

Modélisation et Prédiction des Prix de l'Électricité : Une Perspective de Données Fonctionnelles

BERRAHO Aya & MARGHT Abde-rrahman

Encadré par : Mme Angelina Roche

25 avril 2024

Table des matières

1	Introduction	3
2	Revue de littérature	4
2.1	Aperçu des modèles classiques de prévision des prix de l'électricité	4
2.2	Introduction aux données fonctionnelles	4
3	Présentation des données	5
4	Cadre théorique et modélisation du problème	7
5	Estimation des fonctions de demande de prix journalières	8
5.1	Régression linéaire	9
5.2	Lissage par splines cubiques	12
6	Estimation d'une base fonctionnelle	15
7	Résultats de convergence	19
8	Prévision des prix par le modèle FFM	22
	Références	25
	Annexe	26

1 Introduction

La modélisation et la prévision des prix de l'électricité sont des enjeux cruciaux qui touchent l'économie, la politique énergétique et la gestion des ressources. Cette importance est renforcée par la volatilité accrue des prix due à la transition vers des sources d'énergie renouvelables et à la libéralisation des marchés de l'énergie. Ces dynamiques complexifient les prévisions, rendant les modèles traditionnels moins efficaces et mettant en lumière le besoin d'approches novatrices telles que l'analyse de données fonctionnelles. Cette méthode offre un nouveau cadre pour une compréhension améliorée et une anticipation plus précise des variations de prix, en traitant les données de prix et de demande comme des fonctions continues.

Les modèles classiques, tels que les modèles autoregressifs et à changement de régime de Markov, ont historiquement dominé l'analyse des prix spot de l'électricité. Cependant, ils échouent souvent à capturer les fluctuations extrêmes, exacerbées par des facteurs exogènes comme les aléas météorologiques ou les changements abrupts dans la demande d'électricité. Ces défis sont accentués par les propriétés uniques du marché de l'électricité, notamment l'incapacité à stocker efficacement cette énergie et la nécessité d'un équilibre instantané entre l'offre et la demande.

La stabilité économique et la planification financière sont directement impactées par les fluctuations des prix de l'électricité. Des prévisions précises permettent aux consommateurs et entreprises de mieux gérer les impacts de ces fluctuations sur leurs coûts, renforçant ainsi la planification financière et la stabilité économique. Par ailleurs, en anticipant les variations des prix, les gestionnaires de réseau peuvent optimiser l'utilisation des diverses sources d'énergie. Ceci est crucial pour maximiser l'efficacité des ressources renouvelables et minimiser la dépendance aux combustibles fossiles, contribuant ainsi à une meilleure sécurité énergétique. En effet, une prévision efficace des prix contribue à garantir une alimentation énergétique stable et fiable, évitant les pénuries et gérant efficacement les pics de demande.

De plus, une meilleure compréhension des tendances des prix stimule les investissements dans les nouvelles technologies énergétiques et l'innovation, favorisant les progrès en matière d'efficacité énergétique et de durabilité. Les prévisions jouent également un rôle crucial dans la régulation et la politique énergétique, aidant les régulateurs à ajuster les politiques et les tarifications pour promouvoir une concurrence équitable, encourager les investissements stratégiques et protéger les consommateurs.

Ce mémoire propose de développer un modèle prédictif des prix spot de l'électricité à partir d'une analyse de données fonctionnelles. Le but est d'explorer les propriétés statistiques des séries de prix, d'implémenter un modèle capable de capturer ces dynamiques, et d'évaluer sa performance prédictive. Dans un premier temps, nous présenterons le modèle en définissant les variables et les relations fonctionnelles sous-jacentes. Nous partirons des données discrètes disponibles pour estimer des fonctions prix-demande continues, permettant une compréhension plus fine des dynamiques de marché. Ensuite, pour obtenir les estimations des fonctions prix-demande selon le modèle fonctionnel, nous procéderons à l'estimation d'une base fonctionnelle grâce à l'Analyse en Composantes Principales fonctionnelle. Ce travail préparatoire sera utilisé pour réaliser des prédictions de prix, qui seront comparées aux valeurs réelles pour évaluer la performance du modèle. L'ensemble de cette démarche permettra non seulement de tester l'efficacité de notre approche modélisatrice mais aussi de fournir des outils analytiques pertinents pour les décideurs dans le secteur énergétique.

2 Revue de littérature

2.1 Aperçu des modèles classiques de prévision des prix de l'électricité

Il existe aujourd'hui plusieurs modèles classiques de séries temporelles qui permettent de modéliser et prévoir les variations des prix de l'électricité.

En Europe, les opérateurs de réseau et les fournisseurs d'électricité tel que EDF, en France, et RWE, en Allemagne, utilisent des modèles ARIMA (Auto-Regressive Integrated Moving Average) pour la gestion quotidienne et la planification à court terme de l'approvisionnement en électricité. Ces modèles combinent les approches des modèles autoregressifs (AR) qui utilisent les valeurs des prix observés lors des périodes précédentes, et celles des modèles de moyenne mobile (MA) qui modélisent les prix comme une moyenne des perturbations passées prenant en compte les bruits et les chocs aléatoires. Les prix de l'électricité subissent une forte saisonnalité ce qui impacte considérablement la fiabilité des résultats donnés par le modèle ARIMA. De plus, les prix de l'électricité sont souvent des séries non-stationnaires, leurs statistiques descriptives notamment la moyenne et la variance ne sont pas constantes dans le temps, ce qui n'est pas vraiment compatible avec les modèles ARIMA qui nécessitent que les données soient stationnaires. On doit alors effectuer un pré-traitement des données ce qui rend la tâche relativement complexe.

Les traders d'énergie et les gestionnaires de risque utilisent les modèles de changement de régime de Markov pour capturer les variations brusques des prix dues à des événements inattendus ou des changements de politique énergétique. Ces modèles utilisent une chaîne de Markov pour contrôler les transitions entre différents états. En effet, dans notre cas pour modéliser les prix on peut diviser la série temporelle en plusieurs phases ou régimes, un saut de la chaîne de Markov peut être considéré comme un changement vers une autre phase. Ces modèles comportent également plusieurs désavantages, en effet il peut être difficile de déterminer les bons paramètres pour les matrices de transition et pour les processus sous-jacents de chaque phase. De plus ces modèles supposent que les transitions de régime dépendent uniquement de l'état actuel, ce qui peut ne pas toujours être la réalité des marchés d'électricité, sans oublier qu'identifier correctement quand un régime change réellement d'état peut être difficile en pratique.

2.2 Introduction aux données fonctionnelles

Les données fonctionnelles offrent une perspective riche et nuancée pour modéliser et prévoir les prix spot de l'électricité, un domaine confronté à des défis uniques en raison de la nature non stockable de l'électricité et de l'impact direct de la demande sur les prix. Traditionnellement, les modèles de séries temporelles et les modèles de changement de régime de Markov ont dominé ce champ d'étude. Toutefois, ces modèles peinent à saisir pleinement la complexité et la dynamique inhérentes aux données des prix de l'électricité. En réponse, la perspective des données fonctionnelles émerge comme une approche puissante, traitant les prix spot non pas comme des points isolés, mais comme des fonctions continues dépendantes du temps et d'autres variables explicatives comme la demande.

Cette méthodologie, illustrée par le modèle à facteurs fonctionnels (Functional Factor Model,

FFM), envisage chaque série de prix comme une fonction continue de la demande, capturant ainsi de manière plus naturelle les fluctuations des prix en fonction des variations de la demande. Par rapport aux modèles de série temporelle et de changement de régime de Markov, l'approche fonctionnelle se distingue par sa capacité à intégrer et à analyser les grandes dimensions de données temporelles, offrant une flexibilité et une précision souvent hors de portée pour les modèles traditionnels et ceux à changement de régime.

3 Présentation des données

Pour l'analyse menée dans ce mémoire, nous travaillons sur le marché allemand de l'électricité, le plus grand en terme de consommation en Europe, pour la période du 1er janvier 2006 au 31 décembre 2008. Nous avons récupéré les données détaillant la demande brute, l'apport éolien d'énergie, et les prix de l'électricité sur une base horaire et quotidienne. Ainsi nous calculons la demande nette en soustrayant l'apport éolien à la demande brute. En effet, comme dans la plupart des marchés électriques mondiaux, en Allemagne les sources d'énergie renouvelables bénéficient généralement de garanties d'achat. Par conséquent, les valeurs horaires de la demande brute d'électricité ne sont pas pertinentes, on soustrait alors les apports horaires d'électricité provenant des sources d'énergie renouvelables. Nous considérons uniquement les données d'apport d'énergie éolienne car les influences d'autres sources d'énergie renouvelables sur les prix restent négligeables pour le marché étudié.

