

Projet de Recherche (PRE)

Spécialité : SIM/Finance
Année Scolaire : 2017-2018

Coût de la volatilité de la production intermittente sur les marchés de l'électricité

Mention de confidentialité

Le contenu de ce rapport est non confidentiel

**Auteur :
Samer Said**

**Promotion :
2019**

**Tuteur ENSTA ParisTech :
Francesco Russo**

**Tuteur organisme d'accueil :
René Aid et Yannick Le Pen**

Stage effectué du 14-05-2018 au 03-08-2018

Nom de l'organisme d'accueil : PSL Dauphine

Adresse : Place du Maréchal de Lattre de Tassigny, 75016 Paris, France

Mention de confidentialité

Ce document est non confidentiel. Il peut donc être communiqué à l'extérieur sous format papier mais également diffusé sous format électronique.

Remerciement

Je voudrais exprimer ma gratitude à mes deux tuteurs, René Aïd et Yannick Le Pen, pour m'avoir inclus dans leur équipe pour ce programme d'été et pour leur disponibilité à m'aider chaque fois que j'avais besoin d'aide ou de conseil dans mon projet.

Je suis aussi très chanceux d'avoir Monsieur Francesco Russo comme un enseignant référant, dont la porte était toujours ouverte et qui était toujours patient avec moi. Une reconnaissance particulière à Damien Fessler, pour m'avoir fait découvrir pratiquement tout à la Dauphine, des services informatiques à la bibliothèque de recherche, avoir parfois le déjeuner ensemble et m'expliquer patiemment les nouveautés du domaine. Enfin, je voudrais remercier toute l'équipe du bureau, surtout Miguel Ventura, d'accepter toujours de répondre à mes questions et de me donner des suggestions concernant mon travail. C'était un excellent laboratoire.

Resumé

Le projet de recherche porte sur l'estimation d'une partie des coûts induits par l'accroissement de la volatilité sur les marchés de court-terme de l'électricité du fait de l'introduction massive des énergies renouvelables intermittentes.

La question des coûts induits par ces évolutions dans les moyens de production a été abordée dans la littérature surtout à l'aide du modèle d'équilibre offre-demande. Récemment, des travaux ont utilisé des données de prix de marché infra-journalier pour estimer l'efficacité du marché infra-journalier dans la réduction des erreurs de prévisions [2]. Par ailleurs, un modèle de Meyn et al. 2010 met en relation la flexibilité des moyens de production et les coûts induits par la volatilité de la demande [4].

L'objet de ce mémoire consiste à développer des modèles économétriques pour modéliser l'erreur de prévision de la consommation nette définie après et pour estimer le coût de la volatilité de la production intermittente sur les marchés day-ahead de l'électricité.

Abstract

The research project focuses on estimating a portion of the costs of increasing volatility in the short-term electricity markets as a result of the massive introduction of intermittent renewables.

This problem of the costs induced by these evolutions in the means of production has been addressed in the literature mainly by the supply-demand equilibrium model. Recently, research has used intraday market price data to estimate the efficiency of the intra-day market in reducing forecast errors [2]. In addition, a model by Meyn et al. 2010 relates the flexibility of the means of production and the costs induced by the volatility of demand [4].

The purpose of this thesis is to develop econometric models to model the net consumption defined after and to estimate the cost of intermittent production volatility in the day-ahead electricity markets.

Table des matières

Mention de confidentialité	2
Remerciement	3
Resumé	4
Introduction	10
1 Préparation des données	12
1.1 Présentation des données	12
1.2 Sources de données	14
1.3 Traitement préliminaire des données	14
2 Analyse des données	16
2.1 Description globale de la consommation et de la production éolienne	16
2.2 Analyse des coûts	19
2.2.1 L'erreur de prévision de la consommation nette	19
2.2.2 Coût prévu de production	20
2.2.3 Coût d'ajustement de production	22
2.2.4 Variance de la demande résiduelle journalière	24
2.3 Explications des données aberrantes dans les coûts	29
3 Modélisation et estimation de la volatilité	30
3.1 Tests de racine unitaire	31
3.2 Le modèle ARMA	35
3.2.1 Spécification, estimation et validation du modèle	36
Spécification	36
Opérateur retard :	36
Le processus AR :	36
Le processus MA :	37
Estimation	39

