

Projet d'Économétrie

Modélisation de la consommation d'électricité en France et prévisions à horizon 2025

Emmanuel BELFIORE
Victor THIÉBOT



OB1 - Économétrie II
Professeur : Frédéric LANTZ

Janvier 2024

Table des matières

1	Introduction	2
2	Données utilisées	2
2.1	Description du jeu de données	2
2.1.1	Variables utilisées	2
2.1.2	Quelques statistiques générales	3
2.1.3	Visualisation des données	3
2.2	Rupture temporelle et restriction de l'échantillon	4
2.3	Suppression <i>a priori</i> d'une variable de population	4
3	Estimation d'un modèle de régression linéaire	5
3.1	Expression du modèle	5
3.1.1	Visualisation des nouvelles variables utilisées	6
3.2	Estimation du modèle	6
3.2.1	Modèle 1	6
3.2.2	Modèle 2	6
3.3	Vérification des hypothèses du modèle	7
3.3.1	Autocorrélation	8
3.3.2	Hétéroscédasticité	8
3.3.3	Normalité des résidus	9
3.4	Multicolinéarité	10
3.5	Amélioration du modèle	10
3.6	Choix du modèle utilisé pour les prévisions	11
3.6.1	Critère de qualité	11
3.6.2	Interprétation économique des coefficients du modèle final	11
4	Prévisions à horizon 2025	11
4.1	Obtention de valeurs pour les variables explicatives	11
4.1.1	Modèle pour le prix de l'électricité	12
4.1.2	Modèle pour la rigueur climatique	12
4.1.3	Modèle pour la population	13
4.1.4	Prévisions pour les variables explicatives	13
4.2	Prévisions du modèle	13
4.2.1	Commentaires	14
5	Conclusion	14

1 Introduction

Le but du présent rapport¹ est de comprendre l'évolution de la consommation électrique par habitants français en fonction de plusieurs facteurs qui pourraient à premier abord l'impacter. Nous allons effectivement nous pencher sur l'influence que pourrait avoir le prix unitaire du MWh, le PIB par habitants, l'indice des prix à la consommation, l'indice de rigueur climatique (indice de température calculé à partir de la température moyenne journalière et d'une température de référence).

Afin de produire une analyse des plus cohérentes, nous allons dans un premier temps présenter un aperçu des données que nous allons traiter puis établir quelques statistiques générales accompagnées de visualisations graphiques. Par la suite, nous allons nous consacrer à l'analyse du modèle économétrique en nous souciant d'éliminer les principales sources d'erreurs auxquelles notre modèle pourrait être sujet.

Puis nous allons procéder à l'interprétation économique du modèle déterminé afin de comprendre si effectivement un lien de cause à effet est envisageable entre nos variables explicatives et la consommation électrique française et d'en saisir sa magnitude. Sur la base d'hypothèses que nous justifierons et d'un raisonnement économique rationnel, nous essaierons de comprendre comment ces variables interagissent entre elles et si, oui ou non, nous pouvons leur imputer une importance statistiquement valable dans l'explication de notre variable dépendante.

Pour pousser l'analyse encore un petit plus loin, nous allons aussi nous servir des données à disposition pour prévoir cette consommation d'électricité sur un horizon de trois ans. En nous basant sur l'analyse factuelle du passé, pouvons-nous anticiper l'évolution de cette consommation en partant du principe que le marché de l'électricité ne sera pas sujet à des événements extraordinaires qui viendraient fortement perturber l'une de nos variables d'intérêt ?

2 Données utilisées

2.1 Description du jeu de données

2.1.1 Variables utilisées

On dispose de données sur la consommation annuelle d'électricité en France de 1990 à 2022 (autrement dit, nous avons $n = 33$ observations). Il est composé de 8 grandeurs : la grandeur cible et 7 variables :

- Celec : la consommation d'électricité en France (en GWh), la grandeur cible
- Date : l'année de l'observation, de 1990 à 2022
- IPC(base100=2015) : l'indice annuel des prix à la consommation (base 100 en 2015), renommée par la suite IPC
- PIB2014 : le produit intérieur brut (en milliards d'euros 2014), renommée par la suite PIB
- Pop1 : la population en France métropolitaine
- Pop2 : la population en France métropolitaine et dans les départements d'outre-mer
- Pelec : le prix de l'électricité des ménages (en €/MWh)
- DJU : l'indice de rigueur climatique ("DJU" pour "degrés jours unifiés", il s'agit d'un indice de température calculé à partir de la température moyenne journalière et d'une température de référence, ici 18°C, plus d'informations sur Wikipédia (cliquer ici), données issues du Ministère de la transition écologique, et de la cohésion des territoires (cliquer ici))

On peut visualiser les données à l'aide d'un tableau :

TABLEAU 1 – Aperçu des données

Date	IPC	PIB	Pelec	Pop1	Pop2	DJU	Celec
1990	67	1 480	125	56 577 000	57 996 401	0,96	323 252
1991	70	1 496	121	56 840 661	58 280 135	1,17	346 798

1. Ce document est généré depuis un fichier écrit en R Markdown. Le code source est disponible sur [GitHub](<https://github.com/victor-thiebot/projet-conso-elec>)

TABLEAU 1 – Aperçu des données

Date	IPC	PIB	Pelec	Pop1	Pop2	DJU	Celec
1992	71	1 520	124	57 110 533	58 571 237	1,08	355 874
1993	73	1 510	126	57 369 161	58 852 002	1,08	356 188