Une étape importante du nettoyage des données a été l'élimination des jours non ouvrables de l'ensemble des données analysées. Cela inclut les week-ends, les vacances, et les jours fériés. Au total, 79 jours ont été retirés de l'analyse pour se concentrer uniquement sur les données des jours ouvrables, offrant ainsi une perspective plus précise des dynamiques de marché durant les périodes d'activité régulière.

Les outliers, ou valeurs aberrantes, sont des observations qui s'écartent significativement de la tendance générale des données. Dans notre contexte, nous considérons comme outliers les heures où le prix de l'électricité dépasse les 200 EUR/MWh. Ce seuil a été déterminé en identifiant les prix qui ne sont pas seulement atypiques mais aussi ceux qui pourraient indiquer des conditions anormales du marché, telles que des pénuries extrêmes, des erreurs de saisie de données, ou des événements non anticipés affectant l'offre ou la demande. Ces outliers représentent 0.51% des données et sont exclus de l'analyse pour éviter de fausser les modèles prédictifs et les conclusions de l'étude.

Après avoir appliqué les étapes de nettoyage mentionnées, il reste les données correspondant aux N_t heures de 638 jours ouvrables, où pour tout t entre 1 et $T = 638$, $N_t \leq 24$ désigne le nombre d'heures restant au jour t . Ce nettoyage permet de s'assurer que l'analyse se base sur des données représentatives des conditions normales de marché et non influencées par des facteurs exceptionnels ou non représentatifs. Ces données nettoyées forment la base sur laquelle les analyses subséquentes seront construites, visant à offrir des perspectives précises et fiables sur les dynamiques des prix de l'électricité dans un environnement de marché typique.

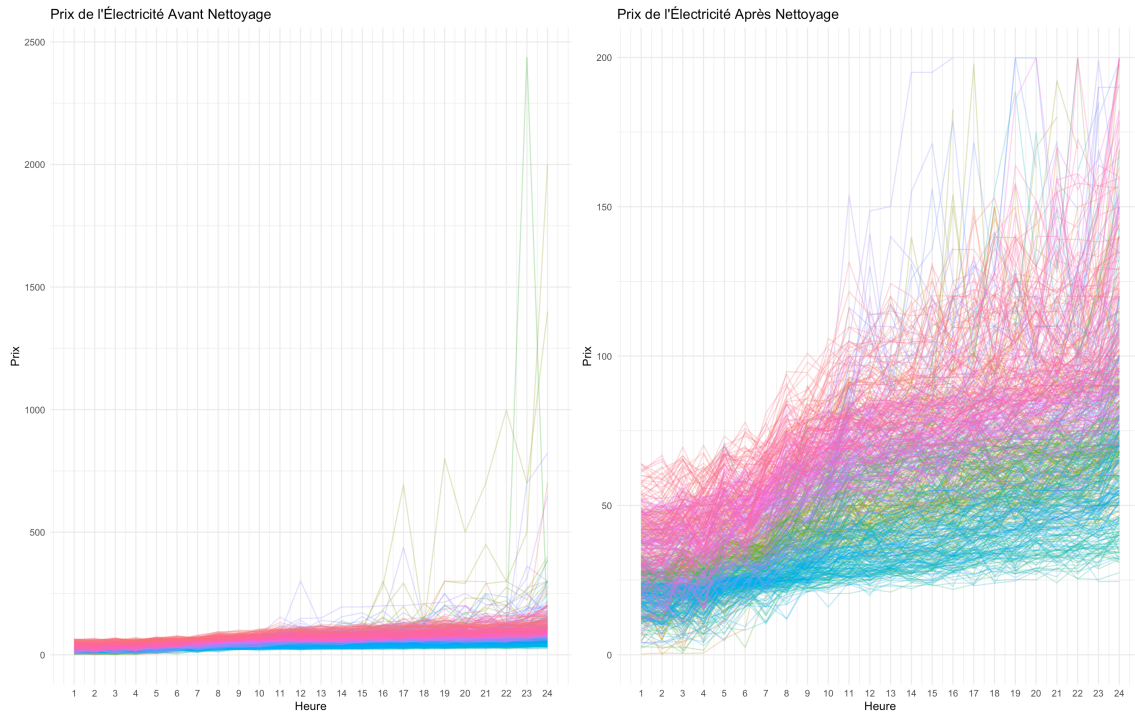


FIGURE 1 – Prix de l’électricité avant et après nettoyage des données (chaque jour est représenté par une courbe de couleur différente et les heures sont triés dans l’ordre croissant de la demande).

Sur le graphique de gauche nous avons affiché les prix de l’électricité en fonction des heures de la journée triées dans l’ordre croissant par rapport à la demande horaire, avant nettoyage des données. sur ce dernier, on n’arrive pas à visualiser l’évolution des prix par rapport à la demande car la présence de valeurs aberrantes trop élevées, rend l’échelle de l’axe verticale trop petite. Sur le graphique de droite, après avoir supprimé les outliers, on a une vision plus claire des données nous permettant de voir que le prix semble fonction croissante de la demande.

4 Cadre théorique et modélisation du problème

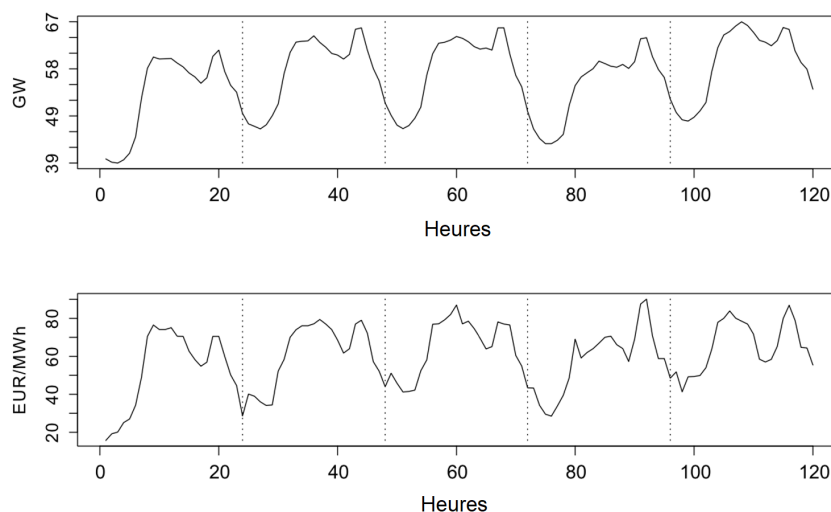


FIGURE 2 – Séries temporelles de la demande (graphique supérieur) et des prix (graphique inférieur) de l'électricité sur une période de 5 jours (3-7 mars 2008)

La figure 2 est issue du document de recherche de Liebl (2013). Elle représente les séries temporelles de la demande d'électricité et des prix spot de l'électricité sur une période de cinq jours ouvrables, du 3 au 7 mars 2008.

Le premier représente la demande d'électricité mesurée en GigaWatts (GW) en fonction des heures sur les 5 jours, on observe des pics clairs pendant les heures de la journée, qui correspondent probablement aux heures de travail et de consommation industrielle élevée, tandis que la demande baisse pendant les heures creuses, telles que la nuit.

De même le deuxième graphique représente les prix spot de l'électricité mesurés en euros par MegaWatheure (EUR/MWh), ceux-ci fluctuent de manière significative sur la période observée, avec des prix généralement plus élevés pendant les périodes de forte demande et des prix plus bas pendant les heures creuses.

En examinant les deux graphiques, on peut observer une corrélation évidente entre la demande d'électricité et les prix spot. Les périodes de forte demande semblent coïncider avec des prix plus élevés, tandis que les périodes de faible demande sont associées à des prix plus bas. Cette corrélation suggère qu'une fonction pourrait être déterminée pour estimer le prix spot de l'électricité à partir de la demande observée.

Ainsi, nous arrivons à notre modèle de relation fonctionnelle entre les prix et la demande d'électricité :

Définition 4.1. *Modèle de relation fonctionnelle entre les prix et la demande d'électricité*

Pour tout $t \in \{1, \dots, T\}$ et pour tout $h \in \{1, \dots, N_t\}$, la relation fonctionnelle est donnée par :

$$y_{t_h} = X_t(u_{t_h}) + \epsilon_{t_h}$$

où

- y_{t_h} représente le prix de l'électricité à l'heure h du jour t ,
- u_{t_h} représente la demande d'électricité à l'heure h du jour t ,
- $\{\epsilon_{t_h}\}$ représente un bruit blanc indépendant et identiquement distribué avec $\text{Var}(\epsilon_{t_h}) = \sigma_\epsilon^2$,
- X_t est la fonction prix-demande du jour t modélisée par un modèle FFM défini ci-dessous.