Test de significativité des coefficients	39
Estimation	40
Test de la significativité des coefficients	40
3.2.2 Tests sur les résidus	41
Tests de Ljung-Box d'absence d'autocorrélation des résidus .	41
Test de l'hypothèse de normalité des résidus	42
Tests d'homoscédasticité	42
3.3 Le modèle ARMA-GARCH	44
3.3.1 Estimation du modèle	44
3.3.2 Tests de validation	46
3.3.3 Choix du modèle	48
3.4 Le coût de la volatilité	48
Conclusion	50
A Annexe A	52
B Annexe B	53

Table des figures

1.1	Extrait du fichier de la réalisation de la consommation totale au pas d'une demi-heure [1].	13
1.2	Extrait du fichier de prix spot day-ahead au pas d'une heure [5]. . .	13
1.3	Un code sur VBA *	15
2.1	Exemple de cycle annuel de consommation d'énergie électrique en 2011 [3].	17
2.2	Exemple de cycle hebdomadaire de consommation électrique en puissance (MW) [3].	17
2.3	Exemple d'une consommation électrique en été le 16/05/2015 en MWh *.	18
2.4	Exemple d'une consommation électrique en hiver le 22/02/2015 en MWh *.	18
2.5	L'évolution de l'erreur de prévision de la consommation nette au long des trois années (2015, 2016, 2017) *.	20
2.6	Exemple du coût prévu de production pour l'année 2015 *.	21
2.7	L'évolution du coût prévu de production au long des trois années (2015, 2016 et 2017) *.	21
2.8	Exemple du coût d'ajustement de production pour l'année 2015 *.	23
2.9	L'évolution du coût d'ajustement de production au long des trois années (2015, 2016 et 2017) *.	23
2.10	L'évolution de la variance de la demande résiduelle journalière au long des trois années (2015, 2016 et 2017) *.	24
2.11	Le coût d'ajustement de production en fonction de la variance de la demande résiduelle journalière en 2017 *.	25
2.12	Les paramètres descriptifs de la régression linéaire *.	25
2.13	Le coût d'ajustement de production en fonction de la variance de la demande résiduelle journalière pour les trois années (2015, 2016 et 2017) *.	26
2.14	Les paramètres descriptifs de la régression linéaire *.	26

2.15	Le coût d'ajustement de production en fonction de la variance de la demande résiduelle journalière (entre 18h et 21h) en 2017 *.	27
2.16	Les paramètres descriptifs de la régression linéaire *.	27
2.17	Le coût d'ajustement de production en fonction de la variance de la demande résiduelle journalière (entre 18h et 21h) pour les trois années (2015, 2016 et 2017) *.	28
2.18	Les paramètres descriptifs de la régression linéaire *.	28
3.1	L'évolution de l'erreur de prévision de la consommation nette au long des trois années (2015, 2016 et 2017) *.	31
3.2	Les paramètres descriptifs de la série X_h *.	31
3.3	Le test ADF pour la série de l'erreur de prévision de la consommation nette *.	33
3.4	Le test DF-GLS pour la série de l'erreur de prévision de la consommation nette *.	33
3.5	Le test M pour la série de l'erreur de prévision de la consommation nette *.	34
3.6	Le test KPSS pour la série de l'erreur de prévision de la consommation nette *.	35
3.7	La fonction d'autocorrélation pour l'erreur de prévision de la consommation nette *.	38
3.8	L'autocorrélation partielle pour l'erreur de prévision de la consommation nette *.	38
3.9	Les coefficients du modèle *.	39
3.10	Les probabilités critiques associées aux coefficients *.	39
3.11	Les coefficients du modèle *.	40
3.12	Les probabilités critiques associées aux coefficients *.	40
3.13	Test d'absence d'autocorrélation jusqu'à l'ordre 10 *.	41
3.14	Test d'absence d'autocorrélation jusqu'à l'ordre 10 *.	41
3.15	Test de la normalité des résidus *.	42
3.16	Cluster de volatilité *.	43
3.17	Test d'un effet ARCH pour les résidus *.	43
3.18	Les coefficients du modèle *.	45
3.19	Les coefficients du modèle *.	46
3.20	L'évolution de la variance conditionnelle *.	47
3.21	L'évolution de la variance conditionnelle *.	47
3.22	Les critères d'information pour ARMA(25,4)-GARCH(1,1) *.	48
3.23	Les critères d'information pour ARMA(25,4)-GARCH(1,2) *.	48
3.24	Le coût d'ajustement de production en fonction de la volatilité *.	49
3.25	Les paramètres descriptifs de la régression linéaire *.	49

B.1	Les coefficients du modèle *.	53
B.2	Les probabilités critiques associées aux coefficients *.	53

Introduction

L'énergie éolienne est la technologie de production d'électricité qui connaît la croissance la plus rapide parmi les sources d'énergie renouvelables, et son développement joue un rôle essentiel dans la transformation actuelle à l'échelle mondiale de l'énergie de combustibles fossiles à l'énergie propre grâce à ses bénéfices écologiques.