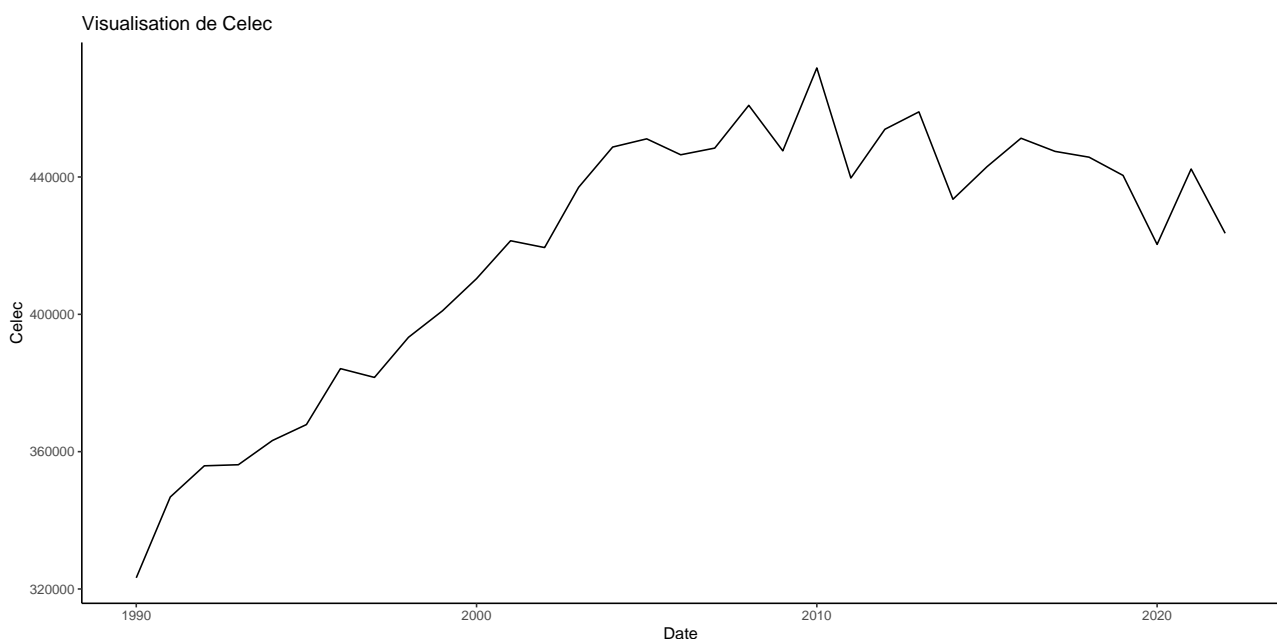
2.1.2 Quelques statistiques générales

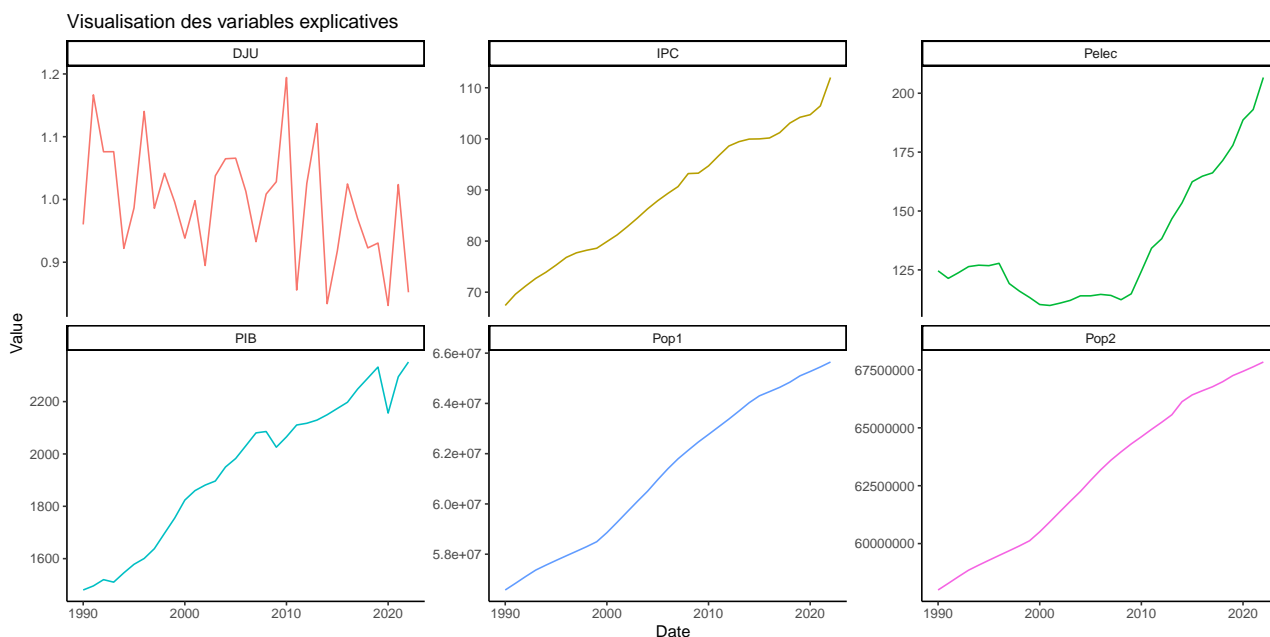
On peut calculer quelques statistiques générales sur les données :

TABLEAU 2 – Statistiques descriptives des données

Variable	Moyenne	Ecart type	Minimum	Maximum
Celec	419301	38948	323252	471762
DJU	1	0.092	0.83	1.2
IPC	89	12	67	112
Pelec	136	28	110	207
PIB	1941	276	1480	2351
Pop1	61208858	3004149	56577000	65646837
Pop2	63012029	3250457	57996401	67842591

2.1.3 Visualisation des données





2.2 Rupture temporelle et restriction de l'échantillon

Au vu de la visualisation de la consommation d'électricité en France, on peut remarquer deux périodes distinctes : une période de croissance marquée entre 1990 et 2008, puis une période de qui semble tendanciellement décroissante entre 2008 et 2022. On peut donc tester la présence d'une rupture temporelle dans la distribution. Si cette hypothèse est vérifiée, on pourra alors estimer deux modèles distincts pour chacune des périodes.