Définition 4.2. *Modèle à facteur fonctionnel (Functional Factor Model, FFM)*

Pour tout $t \in \{1, \dots, T\}$, la fonction $X_t(u)$ est donnée par :

$$X_t(u) = \sum_{k=1}^K \beta_{tk} f_k(u)$$

où

- $K \in \mathbb{N}$ le nombre de fonctions de base utilisées dans le modèle.
- $\{\beta_{tk}\}_t$ sont des séries temporelles (pouvant être non stationnaires) correspondant à des coefficients pour le jour t associées à la base fonctionnelle.
- f_k sont des fonctions de base constantes dans le temps.

X_t est modélisée comme une combinaison linéaire de fonctions de base $f_k(u)$ pondérées par des coefficients β_{tk} , qui varient dans le temps. L'équation d'observation intègre directement cette modélisation en plaçant $X_t(u_{t_h})$ dans un contexte réel où les prix observés y_{t_h} sont expliqués par la fonction prix-demande estimée à la demande observée u_{t_h} , plus un terme d'erreur.

Remarque 4.3. Une difficulté supplémentaire est que pour tout $t \in \{1, \dots, T\}$, la fonction X_t est définie sur un intervalle aléatoire $D(X_t) = [a_t, b_t]$, où pour chaque u_{t_h} les valeurs de la demande sont observées. L'intervalle $[a_t, b_t]$ est inclus dans $[A, B] \subset \mathbb{R}$, où (a_t) et (b_t) sont des séries temporelles telles que $A \leq a_t < b_t \leq B$. Les densités marginales respectives sont telles que $f_a(z_a) > 0$ et $f_b(z_b) > 0$ pour tout $z_a, z_b \in [A, B]$.

Remarque 4.4. La fonction de demande-prix X_t peut être vue comme l'équivalent empirique de la courbe de mérite, estimée de manière non paramétrique à partir des N_t paires de données horaires prix-demande $(y_{t_1}, u_{t_1}), \dots, (y_{t_{N_t}}, u_{t_{N_t}})$.

5 Estimation des fonctions de demande de prix journalières

Notre étude se sépare en deux étapes principales la première est celle de l'estimation des fonctions prix-demande X_t pour chaque jour t de notre période à partir des couples d'observation $(y_{t_1}, u_{t_1}), \dots, (y_{t_{N_t}}, u_{t_{N_t}})$. On a réalisé cette estimation à l'aide de trois différentes méthodes dont on comparera l'efficacité. La première est une régression linéaire, la deuxième est une estimation à noyaux et la dernière est celle utilisée dans l'article de Liebl (2013), la méthode de lissage par splines.

5.1 Régression linéaire

Définition 5.1. Modèle Linéaire

Un modèle linéaire se définit par une équation de la forme :

$$Y = X\beta + \epsilon$$

où $Y \in \mathbb{R}^n$ est le vecteur des observations, $X \in \mathbb{R}^{n \times (p+1)}$ est la matrice de variables explicatives, et $\beta \in \mathbb{R}^p$ est le vecteur de coefficients à estimer. On suppose que la matrice X est injective (i.e., $\text{rg}(X) = p$) et que le vecteur de bruit $\epsilon \in \mathbb{R}^n$ est un n -échantillon centre ($E[\epsilon] = 0$ et $\text{Var}(\epsilon) = \sigma^2 I$).

Remarque 5.2. En pratique, on a y_1, \dots, y_n , n observations indépendantes d'une variable quantitative. Pour chaque observation i , on dispose de p variables explicatives réelles (x_{1i}, \dots, x_{pi}). On cherche à "expliquer" y_i comme une fonction linéaire de ces variables. Ainsi, pour tout i , on écrit

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} + e_i, \quad \forall i = 1, \dots, n$$

où e_i est le terme d'erreur.

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_1^1 & \dots & x_1^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_i^1 & \dots & x_i^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n^1 & \dots & x_n^p \end{pmatrix}, \quad e = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}, \quad \text{et} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

$\underbrace{\hspace{10em}}_{r=p+1 \text{ colonnes}}$

Nous définissons également les notions suivantes :

\hat{y}_i est la valeur prédite par le modèle pour la i -ème observation, donnée par

$\hat{y}_i = X_i \beta$, où X_i est le vecteur ligne i -ème de la matrice X , et

\bar{y} est la moyenne des observations, définie par $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$.

Définition 5.3. Erreur standard des résidus, RSE

L'erreur standard des résidus est la mesure de la quantité de variation non expliquée par le modèle de régression, définie par :

$$RSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - p}}$$

où y_i est la i -ème valeur observée, \hat{y}_i est la i -ème valeur prédite par le modèle, n est le nombre total d'observations et p est le nombre de paramètres estimés dans le modèle (y compris l'ordonnée à l'origine).

Définition 5.4. Somme Totale des carrés, SCT

La SCT est la mesure de la variation totale présente dans les données de la variable dépendante. Elle représente la somme des carrés des écarts entre chaque observation et la moyenne de toutes les observations, et est définie par :

$$SCT = \sum_{i=1}^n (y_i - \bar{y})^2$$

où y_i est la valeur de la i -ème observation et \bar{y} est la moyenne de toutes les observations.

Définition 5.5. Somme des carrés des résidus, SCR

La SCR (RSS en anglais) est la somme des carrés des écarts entre les valeurs observées et les valeurs prédites par le modèle de régression. Elle mesure la quantité totale de variation non expliquée par le modèle et est définie par :

$$SCR = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

où y_i représente la valeur observée et \hat{y}_i la valeur prédite par le modèle pour la i -ème observation.

Définition 5.6. Coefficient de Détermination, R^2

Le Coefficient de Détermination, note R^2 et compris entre 0 et 1, exprime la proportion de la variance de la variable dépendante qui est expliquée par le modèle de régression. Il est défini par :

$$R^2 = 1 - \frac{SCR}{SCT}$$

où SCR est la Somme des carrés des Résidus et SCT est la Somme Totale des carrés.

Remarque 5.7. Les aspects suivants du coefficient de Détermination R^2 sont à considérer :

- Un R^2 de 0 indique que le modèle ne prévoit pas mieux la variable dépendante que ne le ferait un modèle simple basé sur la moyenne des données observées.
- À l'inverse, un R^2 de 1 signifie que le modèle prédit parfaitement la variable dépendante sans aucune erreur résiduelle.

Bien qu'un R^2 élevé suggère que le modèle capte une proportion significative de la variance de la variable dépendante, indiquant potentiellement un bon ajustement du modèle aux données, il ne garantit pas que les prédictions soient exemptes de biais, ni que le modèle soit le plus approprié, ou que des conclusions causales puissent être valablement tirées.

Implémentation R

```
# Vecteur des erreurs standard des résidus pour chaque régression linéaire de X_t
rse <- c()
rsquared <- c()
for(t in 1:T){
  Y.tmp <- as.vector(Y.mat[,t])
  U.tmp <- as.vector(U.mat[,t])
```

```

# Ajuster un modèle linéaire
model <- lm(Y.tmp ~ U.tmp)
rse <- c(rse, summary(model)$sigma)
rsquared <- c(rsquared, summary(model)$r.squared)
}
# Moyenne des erreurs standard des résidus pour chaque régression linéaire de X_t
print(mean(rse))
print(mean(rsquared))

```

Résultats

```

[1] 7.236453 # Moyenne des erreurs résiduelles standard
[2] 0.856932 # Moyenne des R^2