Cependant, l'évolution importante de la pénétration de l'énergie éolienne a forcé le système électrique jusqu'à aujourd'hui à confronter des nombreux défis avec l'intégration des énergies renouvelables et la stabilité de système résultant de la volatilité importante et la prévisibilité faible de l'énergie éolienne. Depuis la déréglementation des marchés de l'électricité en Europe, la conception du marché à court terme a progressivement divisé le trading d'électricité en plusieurs marchés séquentiels avec des horizons temporels différents. Contrairement à ce qui se passe dans de nombreux pays, les mécanismes d'échanges intra-journaliers (connu souvent sous le nom anglais "intraday") sont assez récents sur les marchés de l'électricité européens, notamment en France. Chronologiquement, les producteurs peuvent décider librement de vendre leur production un jour avant le jour de l'opération (sur le marché day-ahead¹ de l'électricité) et/ou de continuer à faire le trade près du moment de livraison physique, si le planning journalier doit être corrigé ou réinitialisé (sur les marchés intra-journaliers² de l'électricité). Comme la production éolienne dépend fortement des conditions météorologiques, qui sont sujettes à une prévisibilité limitée, le calendrier réel de l'énergie éolienne peut dévier de la production prévue sur les marchés day-ahead. Plus proche du temps réel que le trading dans les marchés day-ahead, le trading intra-journalier permet notamment aux producteurs de l'énergie éolienne (et aux autres acteurs du marché) de modifier leurs calendriers de la production journalière en fonction de prévisions actualisées et améliorées après la clôture de marché day-ahead.

1. Le marché où le négoce d'électricité se fait la veille pour le jour suivant, par exemple, l'EPEX Spot à Paris (Spot Market de l'European Power Exchange).

2. Le marché intra-journalier, Intraday en anglais, permet aux acteurs de la bourse de l'électricité de négocier les prix d'achat et de vente durant toute la journée.

Par conséquent, le trading infra-journalier est considéré comme un outil important pour gérer le vent intermittent, et pour favoriser son intégration dans le système électrique.

On a commencé premièrement par la préparation des données. La deuxième partie est consacrée à l'analyse descriptive des données à travers le calcul des quelques variables comme le coût prévu de production, le coût d'ajustement de production etc...

On effectue ensuite les analyses statistiques pour étudier la relation entre ces variables à l'aide des représentations graphiques. Une fois les données bien étudiées, on passe enfin à la partie consacrée à la modélisation et l'estimation afin d'aborder la question du coût de la volatilité de la production intermittente sur les marchés de l'électricité.

Chapitre 1

Préparation des données

1.1 Présentation des données

On dispose de six fichiers des données : Prévision de la consommation totale pour le lendemain, Prévision de la production éolienne pour le lendemain, Prix spot day-ahead¹, Réalisation de la consommation totale, Réalisation de la production éolienne, Prix du règlement des écarts et ces fichiers sont séparément pour les années 2015, 2016 et 2017. Il s'agit aussi de données pour la France seulement en mentionnant que la consommation considérée est la consommation des ménages et industries.

Ces fichiers contiennent des données journalières au pas d'une heure ou demi-heure pour les variables mentionnées précédemment. Les fichiers relatifs à la consommation totale (cf. Figure 1.1) et au prix du règlement des écarts fournissent des données journalières au pas d'une demi-heure de 00 :00 jusqu'à 23 :30, du 01/01 jusqu'au 31/12. Les fichiers relatifs à la production et au prix spot day-ahead (cf. Figure 1.2) fournissent des données journalières (h,t)² au pas d'une heure de 00 :00 jusqu'à 23 :00, de 01/01 jusqu'au 31/12. Ces fichiers sont pour les années 2015, 2016 et 2017, chaque année a ses propres fichiers. Par exemple, on dispose de 17520 observations pour le premier type de fichiers et 8760 observations pour le deuxième.