On peut tester la présence d'une rupture temporelle dans la distribution de la consommation d'électricité en France à l'aide d'un test de Chow.

TABLEAU 3 – Test de Chow pour la rupture de 2008

Statistique de test (Chow)	p-value ($\Pr > F$)
115	1.6e-14

La statistique de test a une valeur de 115.25 et la p-valeur associée est de 0. On peut donc rejeter l'hypothèse d'absence de rupture temporelle dans la distribution de la consommation d'électricité en France.

Comme notre objectif final est de pouvoir faire des prédictions à horizon 2025, on se ramène à n'estimer qu'un seul modèle pour le sous-échantillon démarrant en 2008. On peut donc supprimer les observations antérieures à 2008.

On n'a alors plus que 15 observations.

2.3 Suppression *a priori* d'une variable de population

On peut supprimer la variable Pop1 car elle est redondante avec la variable Pop2 (la population en France métropolitaine et dans les départements d'outre-mer). En effet, la population dans les départements d'outre-mer est négligeable par rapport à la population en France métropolitaine. On peut vérifier cela en calculant le ratio moyen entre les deux populations :

TABLEAU 4 – Ratio moyen entre Pop1 et Pop2

ratio moyen	max	min
0,969	0,972	0,968

Ainsi, Pop1 correspond à un ratio moyen de d'environ 0.969 entre 2008 et 2022, autrement dit un ratio très proche de 1 et quasi-constant. On peut donc la supprimer sans risque de perte d'information. Par la suite, on renommera Pop2 en Pop.

3 Estimation d'un modèle de régression linéaire

3.1 Expression du modèle

On peut proposer un modèle fondé sur un raisonnement économique. On cherche à modéliser la consommation d'électricité *par habitant* en France. Cette consommation dépend du revenu par habitant, du prix de l'électricité et de la rigueur climatique, les valeurs monétaires devant être corrigées de l'inflation (c'est déjà le cas pour le PIB mais on utilise pour cela l'indice des prix à la consommation). On obtient ainsi le modèle suivant (on se référera à ce modèle comme modèle 1) :

$$\ln\left(\frac{C_{\text{élec}}}{\text{Pop}}\right) = \beta_0 + \beta_1 \ln\left(\frac{\text{PIB}}{\text{Pop}}\right) + \beta_2 \ln\left(\frac{P_{\text{élec}}}{\text{IPC}}\right) + \beta_3 \ln(\text{DJU}) + \varepsilon \quad (\text{modèle 1})$$

où : $\varepsilon \sim \mathcal{N}(0, \sigma^2)$

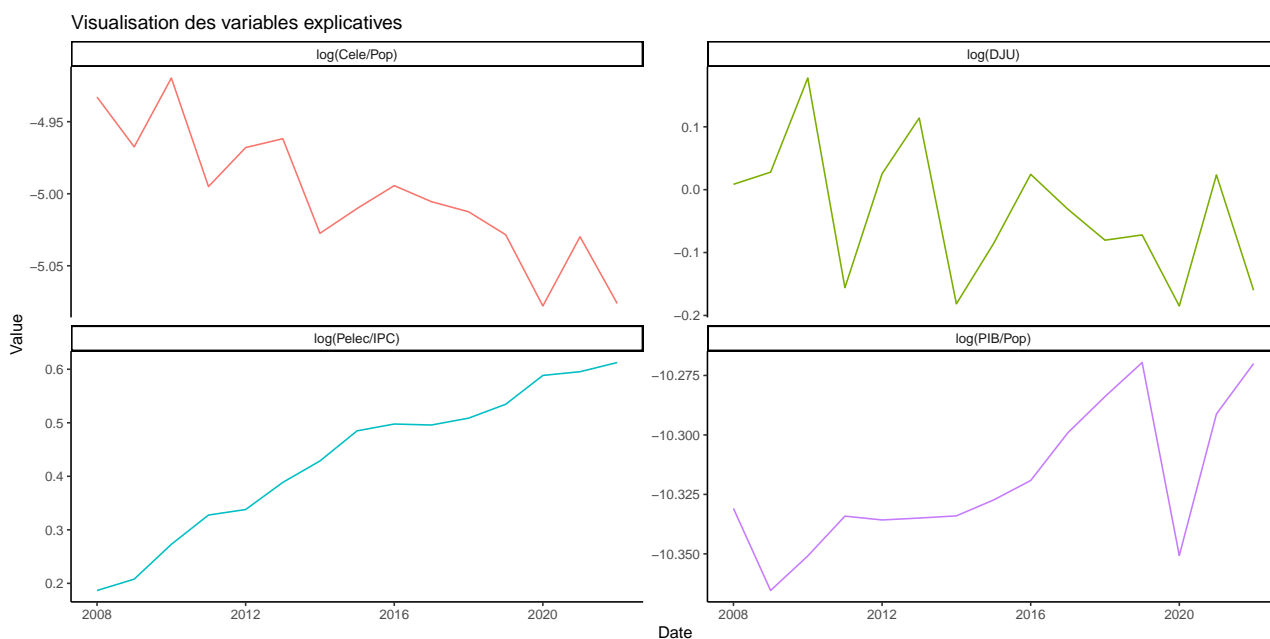
Les coefficients β_i ont alors une interprétation immédiate en terme d'élasticité de la consommation d'électricité par rapport aux variables explicatives (hors le terme constant β_0) :

- β_1 élasticité par rapport au revenu par habitant,
- β_2 élasticité par rapport au prix de l'électricité (en euros 2014),
- β_3 élasticité par rapport à la rigueur climatique.