```

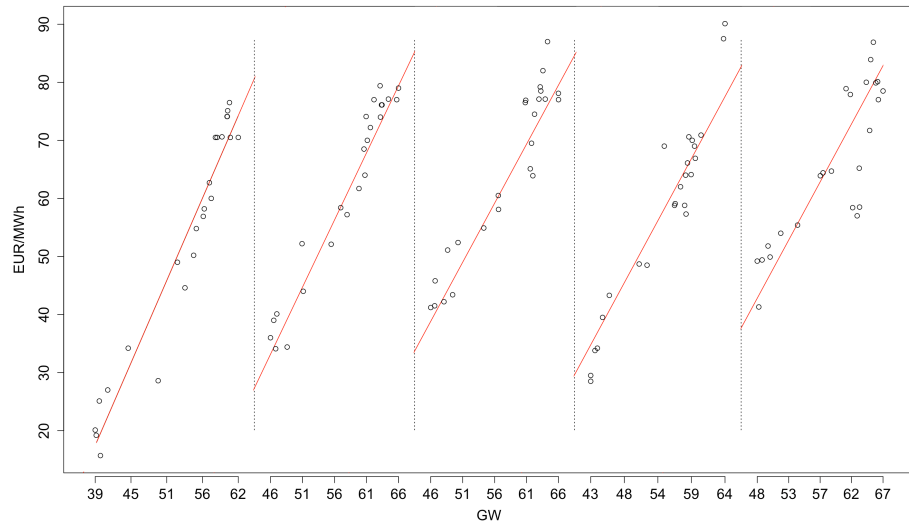


FIGURE 3 – Estimation des fonctions prix-demande par régression linéaire (3-7 Mars 2008)

A partir des observations des couples prix-demande de l'électricité (y_{t_h}, u_{t_h}) , nous avons ajusté un modèle linéaire défini par :

$$y_{t_h} = \beta_0 + \beta_1 u_{t_h} + \epsilon_{t_h}$$

où y_{t_h} représente le prix de l'électricité à l'heure h du jour t , u_{t_h} la demande d'électricité à cette même heure, et ϵ_{t_h} le terme d'erreur, supposé normalement distribué avec une moyenne nulle.

Nous avons tracé la fonction de régression sur une période de cinq jours, du 3 au 7 mars 2008 (figure 3), afin de visualiser l'ajustement du modèle et d'examiner comment les variations quotidiennes de la demande affectent les prix de l'électricité. Cette visualisation aide à comprendre

la dynamique du marché et à évaluer la validité de notre modèle linéaire dans le contexte des données réelles.

5.2 Lissage par splines cubiques

La procédure de lissage par splines cubiques est une méthode de régression non paramétrique utilisée pour estimer des fonctions continues à partir de données bruitées. Les splines cubiques sont des polynômes par morceaux de degré trois définis sur des sous-intervalles du domaine de la variable indépendante. Ces polynômes sont assemblés de manière à ce que la fonction globale soit deux fois continûment différentiable, ce qui garantit une courbe lisse.

Définition 5.8. *Spline cubique*

Soient x_1, \dots, x_n et y_1, \dots, y_n des nombres réels. On appelle spline cubique associée à la famille (x_i, y_i) toute fonction S de classe C^2 , polynomiale de degré au plus 3 sur chaque intervalle $[x_i, x_{i+1}[$, et telle que $S(x_i) = y_i$ pour tout $i = 1, \dots, n$.

Dans notre contexte, chaque fonction prix-demande $X_t(u)$ est estimée à partir d'un ensemble de données de prix y_t et de demandes u_t correspondantes, en minimisant une somme de résidus au carré pondérée par un terme de pénalité. Cette pénalité est proportionnelle au carré de la seconde dérivée de la fonction estimée, intégrée sur l'ensemble du domaine. Mathématiquement, cela revient à résoudre le problème d'optimisation suivant :

$$\min_{X_t \in C^2([a_t, b_t])} \left\{ \sum_{h=1}^{N_t} (y_{t_h} - X_t(u_{t_h}))^2 + \lambda \int_{a_t}^{b_t} (X_t''(u))^2 du \right\}$$

où λ est un paramètre de régularisation qui contrôle l'importance relative de la pénalité, et $X_t''(u)$ est la dérivée seconde de la fonction X_t par rapport à u .

Le résultat de cette minimisation est une fonction lisse $X_t(u)$, deux fois continûment dérivable, qui s'ajuste bien aux données tout en restant suffisamment flexible pour capturer les tendances sous-jacentes sans surajuster le bruit inhérent aux observations.

Afin de résoudre ce problème de minimisation, on cherche d'abord à déterminer le paramètre de lissage λ . Habituellement, le paramètre λ_{opt} est choisi par validation croisée.

Définition 5.9. *Validation croisée*

Technique statistique utilisée pour évaluer et améliorer la performance des modèles prédictifs. Elle consiste à partitionner les données en sous-ensembles, puis à tester le modèle sur un sous-ensemble après l'avoir entraîné sur les autres sous-ensembles. Cette méthode permet de vérifier l'efficacité du modèle en minimisant les risques de surajustement.

Procédure de Validation croisée

La validation croisée est réalisée en plusieurs étapes détaillées ci-dessous :

1. **Partition des données** : divisez l'ensemble de données total en k sous-ensembles (ou "folds") de taille à peu près égale.
2. **Procédure de Validation croisée** :
 - (a) Pour chaque iteration i de 1 à k :

- i. Sélectionnez le i -ème sous-ensemble comme ensemble de test.
- ii. Utilisez les $k - 1$ autres sous-ensembles comme ensemble d'entraînement.
- iii. Entraînez le modèle sur l'ensemble d'entraînement.
- iv. Évaluez le modèle sur l'ensemble de test.
- v. Enregistrez la mesure de performance du modèle, souvent l'erreur, pour cette itération.

3. **Évaluation de la Performance** : calculez la performance moyenne du modèle sur les k itérations pour obtenir une estimation robuste de sa performance.

L'objectif ici est que le compromis entre biais et variance de l'estimation \hat{X}_t soit optimisé asymptotiquement par rapport au critère d'erreur quadratique intégrée moyenne (MISE).

Définition 5.10. *MISE (Mean Integrated Squared Error)*

Le MISE est un critère utilisé pour mesurer la qualité d'un estimateur en statistique. Il représente une mesure globale de l'erreur quadratique de l'estimateur \hat{f} sur tout le domaine, combinant à la fois le biais et la variance de l'estimateur, offrant ainsi un aperçu complet de sa précision et de sa stabilité. Mathématiquement, il est défini par :

$$MISE(\hat{f}) = \mathbb{E} \left[\int (\hat{f}(x) - f(x))^2 dx \right]$$

où \hat{f} est l'estimateur de la fonction f , $f(x)$ est la fonction vraie, et \mathbb{E} désigne l'espérance mathématique.

Selon l'article de recherche de Liebl (2013), il n'est pas nécessaire d'optimiser le MISE de chaque estimateur \hat{X}_t mais plutôt celui des estimateurs $\hat{f}_1, \dots, \hat{f}_K$ de la base f_1, \dots, f_K définie dans la définition (4.2). Dans ce cas, on n'utilisera pas le paramètre de lissage optimal λ_{opt} déterminé par validation croisée mais un paramètre de sous-lissage $\lambda_K < \lambda_{\text{opt}}$, solution du problème de minimisation suivant :

$$CV(\hat{\gamma}) = \min \left\{ \sum_{t=1}^T \left(y_t - \sum_{k=1}^K \hat{\gamma}_{tk} \hat{f}_k(u_t) \right)^2 \right\} \quad \text{sur } [0, +\infty]$$

où $\hat{\gamma}_{tk}$ est l'estimateur des moindres carrés de β_{tk} et $\hat{f}_{k,-t}$ est l'estimateur de f_k déterminé à partir des couples de données (y_{sh}, u_{sh}) avec $s \in \{1, \dots, t-1, t+1, \dots, T\}$.

Notations :

- $\tilde{X}_1, \dots, \tilde{X}_T$ sont les estimateurs de X_t en utilisant le paramètre de sous-lissage λ_K .
- $\hat{X}_1, \dots, \hat{X}_T$ sont les estimateurs de X_t utilisant le paramètre de lissage λ_{opt} .

