

Évolution de la demande en électricité à l'horizon 2030

Thomas Da Costa^{1,3}, Léa Guillaut^{2,3}

¹École Normale Supérieure – Université Paris Sciences et Lettres (ENS - PSL)

²Institut d'Études Politiques de Paris (SciencePo Paris)

³AgroParisTech – Université Paris-Saclay

Résumé -

À faire :

- Weight par la population pour la consommation d'électricité.
- Pour des time-series seules : univariate analysis.
 - **Tester quelles variables sont stationnaires autour d'une trend déterministe** → 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-Rothemberg-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** → Impulse Response Function, construire des chocs sur le prix de l'électricité → 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 a **structural break**? see E2-PC1 + use a **Chow test** in 2009. Cannot do that in 2020 as we need more observations than predictive variables. We also need homoscedasticity in both subsamples. Account for 2020 with predictive Chow test? It is allowed even if we have less observations than predictive variables (*It will actually be easier if there is no structural break detected as, if we want to keep 3 years for a RMSE proximity measurement, we will lack of data*).
- **Anderson-Darling 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 : probablement pas nécessaire, à moins d'une forte multicollinéarité.
- Utiliser des BIC plutôt que AIC pour sélectionner les variables.
- Transformer IPC en taux d'inflation : $\pi = 100 \cdot \frac{IPC_t - IPC_{t-1}}{IPC_{t-1}}$?
- Prévion 2030 : si on utilise des AR process, on peut aisément faire des prévisions par régressions successives.
- Intéressant d'avoir $\log(c_{elec})_t = \alpha + \beta X_t + \gamma X_{t-1} + \chi \log(c_{elec})_{t-1} + \varepsilon$? → **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).

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"??

On n'a *a priori* pas d'information sur la relation analytique entre les variables : on ne suppose aucune contrainte sur les estimateurs.

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.
- Identification 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 $(\varepsilon_t)_t$:

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

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

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.

$$\text{ARDL} : Y_t = \alpha + \sum_{j=1}^m \beta_j X_{t-j} + \sum_{j=1}^m \gamma_j Y_{t-j} + \varepsilon_t.$$

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 OLS Regression

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

$$Y = X\beta + \varepsilon$$

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

Ordinary Least Square (OLS) method is used to estimate the coefficients β . 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 terms follow a normal distribution

If the hypotheses 4, 5 and 6 are verified, the error term is white noise.

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

Si l'hypothèse de non colinéarité n'est pas vérifiée, l'estimation du modèle est impossible (elle nécessiterait d'inverser une matrice singulière) alors que pour toutes les autres hypothèses l'estimation est possible mais donne un estimateur biaisé et/ou non efficace (à variance non minimale) mais il existe des corrections possibles. La normalité des erreurs est quant à elle non obligatoire mais permet de tirer de bonnes propriétés.

1.4.1 QQ plot

The QQ-plot shows that the error terms are normally distributed.

Kolmogorov Smirnov : Various studies have found that, even in this corrected form, the test is less powerful for testing normality than the Shapiro–Wilk test or Anderson–Darling test.

The Shapiro–Wilk test is known not to work well in samples with many identical values. Jarque-Bera is bad for small samples \rightarrow **Anderson-Darling** is better.

Student and **Fisher** tests are used to test the significance of the coefficients : they depends on the normality of the residuals.

1.4.2 homoscedasticity

The Goldfeld–Quandt test is not very robust to specification errors. The **Breusch-Pagan** test is designed to detect only linear forms of heteroskedasticity. White test is more general and can detect a wider range of forms of heteroskedasticity, but cannot be used for small samples.

heteroskedasticity \rightarrow weighted regression.