Une fois le modèle estimé, on pourra facilement repasser à la consommation d'électricité totale en France :

$$C_{\text{élec}} = \text{Pop} \times \exp\left(\beta_0 + \beta_1 \ln\left(\frac{\text{PIB}}{\text{Pop}}\right) + \beta_2 \ln\left(\frac{P_{\text{élec}}}{\text{IPC}}\right) + \beta_3 \ln(\text{DJU}) + \varepsilon\right)$$

3.1.1 Visualisation des nouvelles variables utilisées



3.2 Estimation du modèle

3.2.1 Modèle 1

On estime le modèle 1 par la méthode des moindres carrés ordinaires (MCO). On obtient les résultats suivants :

Coefficient	Estimation	Ecart type	Statistique de test (Student t)	p-value (Pr > t)
β_0	-3.2	1.5	-2.2	0.051
β_1	0.16	0.14	1.1	0.28
β_2	-0.23	0.034	-6.9	2.5e-05
β_3	0.21	0.033	6.2	6.5e-05

D'après les p-valeurs des tests de Student associés aux estimateurs du modèle proposé, on peut considérer que tous les coefficients ne sont pas significativement non nuls.

En particulier, on est tenté de supprimer la variable $\frac{PIB}{Pop}$ car son coefficient β_1 est non significativement différent de 0 et la p-valeur associée est la plus élevée.

Cette proposition peut paraître assez surprenante. En effet, on pourrait penser que la consommation d'électricité par habitant augmente avec le revenu par habitant. Cependant, il faut garder à l'esprit que l'électricité, en tant que bien de consommation, est un bien de première nécessité. Ainsi, la consommation d'électricité par habitant ne dépend pas du revenu par habitant mais plutôt de la population totale. Cela semble donc cohérent de supprimer cette variable.

3.2.2 Modèle 2

On obtient ainsi l'expression d'un modèle plus simple qu'on appellera modèle 2 (on conserve les noms des coefficients du modèle 1) :

$$\ln\left(\frac{C_{elec}}{Pop}\right) = \beta_0 + \beta_2 \ln\left(\frac{P_{elec}}{IPC}\right) + \beta_3 \ln(DJU) + \varepsilon \quad (\text{modèle 2})$$

On obtient alors :

TABLEAU 6 – Résumé du modèle 2

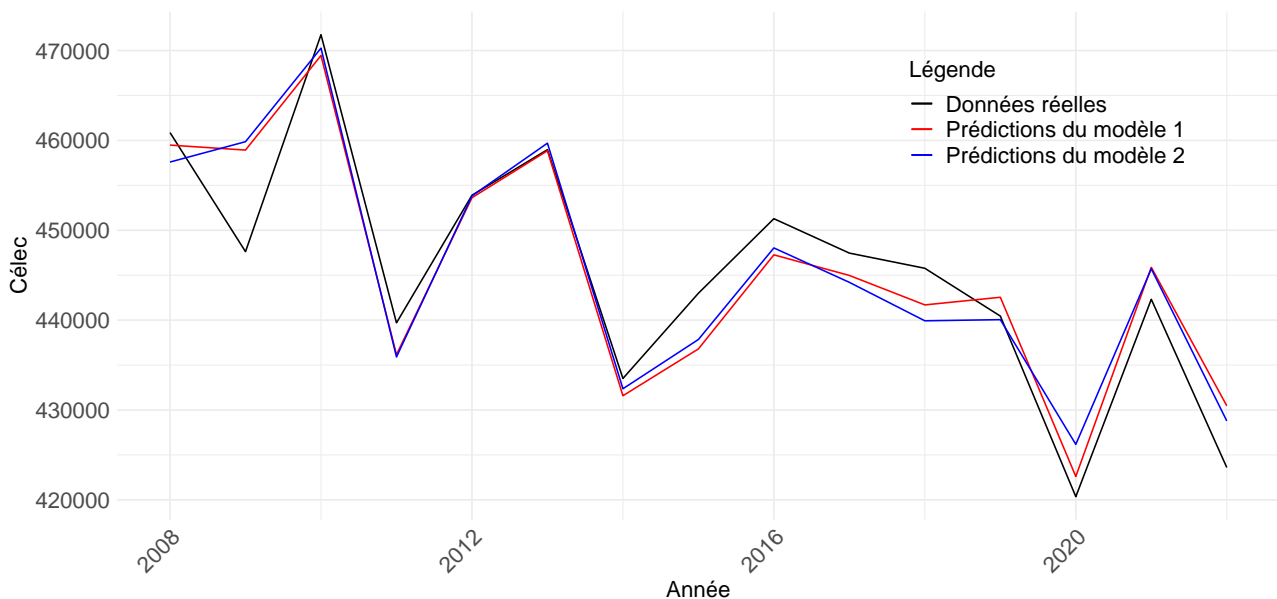
Coefficient	Estimation	Ecart type	Statistique de test (Student t)	p-value ($\Pr > t $)
β_0	-4.9	0.011	-440	1.2e-26
β_2	-0.21	0.026	-8	3.6e-06
β_3	0.21	0.033	6.3	4.2e-05

Le modèle 2 étant un sous-modèle du modèle 1, on peut effectuer un test de Fisher pour tester la significativité globale du modèle 1 (dans R, on utilise la fonction `anova` pour effectuer ce test).

La p-valeur obtenue est de 0.28. On peut donc rejeter l'hypothèse d'absence de significativité globale du modèle 1. On peut donc considérer que le modèle 1 est plus performant que le modèle 2. On peut donc conserver le modèle 1.