1. Le prix d'électricité fixé la veille pour le lendemain.

2. h est l'heure du jour t.

Date	00:00	00:30	01:00	01:30	02:00	02:30	03:00	03:30	04:00
01/01/2015	74493	73481	71506	71505	71134	70856	68840	67315	65749
02/01/2015	70442	68908	66435	66281	65837	65729	63815	62707	61691
03/01/2015	69454	67664	65104	64782	64075	63674	61537	60077	58853
04/01/2015	63198	61548	58984	58353	57594	57013	54873	53363	52125
05/01/2015	64953	63529	61408	61414	61073	61009	59186	58116	57225
06/01/2015	72782	70975	68356	68083	67642	67555	65740	64570	63639
07/01/2015	72638	70921	68256	67837	67502	67125	65230	64147	63028
08/01/2015	71684	69918	67199	66544	66055	65847	63811	62460	61433
09/01/2015	68532	66614	63983	63517	62851	62666	60747	59730	58739
10/01/2015	63512	61281	58132	57477	56517	55756	53621	52201	51057
11/01/2015	60709	58926	56390	55815	55087	54410	52362	50852	49692
12/01/2015	64211	62806	60336	60149	59846	59620	57864	56797	55848
13/01/2015	67741	65945	62874	62623	61993	61532	59943	58610	57642
14/01/2015	66424	64613	61720	61260	60826	60588	58849	57805	56867

FIGURE 1.1 – Extrait du fichier de la réalisation de la consommation totale au pas d'une demi-heure [1].

Delivery day	Hour1	Hour2	Hour3A	Hour3B	Hour4
01/01/2015	53,480	51,930	48,760		42,270
02/01/2015	38,800	36,730	33,520		27,500
03/01/2015	46,970	43,270	41,760		37,540
04/01/2015	36,260	32,280	27,630		23,940
05/01/2015	36,560	34,940	32,190		23,530
06/01/2015	45,560	40,280	39,200		36,220
07/01/2015	40,000	36,690	37,450		32,080
08/01/2015	43,880	35,650	35,870		30,170
09/01/2015	32,060	29,710	28,530		23,750
10/01/2015	19,040	15,960	13,700		11,010
11/01/2015	26,950	24,220	27,080		20,910
12/01/2015	27,590	26,810	25,750		17,860
13/01/2015	33,040	26,560	26,290		17,440
14/01/2015	34,580	29,670	27,490		21,300

FIGURE 1.2 – Extrait du fichier de prix spot day-ahead au pas d'une heure [5].

On remarque qu'il y a deux colonnes pour la troisième heure du jour dans le tableau précédent, une colonne sous le nom de l'heure 3A et l'autre 3B qui est totalement vide sauf pour une seule case qui correspond au jour de changement horaire d'hiver de l'année³.

3. La montre doit aller vers l'arrière d'une heure : à 3 heures, il est 2 heures. Le changement a lieu chaque année le dernier dimanche d'octobre.

1.2 Sources de données

Les données sont extraites de EPEX SPOT⁴ et RTE⁵. le prix spot day-ahead et le prix du règlement des écarts sont mesurés en euro par megawatt heure (€/MWh) en france. En plus, les autres données (les données de prévision pour le lendemain et celles de la réalisation) sont mesurés en megawatt heure (MWh). De l'EPEX SPOT⁶, on a obtenu seulement les prix spot day-ahead et le reste de données est pris de RTE⁷.

1.3 Traitement préliminaire des données

Pour bien commencer, on va décrire les fichiers précédents. Il s'agit comme mentionné avant de six fichiers pour trois années différentes de format excel. Il est important de noter qu'on a effectué nos études avec le langage de programmation et le logiciel dédié aux statistiques R. En effet, pour bien effectuer nos études sur ces fichiers avec R, ces derniers doivent être bien organisés de façon régulière dans le but d'une bonne extraction des données qu'on utilise après dans la mise en oeuvre des algorithmes (la récursivité, l'itération, le parcours de tableaux excel, etc...).

Par conséquent, on a été obligé d'étudier soigneusement les fichiers et de les modifier si nécessaire afin de se débarrasser de toutes les anomalies avant d'aborder la partie d'analyse des données, mais le problème c'est qu'on se dispose des milliers des données⁸. Certainement, il semble très difficile de le faire avec l'interface graphique d'excel seulement, car avec des milliers des données ça va prendre beaucoup de temps, en général ce n'est pas intelligent et on peut le prendre comme une mauvaise impression pour un ingénieur qui doit toujours optimiser les choses.