Implémentation R

```
## estimation spline (avec lambda_opt)
X.smGCVSmth <- vector("list", T)
for (t in 1:T) {
  X.smGCVSmth[[t]] <- sm.spline(x = U.mat[, t][!is.na(U.mat[, t])],
```

```

        y = Y.mat[, t][!is.na(Y.mat[, t])],
        spar = spar.GCV)
}

## estimation spline (avec lambda_K)
X.smUSmth <- vector("list", T)
for (t in 1:T) {
  X.smUSmth[[t]] <- sm.spline(x = U.mat[, t][!is.na(U.mat[, t])],
                             y = Y.mat[, t][!is.na(Y.mat[, t])],
                             spar = CV.USmth.spar)
}

for (t in 1:T) {
  Y.tmp <- as.vector(Y.mat[, t])
  U.tmp <- as.vector(U.mat[, t])
  # Ajuster un modèle spline
  y <- as.vector(predict(X.smUSmth[[t]], U.tmp))

  residuals <- y - Y.tmp
  # Somme des Carrés Résiduels
  SCR <- sum(residuals^2)
  # Somme des Carrés Totaux
  SCT <- sum((Y.tmp - mean(Y.tmp))^2)
  # R-carré
  res_tmp <- mean(sd(residuals))
  rsquared_tmp := 1 - SCR / SCT

  rse_spline <- c(rse_spline, res_tmp)
  rsquared_spline <- c(rsquared_spline, rsquared_tmp)
}
print(mean(rse_spline))
print(mean(rsquared_spline))

```

Résultats

```

[1] 5.943197 # Moyenne des erreurs résiduelles standard
[2] 0.9417929 # Moyenne des R^2

```

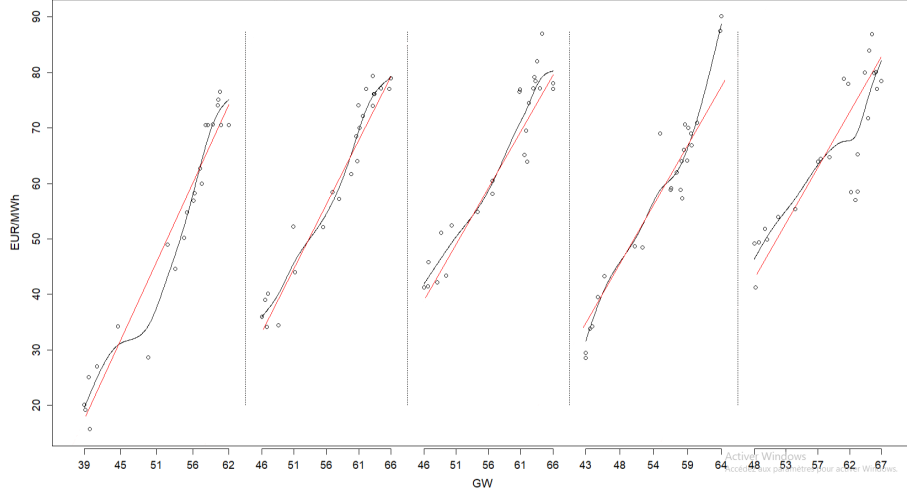


FIGURE 4 – Estimation des fonctions prix-demande par spline cubique (3-7 Mars 2008)

Nous avons ajouté, à la figure 3, les courbes des fonctions prix-demande estimées par la méthode spline. On remarque que cette technique capture mieux les tendances variées au sein d'un même jeu de données, qu'une simple régression linéaire. En effet, avec leur capacité à s'adapter localement aux données et à réguler le lissage, les splines permettent un ajustement plus naturel qui respecte la continuité et la différentiabilité de la fonction estimée.

Ces résultats graphiques se vérifient aussi par le calcul du coefficient de détermination R^2 . Ce dernier est plus élevé pour la méthode des splines, 0.94 contre 0.86 pour la régression linéaire. Ce qui indique potentiellement un meilleur ajustement du modèle aux données. De plus, la moyenne des erreurs résiduelles standard est plus petite pour la méthode des splines cubiques (5.94 contre 7.24) lors de la réévaluation des estimations de la fonction X_t aux points $(u_{t_h})_{h=1,\dots,N_t}$.

6 Estimation d'une base fonctionnelle

Proposition 6.1. *Il découle du modèle fonctionnel défini dans la définition (numéro), que toute base orthonormale $\{f_1, \dots, f_K\}$ de ce modèle doit résoudre le problème de minimisation suivant :*

$$\sum_{t=1}^T \left\| X_t - \sum_{k=1}^K \beta_{tk} f_k \right\|_2^2 = \min_{B_K} \sum_{t=1}^T \min_{\gamma_{t1}, \dots, \gamma_{tK} \in \mathbb{R}} \left\| X_t - \sum_{k=1}^K \gamma_{tk} g_k \right\|_2^2$$

où $B_K = \{g_1, \dots, g_K\}$ tel que $g_1, \dots, g_K \in L^2[A, B]$ et $\|\cdot\|_2$ désigne la norme fonctionnelle L^2 définie par $\|x\|_2 = \left(\int_A^B x^2(u) du \right)^{1/2}$, pour tout $x \in L^2[A, B]$.

Remarque 6.2. Comme vu précédemment, les séries temporelles $(\beta_{t1}), \dots, (\beta_{tK})$ ne sont pas stationnaires, ce qui implique que les fonctions X_t peuvent être d'ordres de grandeur très différents. Dans ce cas, certaines fonctions peuvent présenter des amplitudes très élevées comparées à d'autres, ce qui pourrait fausser les résultats de l'analyse, notamment en donnant un poids disproportionné aux fonctions de plus grande amplitude lors du calcul de la norme L^2 dans le problème ci-dessus.

Une solution à ce problème est de standardiser les fonctions X_t . En standardisant les fonctions, pour qu'elles aient toutes une norme unitaire, on s'assure que chaque fonction contribue également au processus d'analyse, permettant ainsi une comparaison plus juste et une interprétation plus précise des composantes principales ou des fonctions de base extraites.

On pose alors $X_t^* = X_t / \|X_t\|_2$ pour tout $t \in \{1, \dots, T\}$.

En remplaçant dans le problème de minimisation on obtient :

$$\sum_{t=1}^T \left\| X_t^* - \sum_{k=1}^K \beta_{tk}^* f_k^* \right\|_2^2 = \min_{B_K} \sum_{t=1}^T \min_{\gamma_{t1}, \dots, \gamma_{tK} \in \mathbb{R}} \left\| X_t^* - \sum_{k=1}^K \gamma_{tk} g_k \right\|_2^2.$$

Dans cette minimisation, on remplacerait les X_t par leurs estimateurs sous-lissés \tilde{X}_t déterminés dans la section 5.2, pour effectuer une ACP fonctionnelle et trouver une base fonctionnelle solution du problème.

Définition 6.3. *Analyse en composantes principales fonctionnelles (ACPf)*

L'analyse en composantes principales fonctionnelles est une méthode d'analyse statistique qui permet de résumer l'information contenue dans un ensemble de données fonctionnelles, c'est-à-dire des données qui varient en continu sur un intervalle, comme des courbes ou des surfaces. Elle permet de produire un espace de fonctions de plus petite dimension qui reproduisent un maximum d'information sur les données étudiées.

Étapes principales de l'ACPf :

1. Estimation de la fonction de covariance des données fonctionnelles.
2. Décomposition en valeurs propres de l'opérateur de covariance, définie grâce à la fonction de covariance.
3. Sélection des composantes : les premières fonctions propres (associées aux plus grandes valeurs propres) sont utilisées pour représenter les fonctions observées de manière concise.

Définition 6.4. *Opérateur de covariance*

Pour tout $u \in L^2[A, B]$

$$(\Gamma_T x)(u) = \int_A^B \gamma_T(u, v) x(v) dv \quad \text{pour tout } x \in L^2[A, B],$$

où $\gamma_T(u, v)$ est la fonction de covariance définie dans le cas général par :

$$\gamma_T(u, v) = T^{-1} \sum_{t=1}^T X_t(u) X_t(v).$$

Remarque 6.5. La définition classique de la fonction de covariance γ_T ne peut pas être utilisée ici car elle ne s'applique que pour un domaine déterministe. Étant donné que chaque fonction X_t est définie sur un domaine aléatoire $[a_t, b_t]$, on utilisera donc une version non paramétrique de γ_T .

Définition 6.6. Noyau en dimension 2

On appelle noyau toute fonction $K_2 : \mathbb{R}^2 \rightarrow \mathbb{R}$ intégrable telle que

$$\int_{\mathbb{R}^2} K_2(x, y) dx dy = 1$$

Exemple 6.7. Un exemple de noyau en 2 dimension est le noyau tricube défini par :

$$K(x, y) = \frac{4900}{6561} (1 - |x|^3)^3 (1 - |y|^3)^3 \text{ pour } \|(x, y)\| < 1.$$

Ce noyau est particulièrement apprécié pour ses propriétés de lissage en raison de sa décroissance rapide et de sa continuité jusqu'à la troisième dérivée.

Définition 6.8. Estimateur par polynômes locaux d'ordre ℓ en dimension 2

Considérons un ensemble de points $(x_i, y_i)_{i=1, \dots, n}$ avec leurs réponses correspondantes z_i pour chaque point. Soit un noyau positif K et une fenêtre $h > 0$.