1.4.3 Autocorrelation

Durbin–Watson statistic (or Durbin's h statistic), which is only valid for nonstochastic regressors and for testing the possibility of a first-order autoregressive model (e.g. AR(1)) for the regression errors. The Breusch–Godfrey test has none of these restrictions, and is statistically more powerful than Durbin's h statistic. The Breusch–Godfrey test is considered to be more general than the **Ljung-Box** test because the latter requires the assumption of strict exogeneity, but the Breusch–Godfrey test does not. However, the **Breusch–Godfrey** test requires the assumptions of stronger forms of predeterminedness and conditional homoscedasticity.

1.4.4 Multicollinearity

Commonly occurs in models with large numbers of parameters.

Use of Variance Inflation Factor (VIF) to detect multicollinearity. No clear threshold for VIF, and there is often a misunderstanding on how to deal with multicollinearity.

If multicollinearity is detected, the following methods can be used :

- Remove one of the variables or combine them (not recommended)
- Use principal component analysis
- Use ridge regression : regularization method.
- Use LASSO : variable selection and regularization method.
- Use Elastic Net : linear combination of LASSO and ridge regression.

1.5 Non-linear regression

Mention the possible use of a Markov Switching process to account for IRC (up or down), or a price effect (low growth or strong growth) — use Catherine Doz's course for theoretical background.

What is shown by Lantz : Gauss-Newton method (Taylor linearization before OLS), or polynomial regression by minimising Mallows C_p to know the polynomial order.

1.6 Forecast (see p.115)

Bootstrapping is useful when the error terms are non-normal. "L'application des méthodes de bootstrap sur les modèles de régression permet d'approximer la distribution des erreurs de prédiction par leur distribution empirique lorsque celle-ci est inconnue. Le bootstrap est ainsi particulièrement utile lorsque les échantillons de données sont de petite taille et qu'il n'est pas possible de postuler que les erreurs ont une distribution gaussienne"

Guess for a growth rate with 'predict' and add fluctuations based on the regression slope and the IRC fluctuation ?

- Estimation MCO du modèle
- Prévision MCO
- Initialisation de la boucle bootstrap
- Boucle bootstrap
- Construction de l'intervalle de prédiction bootstrap

Les mesures d'erreur quadratique moyenne et d'erreur moyenne absolue permettent de mesurer l'écart entre les prévisions et les observations lorsqu'on effectue des prévisions sur des données rétrospectives. Les indicateurs obtenus à partir de la statistique U de Theil sont utilisés sur des prévisions rétrospectives afin d'évaluer si les erreurs de prévisions retranscrivent un effet de biais, de variance ou, de préférence, un effet de covariance.

1.7 À reproduire ! (p91 du poly)

"Modèle de consommation d'électricité aux Etats-Unis expliquant le logarithme de la consommation d'électricité par habitant par le logarithme du revenu par habitant en monnaie constante et le logarithme du prix de l'électricité en monnaie constante."

"En effet, d'un point de vue économique, la consommation d'un bien dépend d'un effet de revenu et d'un effet de prix auxquels peuvent s'ajouter, le cas échéant, d'autres effets. Ici, on conserve la spécification la plus simple où l'utilité d'un bien dépend d'un effet de richesse et d'un effet prix."

Dans notre cas, on ajoute l'effet de l'IRC et approxime les revenus par le PIB¹.

1. Il existe une relation positive entre le PIB par habitant et le revenu des ménages par habitant, modulée par les politiques fiscales, la répartition des revenus, les structures économiques et la dynamique du marché du travail.

1. Relation Positive :

- Un PIB/habitant élevé est généralement associé à un revenu des ménages plus élevé, car la croissance économique favorise l'emploi, les salaires et les opportunités économiques.
- Plus l'économie produit de biens et services, plus les revenus distribués dans l'économie sont susceptibles d'augmenter.

$$\ln \frac{C_{elec}}{pop} = \alpha + \beta_1 \ln \frac{PIB}{pop} + \beta_2 \ln \frac{prix_{elec}}{IPC} + \beta_3 IRC + \varepsilon \quad (1.1)$$

Attention : le PIB est en euros constants de 2020 : on doit le convertir en euros constants de 2015. $PIB_{2015} = PIB_{2020} \times \frac{IPC_{2015}}{IPC_{2020}}$.