Comme ce résultat va à l'encontre de ce qui a été dit plus haut, puisqu'il suggère qu'il faudrait en réalité conserver le modèle 1 malgré la non-significativité du coefficient β_1 , on va tout de même conserver le modèle 2 pour la suite de l'analyse. Par la suite, les analyses seront donc effectuées conjointement pour les deux modèles. On pourra alors comparer les résultats obtenus.

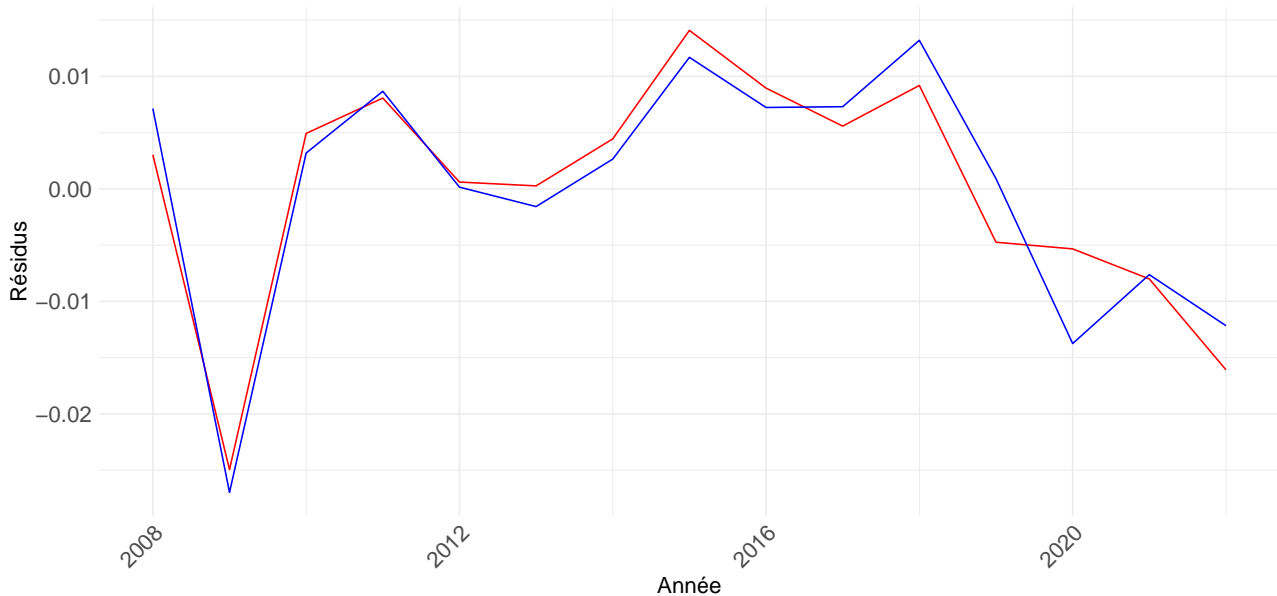
Visualisation de Célec avec les modèles 1 et 2



3.3 Vérification des hypothèses du modèle

Dans cette partie, on va effectuer des tests sur les résidus des deux modèles. Pour se donner une première intuition des problèmes qu'on pourrait mettre en évidence, on peut visualiser les résidus des deux modèles :

Visualisation des résidus des modèles 1 et 2



3.3.1 Autocorrélation

Étant donné qu'on travaille avec des données temporelles, il est nécessaire de vérifier l'absence d'autocorrélation des résidus. En effet, si les résidus sont autocorrélés, les estimateurs du modèle ne sont plus efficaces et les tests de Student associés aux estimateurs ne sont plus valides.

Or, la visualisation précédente semble indiquer la présence d'une autocorrélation d'ordre 1 (au moins) des résidus.

On effectue donc un test de Durbin-Watson pour vérifier cela.

TABLEAU 7 – Test de Durbin-Watson pour l'autocorrélation d'ordre 1 des résidus

Modèle	Statistique de test (Durbin-Watson)	p-value ($\Pr > DW $)
Modèle 1	1.4	0.1
Modèle 2	1.6	0.19

Dans les deux cas, la statistique de test a une valeur inférieure à la valeur seuil pour un risque de première espèce de 5%.

On peut donc rejeter l'hypothèse d'absence d'autocorrélation d'ordre 1 des résidus et ainsi considérer que les estimateurs du modèle ne sont pas efficaces et que les tests de Student associés aux estimateurs ne sont pas valides.

3.3.2 Hétéroscédasticité

On peut également vérifier l'absence d'hétéroscédasticité des résidus pour la même raison : si les résidus sont hétéroscédastiques, les estimateurs du modèle ne sont plus efficaces et les tests de Student associés aux estimateurs ne sont plus valides.

On effectue donc un test de Breusch-Pagan pour vérifier l'absence d'hétéroscédasticité des résidus. On obtient alors :