On définit, pour tout $(x, y) \in \mathbb{R}^2$:

$$\hat{\theta}(x, y) = \arg \min_{\theta = (\theta_{00}, \dots, \theta_{\ell\ell}) \in \mathbb{R}^{(\ell+1) \times (\ell+1)}} \sum_{i=1}^n K\left(\frac{x_i - x}{h}, \frac{y_i - y}{h}\right) \left[z_i - \sum_{k=0}^{\ell} \sum_{j=0}^{\ell-k} \frac{\theta_{kj} (x_i - x)^k (y_i - y)^j}{k! j!} \right]^2$$

On pose $\hat{\theta} = (\hat{\theta}_{00}, \dots, \hat{\theta}_{\ell\ell})$.

L'estimateur par polynômes locaux d'ordre ℓ est alors défini par : $\hat{r}_n^\ell(x, y) = \hat{\theta}_{00}$

Dans notre contexte, on prend pour tout $(u, v) \in [A, B]$ $\gamma_T(u, v) = \hat{\theta}_{T,00}$ où $\hat{\theta}_{T,00}$ est déterminé en résolvant le problème suivant :

$$\min_{\theta_T = (\theta_{T,00}, \theta_{T,11}, \theta_{T,12}) \in \mathbb{R}^3} \sum_{t=1}^T \sum_{i,j=1}^{N_t} K_2\left(\frac{u_{ti} - u}{h}, \frac{u_{tj} - v}{h}\right) \times \{X_t^*(u_{ti})X_t^*(u_{tj}) - f(\theta_T, (u, v), (u_{ti}, u_{tj}))\}^2$$

où

- $(u_{ti})_{i=1, \dots, N_t}$ suite des valeurs observées de la demande en électricité pour tout $t \in \{1, \dots, T\}$
- h est le paramètre de lissage déterminé par validation croisée.
- $\kappa_2 : \mathbb{R}^2 \rightarrow \mathbb{R}$ est un noyau multiplicatif en dimension 2 telle que $\kappa_2(x_1, x_2) = \kappa(x_1)\kappa(x_2)$ avec κ un noyau univarié standard (noyau tricube défini en (6.7)).
- $f : \mathbb{R}^3 \times \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$ tel que $f(\theta_T, (u, v), (u_{ti}, u_{tj})) = \theta_{T,00} + \theta_{T,11}(u - u_{ti}) + \theta_{T,12}(v - u_{tj})$

On détermine $\hat{\gamma}_T$ en minimisant l'équation ci-dessus après avoir remplacé X_t^* par $\tilde{X}_t^* = \frac{\tilde{X}_t}{\|\tilde{X}_t\|^2}$. Par conséquent, on définit $\hat{\Gamma}_T$ par :

$$(\hat{\Gamma}_T x)(u) = \int_A^B \hat{\gamma}_T(u, v) x(v) dv \quad \text{pour tout } x \in L^2[A, B],$$

Ainsi, on désigne les premières K valeurs propres ordonnées de $\hat{\Gamma}_T$ par $\hat{\lambda}_{T,1} > \dots > \hat{\lambda}_{T,K}$ et les fonctions propres correspondantes par $\hat{e}_{T,1}, \dots, \hat{e}_{T,K}$.

Remarque 6.9. *En pratique, on ne réalise pas directement une ACP fonctionnelle, on utilise les fonctions \tilde{X}_t estimés dans la section 5.2 afin de recréer des valeurs discrètes et réaliser une ACP classique. Par ailleurs, on sélectionne 24 points $u \in [A, B]$, pour chaque \tilde{X}_t , on extrait les valeurs de la fonction aux 24 points de demande choisis, ce qui donne un vecteur de valeurs pour chaque t , qui est une représentation discrète des fonctions continues originales. Avec ces vecteurs, on réalise une ACP classique, ainsi, les vecteurs propres issus de cette ACP correspondent aux estimations ponctuelles des fonctions propres aux points de rediscrétisation. Pour obtenir des fonctions propres continues sur tout l'intervalle de demande, on réalise une interpolation de ces vecteurs propres en utilisant de nouveau la méthode des splines cubiques. Les fonctions propres obtenues servent de base pour représenter les fonctions prix-demande dans un espace de dimension réduite, capturant toujours l'essentiel de la variation des données.*

Valeurs propres	Taux d'inertie	Taux d'inertie cumulés
$\hat{\lambda}_{T,1}$	92.61	92.61
$\hat{\lambda}_{T,2}$	7.33	99.94
$\hat{\lambda}_{T,3}$	0.03	99.96

D'après le code en annexe de l'article de Liebl (2013), après la réalisation de l'ACP classique on conserve les trois plus grandes valeurs propres $\{\hat{\lambda}_{T,1}, \hat{\lambda}_{T,2}, \hat{\lambda}_{T,3}\}$ car la somme de leurs taux d'inertie donne 99,96%.

Après avoir déterminé les estimations des fonctions propres $\{\hat{e}_{T,1}, \dots, \hat{e}_{T,K}\}$, on estime les coefficients β_{tk} définie en (4.2) par la méthode des moindres carrés ordinaire.

Définition 6.10. *Estimateur des moindres carrés ordinaire*

On appelle estimateur des moindres carrés ordinaire, le minimiseur $\hat{\beta} \in \mathbb{R}^p$ du risque empirique :

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^p} \|Y - X\beta\|^2,$$

où $\|Y - X\beta\|^2$ définit la norme euclidienne au carré de l'erreur de prédiction.

Ainsi on défini :

— la base $\{\hat{f}_1, \dots, \hat{f}_K\}$ déterminé par ACP fonctionnelle tel que $\{\hat{f}_1, \dots, \hat{f}_K\} = \{\hat{e}_{T,1}, \dots, \hat{e}_{T,K}\}$

— l'estimateur des moindres carrés $\hat{\beta}_{tk}$ du coefficient β_{tk} pour tout $t \in \{1, \dots, T\}$ et $k \in \{1, \dots, K\}$

Finalement on peut estimer les fonctions prix-demande selon le modèle FFM défini en (4.2) tel que :

$$\hat{X}_t^f = \sum_{k=1}^K \hat{\beta}_{tk} \hat{f}_k \quad \text{pour tout } t \in \{1, \dots, T\}$$

Les estimations par splines cubiques, définies dans la section (5.2), \tilde{X}_t sont beaucoup moins efficaces que les estimateurs \hat{X}_t^f puisque ces derniers utilisent les informations de l'ensemble des données.

Définition 6.11. *Critère d'Information d'Akaike, AIC*

Le Critère d'Information d'Akaike (AIC) mesure la qualité d'un modèle statistique. Il est défini comme suit :

$$AIC = 2k - 2\ln(L),$$

où k est le nombre de paramètres dans le modèle et L le maximum de la fonction de vraisemblance du modèle. L'AIC pénalise l'ajout de paramètres supplémentaires au modèle, favorisant ainsi un modèle qui fournit un bon ajustement tout en restant simple.

K	AIC
1	34909.98
2	34361.82
3	34431.05

De nouveau, selon le code en annexe de Liebl (2013), on récupère l'AIC du modèle FFM pour $K = 1, 2$ et 3 . On conclut que le modèle minimisant l'AIC est le modèle fonctionnel à deux facteurs. Dans la suite la base qui nous permettra de réaliser nos prédictions

7 Résultats de convergence

Dans cette section on suppose que les fonctions X_t sont définies sur un domaine déterministe $[A, B]$. Dans ce cas, pour la réalisation de l'ACP fonctionnelle on utilisera directement l'estimation de opérateur de covariance $\hat{\Gamma}_T$ définie par :

$$(\hat{\Gamma}_T x)(u) = \int_A^B \hat{\gamma}_T(u, v) x(v) dv \quad \text{pour tout } x \in L^2[A, B],$$

où $\hat{\gamma}_T$ est la fonction de covariance empirique donnée par la définition (6.4), en remplaçant X_t par \tilde{X}_t , estimation réalisée dans la section 5.2 :

$$\hat{\gamma}_T(u, v) = \frac{1}{T} \sum_{t=1}^T \tilde{X}_t(u) \tilde{X}_t(v).$$

Pour le résultat suivant on posera l'hypothèse que :

$$\hat{\gamma}_T \xrightarrow{L^2} \gamma_T \text{ lorsque } N_t \rightarrow +\infty$$

Proposition 7.1.

Soit $T, K \in \mathbb{N}$ tel que $T \geq K$.