Les coefficients β_1 et β_2 sont, respectivement, l'élasticité de la consommation d'électricité par rapport au revenu et l'élasticité de la consommation d'électricité par rapport au prix. Ces coefficients mesurent la variation en pourcentage de la consommation d'électricité par rapport, respectivement, à une variation en pourcentage du revenu et la variation et à une variation en pourcentage des prix.

On fait ensuite un test de Chow sur 2009 (crise économique ? ; mais **se renseigner aussi sur la mise en place de la libéralisation des prix**²).

Cusum est peu puissant (on accepte facilement l'hypothèse nulle). On peut ceci dit ajouter un test de Chow pas à pas entre 2007 et 2012.

Puis test de Cusum-square pour attester du structural break. On refait une régression sur 2009-2020, et on renouvelle le test de Cusum-square pour vérifier la stabilité temporelle du modèle.

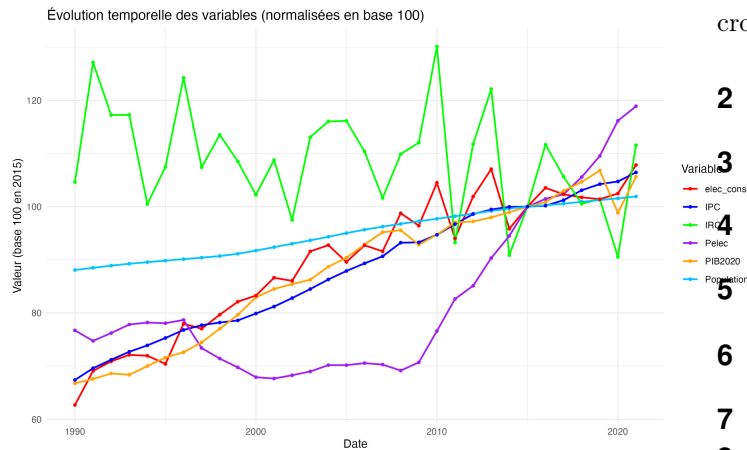


FIGURE 1 – Évolution des variables en base 100.

croissance des prix dépasse celui de l'inflation.

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

1.8 Our results

Graphique 1 : évolution des variables en base 100. On peut déjà intuitivement voir 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

2. Différences en raison des prélèvements et transferts

- Taux d'imposition et de cotisations sociales : Un PIB élevé peut ne pas se traduire directement par un revenu élevé des ménages si les impôts sont importants.
- Redistribution : Des politiques de redistribution (prestations sociales, aides, etc.) peuvent influencer la relation entre les deux.
- Revenus des entreprises : Une partie du PIB est réinvestie par les entreprises sous forme de profits et ne se traduit pas directement en revenu des ménages.

3. Effets du marché du travail

- Si le taux de chômage est élevé, le PIB/habitant peut être relativement stable, mais le revenu des ménages peut être faible en raison d'un accès limité aux revenus du travail.
- La productivité du travail joue un rôle clé dans le lien entre le PIB et les salaires.

4. Effets sectoriels

- Un PIB dominé par des secteurs capitalistiques (finance, industrie lourde) peut générer un écart important entre PIB/hab et revenu/hab car les revenus du capital sont concentrés et peu redistribués aux ménages.
- À l'inverse, une économie fondée sur les services à forte intensité de main-d'œuvre peut avoir un lien plus direct entre PIB et revenu des ménages.

2.

- juillet 2007 : éligibilité de tous les consommateurs (dont les clients résidentiels) au marché concurrentiel de l'électricité.
- Entrée en vigueur du traité de Lisbonne le 1er décembre 2009.
- 7 décembre 2010 : loi NOME (Nouvelle Organisation du Marché de l'Électricité) qui vise à favoriser la concurrence sur le marché de l'électricité (*e.g.* mise en place de l'accès régulé à l'électricité nucléaire historique).