TABLEAU 8 – Test de Breusch-Pagan pour l'hétéroscédasticité des résidus

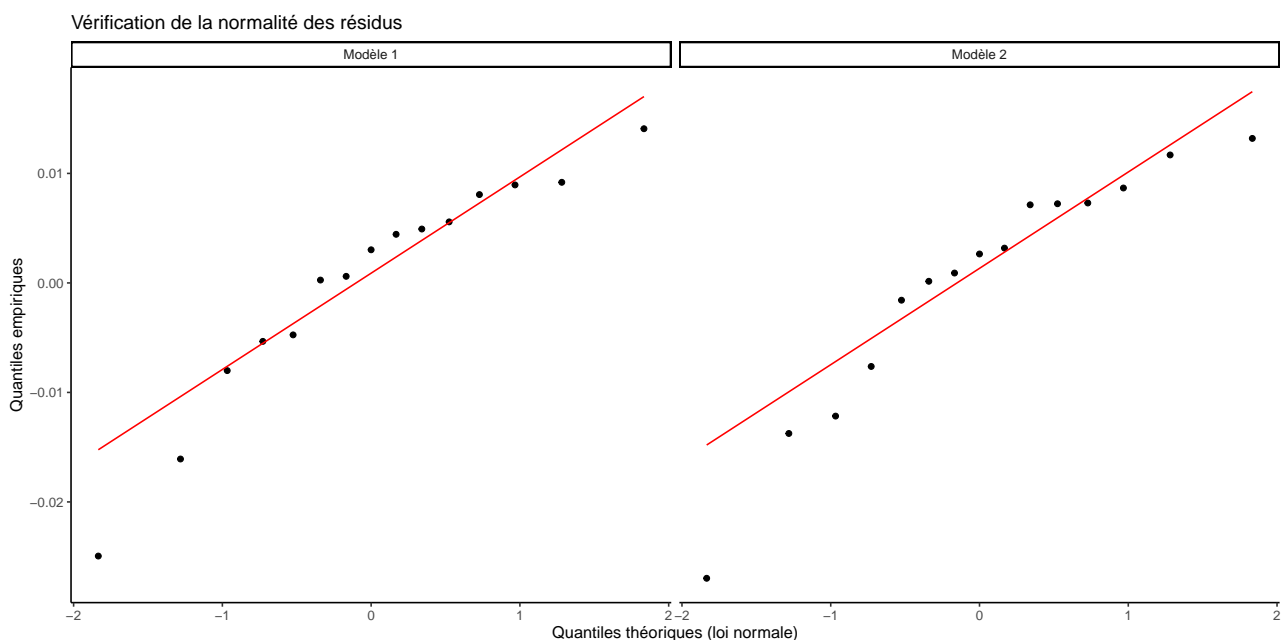
Modèle	Statistique de test (Breusch-Pagan)	Degrés de liberté	p-value (Pr > BP)
Modèle 1	0.96	3	0.81
Modèle 2	1.5	2	0.47

Dans les deux cas, on rejette l'hypothèse d'hétéroscédasticité des résidus.

3.3.3 Normalité des résidus

Il est important de vérifier la normalité des résidus dans le cas étudié ici. En effet, l'échantillon à notre disposition est assez petit (15 observations) et il est donc nécessaire de vérifier que les résidus suivent une loi normale.

Pour ce faire, on peut commencer par se donner de l'intuition avec un QQ-plot des résidus :



Sans surprise, les résidus ne semblent pas suivre une loi normale. Ce n'est pas surprenant car on ne dispose que de peu de données.

Vérifions à présent cela pour nos deux modèles à l'aide du test de Kolmogorov-Smirnov :

TABLEAU 9 – Test de Kolmogorov-Smirnov pour la normalité des résidus

Modèle	Statistique de test (Kolmogorov-Smirnov)	p-value (Pr > KS)
Modèle 1	0.49	0.00066
Modèle 2	0.49	0.00065

Les p-valeurs des tests de Kolmogorov-Smirnov associés aux résidus des deux modèles sont faibles, ce qui nous impose en fait de rejeter l'hypothèse de normalité des résidus (toujours avec un risque de première espèce de 5%).

On peut donc considérer que les résidus ne suivent pas une loi normale.

On peut néanmoins se rassurer en constatant que les moyennes des résidus sont très proches de 0 :

- $-4,2\text{e-}19$ pour le modèle 1
- $-1\text{e-}19$ pour le modèle 2

On peut donc considérer que les estimateurs du modèle sont non biaisés et que la normalité des résidus n'est pas un problème majeur.

3.4 Multicolinéarité

L'échec des deux modèles à passer le test d'autocorrélation (et de normalité) des résidus peut s'expliquer par la présence de multicolinéarité entre les variables explicatives. En effet, on peut remarquer que les variables utilisées dans les deux modèles sont fortement corrélées entre elles. On peut visualiser cela à l'aide d'une matrice de corrélation des variables explicatives :

TABLEAU 10 – Matrice de corrélation des variables explicatives

Variables	log(PIB/Pop)	log(Pelec/IPC)	log(DJU)
log(PIB/Pop)			
log(Pelec/IPC)	0,67		
log(DJU)	-0,28	-0,49	

3.5 Amélioration du modèle

On l'a vu, on est à la fois en présence de multicolinéarités et de résidus autocorrélés.

On choisit donc une solution qui permet de résoudre ces deux problèmes² : la régression ridge.

On estime ainsi un modèle de régression ridge pour chacun des deux modèles. Dans R, on utilise pour cela la fonction `glmnet` du package `glmnet`. Comme on a besoin d'obtenir un coefficient optimal de pénalisation, on utilise la fonction `cv.glmnet` qui permet d'obtenir ce coefficient optimal par validation croisée (*K-fold cross-validation*).

Cette analyse nous permet d'obtenir les coefficients optimaux de pénalisation pour chacun des deux modèles :

- $3.8\text{e-}03$ pour le modèle 1
- $3.8\text{e-}03$ pour le modèle 2

On peut donc estimer les deux modèles ridge correspondants :

TABLEAU 11 – Estimation des modèles ridge

Modèle	β_0	β_1	β_2	β_3	R^2 ajusté
Modèle 1 (ridge)	-4.3	0.054	-0.2	0.2	0.94
Modèle 2 (ridge)	-4.9		-0.19	0.2	0.94

2. qui est en tout cas indiquée en cas de multicolinéarité et qui peut probablement aider à résoudre, par suite, l'autocorrélation observée

3.6 Choix du modèle utilisé pour les prévisions

3.6.1 Critère de qualité

On peut juger de la qualité des différents modèles en s'appuyant sur l'erreur quadratique moyenne (RMSE) :

TABLEAU 12 – Erreur quadratique moyenne (RMSE) des différents modèles

Modèle	RMSE
Modèle 1	0.01003
Modèle 2	0.0106
Modèle 1 (ridge)	0.01058
Modèle 2 (ridge)	0.01087

Ce critère semble placer tous les modèles sur à peu près le même niveau de qualité.