Soient $e_{T,1}, \dots, e_{T,K}$ les K fonctions propres associées aux K plus grandes valeurs propres ordonnées $\lambda_{T,1} > \dots > \lambda_{T,K}$ de l'opérateur de covariance Γ_T , défini en (6.4), supposé compact, et $\hat{e}_{T,1}, \dots, \hat{e}_{T,K}$ les K fonctions propres associées aux K plus grandes valeurs propres ordonnées $\hat{\lambda}_{T,1} > \dots > \hat{\lambda}_{T,K}$ de l'estimateur de l'opérateur de covariance $\hat{\Gamma}_T$ défini ci-dessus. $\{\hat{X}_1^f, \dots, \hat{X}_T^f\}$ sont les estimations des fonctions prix-demande réalisées par FFM définies à la fin de section 6.

On a, pour tout $t \in \{1, \dots, T\}$,

$$\hat{X}_t^f \xrightarrow[N_t \rightarrow +\infty]{L^2} X_t^f$$

Démonstration.

Dans un premier temps, le but est d'utiliser la convergence de l'estimation de l'opérateur de covariance au sens L^2 pour montrer la convergence des estimations des fonctions propres au sens L^2 .

Nous montrons que $\hat{\Gamma}_T \xrightarrow{L^2} \Gamma_T$ au sens L^2 . On a :

$$\begin{aligned} |||\hat{\Gamma}_T - \Gamma_T|||^2 &= \sup_{x \in L^2([A,B]), x \neq 0} \frac{||\hat{\Gamma}_T x - \Gamma_T x||^2}{||x||^2} \\ &= \sup_{x \in L^2([A,B]), x \neq 0} \frac{\int_A^B \left(\int_A^B (\hat{\gamma}_T - \gamma_T)(u, v) x(v) dv \right)^2 du}{||x||^2} \\ &\leq \sup_{x \in L^2([A,B]), x \neq 0} \frac{\int_A^B \left(\int_A^B (\hat{\gamma}_T - \gamma_T)^2(u, v) dv \right) \left(\int_A^B x^2(v) dv \right) du}{||x||^2} \\ &= ||\hat{\gamma}_T - \gamma_T||_2^2. \end{aligned}$$

La seconde inégalité s'obtient par l'inégalité de Cauchy-Schwarz.

Ainsi, la convergence de $\hat{\gamma}_T$ vers γ_T au sens L^2 implique la convergence de $\hat{\Gamma}_T$ vers Γ_T au sens L^2 .

À présent on souhaite montrer que $\hat{e}_{T,k} \xrightarrow[N_t \rightarrow +\infty]{} e_{T,k}$ au sens L^2 . Pour cela, on utilise l'inégalité de Bosq, citée en annexe. On a supposé que Γ_T est un opérateur compact, il reste à montrer qu'il est auto-adjoint et positif sur l'espace de Hilbert $L^2[A, B]$.

Soit $x \in L^2[A, B]$,

$$\begin{aligned}
\langle \Gamma_T x, x \rangle &= \int_A^B \Gamma_T x(u) x(u) du \\
&= \int_A^B \left(\int_A^B \gamma_T(u, v) x(v) dv \right) x(u) du \\
&= \int_A^B \int_A^B \frac{1}{T} \sum_{t=1}^T X_t(u) X_t(v) x(v) x(u) dv du \\
&= \frac{1}{T} \sum_{t=1}^T \langle X_t, x \rangle_2^2 \geq 0 \quad \text{comme somme de termes positifs.}
\end{aligned}$$

La troisième égalité s'obtient par le théorème du Fubini et la définition de γ_T .

De même,

soit $x, y \in L^2[A, B]$,

$$\begin{aligned}
\langle \Gamma_T x, y \rangle &= \int_A^B \Gamma_T x(u) y(u) du \\
&= \int_A^B \left(\int_A^B \gamma_T(u, v) x(v) dv \right) y(u) du \\
&= \int_A^B \int_A^B \gamma_T(u, v) x(v) dv y(u) du
\end{aligned}$$

Or, $\gamma_T(u, v) = \gamma_T(v, u)$ pour tout $u, v \in [A, B]$, donc :

$$\langle \Gamma_T x, y \rangle = \langle \Gamma_T y, x \rangle$$

La troisième égalité s'obtient par le théorème du Fubini.

Ainsi Γ_T est auto-adjoint et positif sur $L^2[A, B]$.

On a supposé que toutes les valeurs propres sont différents, ainsi, on a :

$$\|\hat{e}_{T,k} - e_{T,k}\|_2 \leq \frac{2\|\hat{\Gamma}_T - \Gamma_T\|_2}{|\lambda_i - \lambda_{i+1}|},$$

Puisque $\|\hat{\Gamma}_T - \Gamma_T\|_2 \xrightarrow{N_t \rightarrow +\infty} 0$, on peut en déduire que $\|\hat{e}_{T,k} - e_{T,k}\|_2 \xrightarrow{N_t \rightarrow +\infty} 0$

De plus on sait que l'estimateur des moindres carrés est consistant donc $\hat{\beta}_{tk} \xrightarrow{N_t \rightarrow +\infty} \beta_{tk}$ presque sûrement.

Les fonctions propres $\hat{e}_{T,k}$ sont déterminées à un changement de signe près. On pose $\hat{e}_{T,k} = \hat{f}_k$ et on suppose que les signes corrects de $\hat{e}_{T,k}$ sont choisis.

On a $\hat{X}_t^f = \sum_{k=1}^K \hat{\beta}_{tk} \hat{f}_k$, par ailleurs, on a montré que $\hat{\beta}_{tk} \rightarrow \beta_{tk}$ ps et que $\hat{e}_{T,k} \rightarrow e_{k,T}$ au sens L^2 .

On a donc $\hat{\beta}_{tk}\hat{f}_k \xrightarrow[N_t \rightarrow +\infty]{L^2} \beta_{tk}f_k$ et par continuité de la fonction somme on peut conclure que $\hat{X}_t^f \xrightarrow[N_t \rightarrow +\infty]{} X_t^f$ au sens L^2 .

□

Remarque 7.2. Si on souhaite obtenir ce résultat en ne supposant plus que le domaine de définition des X_t est déterministe mais aléatoire (comme dans le cas de notre étude) : $\mathcal{D}(X_t) = [a_t, b_t]$ $\forall t \in \{1, \dots, T\}$. Avec nos hypothèses faites en (4.3), on a que :

$$\forall \epsilon > 0, \mathbb{P}(a_t \in [A, A + \epsilon]) > 0 \text{ et } \mathbb{P}(b_t \in [B - \epsilon, B]) > 0$$

Donc il existe 2 séries temporelles $(a_s)_s$ et $(b_s)_s$ tels que A et B sont respectivement des points d'adhérence de $(a_s)_s$ et $(b_s)_s$. Lorsque $T \rightarrow +\infty$, on peut donc trouver plus de K fonctions X_t qui sont définies en tout point $u \in [A, B]$

8 Prédiction des prix par le modèle FFM

Dans la section 6, nous avons développé un modèle à facteurs fonctionnels (FFM) pour modéliser les prix de l'électricité, en intégrant des méthodes de statistiques et d'analyse fonctionnelle. Cette section se consacre à la prédiction des prix de l'électricité en utilisant le cadre méthodologique établi.

La prédiction des prix commence par la prédiction des coefficients utilisés dans le modèle FFM, celle-ci est faite à partir des estimations par la méthode des moindres carrés réalisée dans la section précédente. Cette étape utilise un modèle SARIMA (Seasonal AutoRegressive Integrated Moving Average) pour prédire les scores associés à chaque fonction de base pour les jours futurs. Ce modèle capte les structures saisonnières et autoregressives présentes dans les séries temporelles des coefficients, fournissant ainsi une base solide pour la prédiction.

Soit $S \in \{1, \dots, T\}$. Pour notre étude de prédiction, nous divisons l'ensemble de données en un échantillon d'apprentissage pour les jours $t \in \{1, \dots, S\}$ et un échantillon test pour les jours $t \in \{S + 1, \dots, T\}$. L'échantillon d'apprentissage est utilisé pour estimer les paramètres et l'échantillon test pour évaluer la performance de la prédiction. Par ailleurs, en parallèle à l'estimation des coefficients, la demande future d'électricité, notée $u_{S+\ell, h}$ pour tout $\ell \in \{1, \dots, 20\}$, est prédite en utilisant la méthode de persistance. Cette approche simple suppose que la valeur future sera égale à la dernière valeur observée, ce qui est une hypothèse courante pour les prévisions à très court terme dans les études de marchés de l'électricité, où les patterns quotidiens peuvent souvent se répéter.

