4. choice of (p,q) : compute $Y_t = (1-L)^d X_t$ and apply to Y_t the procedure which has been presented for ARMA(p,q). Estimate an ARMA model for Y_t.

1.2.2 Multivariate time series

Let's consider a vector of time series $(X_t)_t$ with $X_t = (X_{1t}, X_{2t}, \dots, X_{kt})$. We suppose that $(X_t)_t$ is a stationary process.

Wold theorem : If $(X_t)_t$ is a stationary process and $(\varepsilon_t)_t$ is a white noise, then $(X_t)_t$ can be written as a linear combination of $(\varepsilon_t)_t$:

$$X_t = m + \sum_{i=0}^{\infty} A_i \varepsilon_{t-i}, \quad A_0 = I, \sum_{i=1}^{\infty} A_i < \infty$$

VAR(p) : $X_t = \mu + \sum_{i=1}^{p} \Phi_i X_{t-i} + \varepsilon_t \Leftrightarrow \Phi(L)X_t = \mu + \varepsilon_t$.

It's not clear that IRC is a good variable to include in the VAR model : it does not seem to be an AR(1) process, or at least not with our granularity.

To do an IRF : Cholesky decomposition to have an orthogonalized impulse response function.

1.3 Notes Ferrara - Doz

1. Data analysis
2. Model specification
3. Parameter estimation
4. Model validation by tests
5. Macro use of the model for forecasting and policy analysis

Bootstrap on residuals is valid if the residuals are white noise and the process is stationary.

ARDL : $Y_t = \alpha + \sum_{j=1}^{m} \beta_j X_{t-j} + \sum_{j=1}^{m} \gamma_j Y_{t-j} + \varepsilon_t$.
The model specification is generally carried out using information criteria.

About Structural VAR : Structural shocks are supposed to be white noise processes and orthogonal to each others.

we could use short-run restrictions with Cholesky decomposition, but also Local Projection à la Jordà (2005) or sign (long-run) restrictions à la Uhlig (2005).

1.4 Regression

For $n = 32$ observations, $k = 6$ variables, we use the following linear regression model :

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Ordinary Least Square (OLS) method is used to estimate the coefficients $\boldsymbol{\beta}$. It holds over the following hypothesis :

1. More observations than explanatory variables
2. Absence of multicollinearity $\rightarrow$ useful to check if we are using lags.
3. Explanatory variables rely on data and the error term is random
4. The expected value of the error term is zero
5. Errors are not autocorrelated
6. Errors are homoscedastic
7. The error term follows a normal distribution

If the hypotheses 4, 5 and 6 are verified, the error term is white noise. The Durbin-Watson test is used to check the absence of autocorrelation. The Breusch-Pagan test is used to check the homoscedasticity of the error term. The Shapiro-Wilk test is used to check the normality of the error term.

The R^2 coefficient is used to evaluate the goodness of fit of the model. It measures the proportion of the variance in the dependent variable that is predictable from the independent variables. The adjusted R^2 is used to compare the goodness of fit of models with different numbers of variables.

Confidence interval for $\beta_j \in \{\beta_1, \beta_2, \ldots, \beta_k\}$ is given by a Student's t-distribution with $n - k - 1$ degrees of freedom. (*is it useful ?*).

1.5 Nos résultats

Graphique 1 : évolution des variables en base 100. On peut déjà intuiter une influence de la croissance du PIB, de la population et de l'indice des prix à la consommation (IPC) sur la demande d'électricité. On remarque aussi comme les fluctuations de l'Indice de Rigueur Climatique (IRC) viennent influencer la consommation d'électricité des ménages. On note enfin que la croissance forte du prix de l'électricité depuis 2009 semble corrélée à la baisse tendancielle du taux de croissance de la consommation d'électricité. Soulignons qu'à partir de cette période, le taux de croissance des prix dépasse celui de l'inflation.

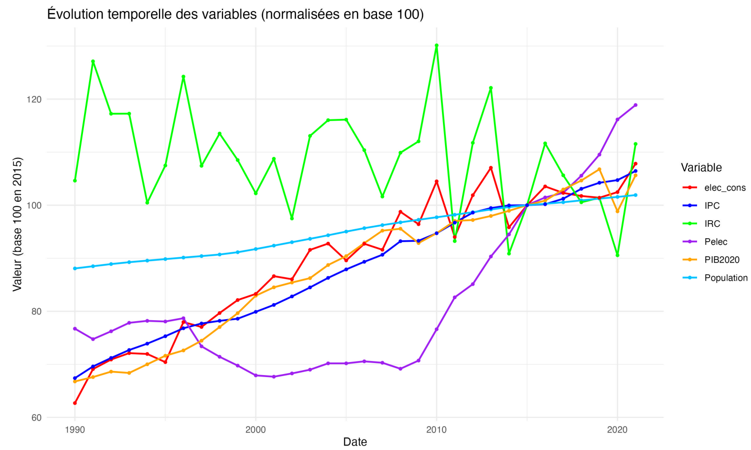

FIGURE 1 – Évolution des variables en base 100.

2 Introduction

3 Les élastomères, approche théorique

4 Matériels et méthodes

5 Résultats et discussions

6 À propos du prototype

7 Conclusion

8 Documents complémentaires