Toutefois, comme les modèles estimés par MCO ne vérifient pas toutes les hypothèses du modèle linéaire, on peut considérer que les modèles estimés par régression ridge sont de meilleure qualité.

Ensuite, pour choisir entre les deux modèles estimés par régression ridge, on adopte un raisonnement de parsimonie : on choisit le modèle le plus simple, c'est-à-dire celui qui a le moins de variables explicatives. **Ainsi, on choisit le modèle 2 (ridge) comme modèle final.**

3.6.2 Interprétation économique des coefficients du modèle final

En se penchant sur l'interprétation économique du coefficient B2, on peut tout de suite remarquer que la relation entre le prix de l'électricité et sa consommation. Cela va de soit puisque plus un bien est cher, plus sa consommation diminuera. La magnitude de la relation, exprimée sous forme d'élasticité par notre modèle logarithmique nous indique que si le prix du MWh augmente de 1%, sa consommation diminuera de 0,21%. Nous avons une p-valeur très proche de zéro ce qui nous mènerait à penser que notre estimateur est statistiquement significatif. En revanche, nous avons vu que toutes les hypothèses nécessaires à garantir la validité de notre modèle ne sont pas vérifiées et, par conséquent, nous ne pouvons pas nous projeter dans des affirmations formelles.

Le coefficient B3 quant à lui est positif ce qui signifierait en effet que si l'indice de rigueur climatique augmente, la consommation électrique française ferait de même. En fait, comme cet indice représente la différence entre la température moyenne et une température de base fixe (18 degrés), alors plus la température diminue, plus il aura tendance à diminuer. Par conséquent on comprend mieux son impact potentiel dans notre modèle. Il nous apprend que si cet indice augmente de 1% alors, la consommation électrique augmentera de 0,21%. A nouveau sa p-valeur est proche de zéro ce qui nous tenterait d'affirmer formellement que notre estimateur est significativement différent de zéro mais comme mentionné ci-dessus, les hypothèses nécessaires ne sont pas vérifiées.

4 Prévisions à horizon 2025

4.1 Obtention de valeurs pour les variables explicatives

Pour faire des prévisions à horizon 2025, il faut obtenir des valeurs pour les variables explicatives du modèle. Plusieurs possibilités s'offrent à nous :

- chercher des données pour les années 2023, 2024 et 2025 à partir de sources fiables (par exemple, les prévisions de l'INSEE ou de l'OCDE) ;
- estimer des modèles de régression linéaire n'intégrant qu'un terme constant et un terme temporel (pour tenir compte de la tendance) et utiliser ces modèles pour faire des prévisions à horizon 2023, 2024 et 2025.

On va ici utiliser la deuxième méthode. Autrement dit, chaque variable explicative x_i sera associée à un modèle de régression linéaire de la forme :

$$x_i = \beta_0 + \beta_1 \text{Date} + \varepsilon$$

On précise qu'on considère comme variables explicatives les variables transformées utilisées dans le modèle 2 (ridge) :

- $\ln\left(\frac{P_{\text{elec}}}{\text{IPC}}\right)$
- $\ln(\text{DJU})$

Par simplicité, on ne vérifie pas les hypothèses du modèle linéaire pour ces modèles de régression linéaire. On se contente d'obtenir des prévisions pour les années 2023, 2024 et 2025 en jugeant qu'on peut utiliser un modèle sur base du test de Student associés aux estimateurs du modèle³.

4.1.1 Modèle pour le prix de l'électricité

On commence par estimer un modèle de régression linéaire pour la variable $\ln\left(\frac{P_{\text{elec}}}{\text{IPC}}\right)$:

TABLEAU 13 – Résumé du modèle pour $\ln(\text{Pelec}/\text{IPC})$

Coefficient	Estimation	Ecart type	Statistique de test (Student t)	p-value (Pr > t)
β_0	-61	3.4	-18	1.4e-10
β_1	0.03	0.0017	18	1.3e-10

4.1.2 Modèle pour la rigueur climatique

On estime ensuite un modèle de régression linéaire pour la variable $\ln(\text{DJU})$:

TABLEAU 14 – Résumé du modèle pour $\ln(\text{DJU})$

Coefficient	Estimation	Ecart type	Statistique de test (Student t)	p-value (Pr > t)
β_0	22	12	1.8	0.089
β_1	-0.011	0.006	-1.8	0.088

Comme les coefficients ne semblent pas significatifs, on se contentera d'utiliser la valeur moyenne de $\ln(\text{DJU})$ pour faire des prévisions pour les années 2023, 2024 et 2025.

L'indice DJU est en effet calculé comme un écart par rapport à une donnée de référence. Par conséquent, plus la température augmente, plus cet écart diminue et plus l'indice diminue. Nous aurions plutôt tendance à penser que les températures ces dernières années auraient une tendance haussière due au changement climatique (et donc l'indice une tendance baissière).

Toutefois, cette diminution ne se traduit pas clairement sur notre graphique. En effet, nous devinons une tendance baissière mais qui demeure très faible par rapport à ce qui pourrait être attendu. Ceci pourrait être expliqué par la transformation successive des données de températures réelles en un indice qui lissé de surcroît par le logarithme. En effet, la transformation logarithmique a tendance à lisser les données et donc à diminuer les écarts entre les valeurs. C'est pourquoi nous avons une tendance baissière mais très faible, donc un modèle qui ne semble pas significatif.