Une fois les scores futurs et les estimations de la demande obtenus, ils sont combinés pour prédire les prix futurs de l'électricité. Cela se fait en utilisant la relation fonctionnelle établie par le modèle FFM :

$$\hat{X}_{S+\ell}^f(u_{S+\ell, h}) = \hat{\beta}_{S+\ell, 1}\hat{f}_1(u_{S+\ell, h}) + \hat{\beta}_{S+\ell, 2}\hat{f}_2(u_{S+\ell, h}),$$

où $\{\hat{f}_1, \hat{f}_2\}$ est la base déterminée grâce à l'ACP fonctionnelle de la section 6, $\hat{\beta}_{S+\ell, k}$ sont les coefficients futurs prédits pour chaque base fonctionnelle \hat{f}_k , pour tout $k \in \{1, 2\}$, et $u_{S+\ell, h}$ est la

demande prévue à l'heure h du jour $S + \ell$. Cette équation assemble les composantes pour donner une estimation du prix à l'heure h du jour $S + \ell$.

Implémentation R

```
T <- 717
#Separation de notre échantillon d'apprentissage et de test
S <- 711
#Nombre de jours à prédire
nb <- 1
#Pourcentage de notre échantillons de test
round((T-S)/T, digits = 3)*100

y <- NULL
y_estim <- NULL
#Choix de la méthode persistance
fin <- T-S-nb
for(i in 1:fin){
  ts.1 <- ts(ts.beta1[1:S])
  ts.2 <- ts(ts.beta2[1:S])
  model1 <- try(arima(ts.1, order=pdq1, seasonal=list(order=PDQ1, period=5), method="ML"))
  if(is.character(model1)){
    model1 <- arima(ts.1, order=pdq1, seasonal=list(order=PDQ1, period=5))
  }
  model2 <- try(arima(ts.2, order=pdq2, seasonal=list(order=PDQ2, period=5), method="ML"))
  if(is.character(model2)){
    model2 <- arima(ts.2, order=pdq2, seasonal=list(order=PDQ2, period=5))
  }
  ## "dernier_jour_observe": Dernier jour observe de l'échantillon d'apprentissage
  dernier_jour_observe <- S
  #Permet de trouver le dernier jour observe de l'échantillon d'apprentissage
  while(is.na(ord.mat.withNAs[1,dernier_jour_observe])){
    dernier_jour_observe <- dernier_jour_observe-1
  }
  #vecteur avec les vrais prix
  y <- c(y,
    if(is.na(Dates.withNAs[c(S+nb)])){
      rep(NA,24)
    }else{
      Y.mat.orig.withNAs[,c(S+nb)][!is.na(Y.mat.orig.withNAs[,c(S+nb)])]
    })
  #vecteur avec les estimations des prix
  y_estim <- c(y_estim,
    if(is.na(Dates.withNAs[c(S+nb)])){
      rep(NA,24)
    }else{
```

```

      c(c(predict(model1, nb.ahead=nb)$pred)[nb] *sgn1*
        c(predict(e.fun.hat.rot,
                  U.mat.orig.withNAs[,c(dernier_jour_observe)]
                  [!is.na(U.mat.orig.withNAs[,c(dernier_jour_observe)])])
        ),1]) +
      c(predict(model2, nb.ahead=nb)$pred)[nb] *sgn2*
      c(predict(e.fun.hat.rot,
                U.mat.orig.withNAs[,c(dernier_jour_observe)]
                [!is.na(U.mat.orig.withNAs[,c(dernier_jour_observe)])])
      ),2]))}

  S  <- S+1
}
#Création d'un data frame à partir des vecteurs y et y_estim
data <- data.frame(day = 1:length(y), Réel = y, Estimé = y_estim)

```

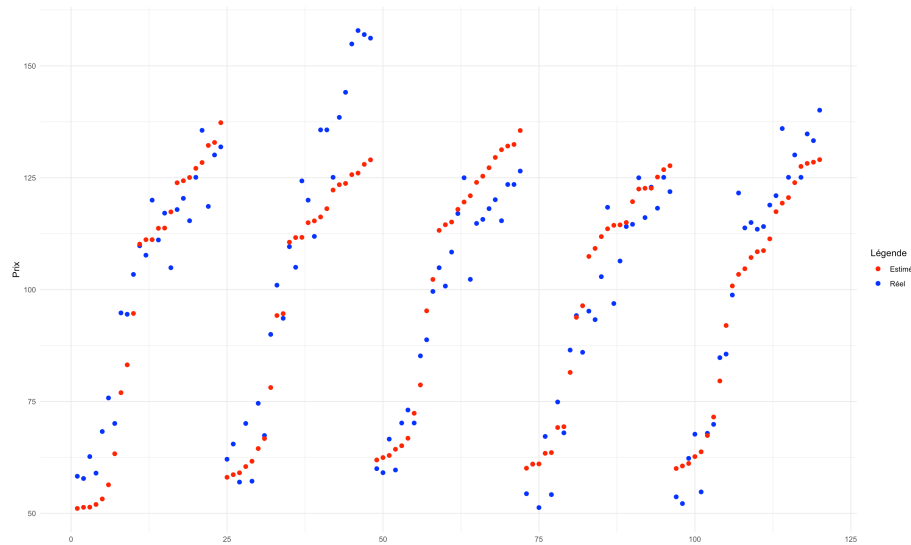


FIGURE 5 – Comparaison entre les prix prédis par FFM et leurs valeurs observés (3-7 Mars 2008)

Références

- [1] Dominik Liebl (2013) *Modeling and forecasting electricity spot prices: A functional data perspective*
- [2] Ekaterina Popovska, Galya Georgieva-Tsaneva (2022) *ARIMA Model for Day-Ahead Electricity Market Price Forecasting*
- [3] Michael Bierbrauer, Stefan Trueck, Rafał Weron(2004)
- [4] Frédéric Planchet (2023) *Méthodes de lissage et d'ajustement*
- [5] Carl de Boor (1978) *A Practical Guide to Spline*
- [6] Alexandre Tsybakov (2008) *Introduction to Nonparametric Estimation*
- [7] Ryad Belhakem, Franck Picard, Vincent Rivoirard, Angelina Roche (2021) *Minimax estimation of Functional Principal Components from noisy discretized functional data*
- [8] Fang Yao, Hans-Georg Müller, Jane-Ling Wang (2005) *Functional Data Analysis for Sparse Longitudinal Data*

Annexe

Norme d'opérateur :

Soient E et F deux espaces vectoriels normés respectivement munis des normes $\|\cdot\|_E$ et $\|\cdot\|_F$.
Soit f une application linéaire de E dans F . Considérons

$$N(f) := \sup_{v \neq 0} \frac{\|f(v)\|_F}{\|v\|_E}.$$

Si $N(f) < +\infty$, on dit que $N(f)$ est la norme de l'application f , subordonnée aux normes $\|\cdot\|_E$ et $\|\cdot\|_F$, on note $\|f\| = N(f)$.

Inégalité de Cauchy-Schwarz :

Soit $(E, \langle \cdot, \cdot \rangle)$ un espace vectoriel réel muni d'un produit scalaire. Alors, pour tous $x, y \in E$,

$$|\langle x, y \rangle| \leq \|x\| \|y\|$$

où $\langle \cdot, \cdot \rangle$ représente le produit scalaire dans E , et $\|x\|$ et $\|y\|$ sont les normes de x et y , respectivement, définies par $\|x\| = \sqrt{\langle x, x \rangle}$ et $\|y\| = \sqrt{\langle y, y \rangle}$.

Convergence au sens L^2 :

On dit que la suite (X_n) converge vers X au sens L^2 , ou encore en norme L^2 , si, pour tout n , X_n et X ont un moment d'ordre 2 fini et si

$$\lim_{n \rightarrow \infty} \|X_n - X\|_2 = 0$$

ou de manière équivalente, si

$$\lim_{n \rightarrow \infty} \mathbb{E} [|X_n - X|^2] = 0.$$

Dans ce cas, on note

$$X_n \xrightarrow{L^2} X.$$

Inégalité de Bosq :

Soient T un opérateur compact, auto-adjoint, et positif sur un espace de Hilbert H , et $\{\lambda_i\}$ les valeurs propres de T avec les fonctions propres correspondantes $\{\phi_i\}$. Supposons que \hat{T} soit une estimation empirique de T basée sur des données, avec les estimations des valeurs propres $\{\hat{\lambda}_i\}$ et des fonctions propres $\{\hat{\phi}_i\}$.

L'inégalité de Bosq est exprimée comme suit :

$$\|\hat{\phi}_i - \phi_i\| \leq \frac{2\|\hat{T} - T\|}{|\lambda_i - \lambda_{i+1}|},$$

où λ_{i+1} est la valeur propre suivante après λ_i dans l'ordre décroissant des valeurs propres.