3. Évidemment, si les hypothèses du modèle ne sont pas vérifiées, alors on ne peut pas se fier à la statistique de Student. Une étude approfondie permettrait d'aboutir à des résultats plus robustes mais on cherche ici seulement à obtenir des valeurs qui semblent cohérentes et avec lesquelles on pourra travailler.

4.1.3 Modèle pour la population

La population est une variable explicative du modèle mais elle n'est pas transformée. On va donc estimer un modèle de régression linéaire pour la variable Pop :

TABLEAU 15 – Résumé du modèle pour Pop

Coefficient	Estimation	Ecart type	Statistique de test (Student t)	p-value ($Pr > t $)
β_0	-5e+08	20892007	-24	3.9e-12
β_1	281199	10368	27	7.9e-13

4.1.4 Prévisions pour les variables explicatives

On obtient alors les prévisions pour les variables explicatives :

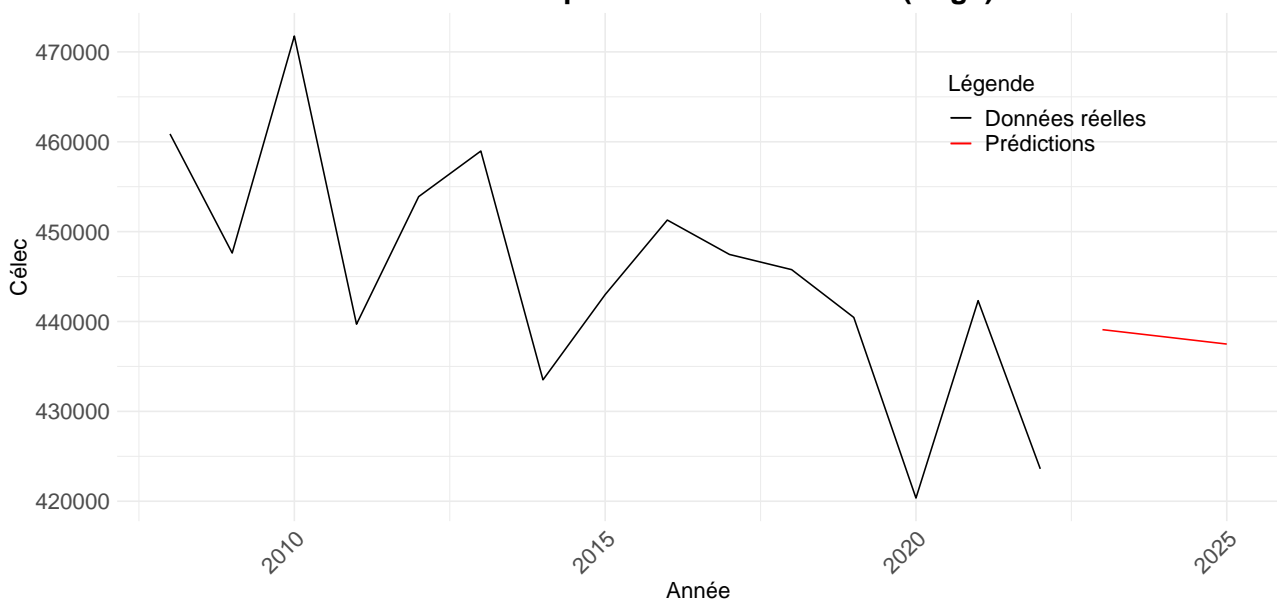
TABLEAU 16 – Prévisions pour les variables explicatives

Date	log_Pelec_IPC	log_DJU	Pop
2 023	0,6743199	-0,0366945	68 364 184
2 024	0,7047071	-0,0366945	68 645 383
2 025	0,7350943	-0,0366945	68 926 582

4.2 Prévisions du modèle

On peut alors faire des prévisions pour les années 2023, 2024 et 2025 :

Visualisation de Célec avec les prévisions du modèle 2 (ridge)



4.2.1 Commentaires

On peut observer que les prévisions semblent suivre la tendance des données réelles, avec une légère augmentation de la consommation électrique par rapport aux années post-covid (mais tout de même inférieure à la consommation électrique avant la crise sanitaire).

On précise également que la régression ridge ne permet pas de construire directement d'intervalle de confiance pour les prévisions. En effet, la régression ridge est une méthode de régression pénalisée qui permet de réduire la variance des estimateurs mais qui ne permet pas de construire d'intervalle de confiance. Une solution aurait consisté à utiliser la méthode de bootstrap pour construire des intervalles de confiance mais on n'est pas parvenu à obtenir des résultats satisfaisants. En particulier, la méthode bootstrap risquait de surestimer les incertitudes en propageant celles faites sur les prévisions des variables explicatives. On a donc préféré ne pas utiliser cette méthode.

5 Conclusion

En conclusion, nous voyons que le modèle final que nous avons choisi d'analyser est très simple. Nous expliquons le prix de l'électricité par habitants français par deux variables. D'une part son prix ajusté par la tendance inflationniste et de l'autre un indice climatique afin de prendre en compte la variation de température qui effectivement impactera sans aucun doute la consommation électrique française.

Nous avons pu observer que les résultats obtenus représentent bel et bien ceux attendus : une élasticité de la demande en électricité négative et une corrélation positive quant à l'indice de rigueur climatique. Ces deux résultats possèdent une significativité statistique très forte mais subissent également le manque de validité des hypothèses que le modèle doit vérifier afin de conclure formellement la corrélation entre notre variables dépendante et les variables explicatives.