Ainsi, on s'est dirigé vers la programmation avec le VBA⁹ qui est une implémentation de Microsoft Visual Basic qui est intégrée dans toutes les applications

4. La bourse européenne de l'électricité EPEX SPOT SE est une bourse de l'électricité au comptant (spot) pour l'Allemagne, la France, la Grande Bretagne, les Pays-Bas, la Belgique, l'Autriche, la Suisse et le Luxembourg.

5. RTE, réseau de transport d'électricité, est une entreprise de service qui gère le réseau public de transport d'électricité haute tension en France métropolitaine.

6. <https://www.epexspot.com>

7. <https://data.rte-france.com>

8. Une donnée pour chaque heure ou chaque demi-heure de chaque jour de l'année.

9. Visual Basic for applications.

de Microsoft Office, en particulier le microsoft excel. Sous ce dernier, on a travaillé avec les macros-commandes qui sont une suite de commandes. De plus, l'enregistreur de macros sous Microsoft Excel (Onglet Développeur / Enregistrer une Macro) permet de créer facilement une procédure en code VBA. Toute la séquence d'action effectuée entre le début et la fin de l'enregistrement est enregistrée dans une procédure VBA, qui pourra être réexécutée à l'identique. Il est possible de modifier ce code ou de programmer directement dans la VBE (Onglet Développeur / Visual Basic).

L'utilisation de macro-commandes permet des gains de temps considérables par la faveur de l'automatisation de tâches complexes et souvent répétitives (par exemple : actualisation des données d'un tableau croisé dynamique, concaténation de plusieurs fichiers). Dans notre cas, on a utilisé les macros pour détecter les lignes aberrantes pour les supprimer et le plus important on doit détecter les données manquantes dans nos fichiers (exemple : on ne trouve pas les données du 21 mai au 25 mai, où il faut intervenir et ajouter des lignes vides pour prendre en considération ces données en les marquant comme des données manquantes (NA)).

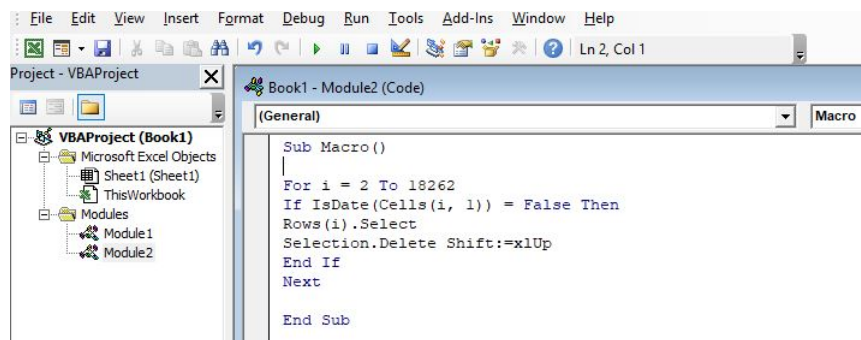


FIGURE 1.3 – Un code sur VBA *¹⁰.

La figure précédente montre un exemple de code écrit sur VBA servant à détecter en premier lieu les lignes aberrantes, puis les supprimer¹¹.

10. Cette figure est faite par le réalisateur de ce rapport.

11. La première colonne de nos fichiers est occupée par les dates, c'est pourquoi on vérifie si le contenu de chaque première cellule de chaque ligne est de format date ou non.

Chapitre 2

Analyse des données

On a utilisé des techniques de séries temporelles pour répondre aux questions de recherche soulevées dans la section d'introduction mais avant, nous avons fait nos analyses descriptives sur les données.

Les analyses descriptives sont la base de toute analyse de données. En effet, avant d'approfondir l'analyse dans les détails, il faut commencer par la description globale des variables. Il faut utiliser les outils de représentation graphiques qui permettent de modéliser les variables sous forme synthétique.

2.1 Description globale de la consommation et de la production éolienne

La méthode de prévision s'appuie généralement sur le caractère cyclique de la consommation comme dans notre cas avec la méthode de RTE. On peut identifier trois cycles temporels :

- Le cycle annuel
- Le cycle hebdomadaire
- Le cycle journalier

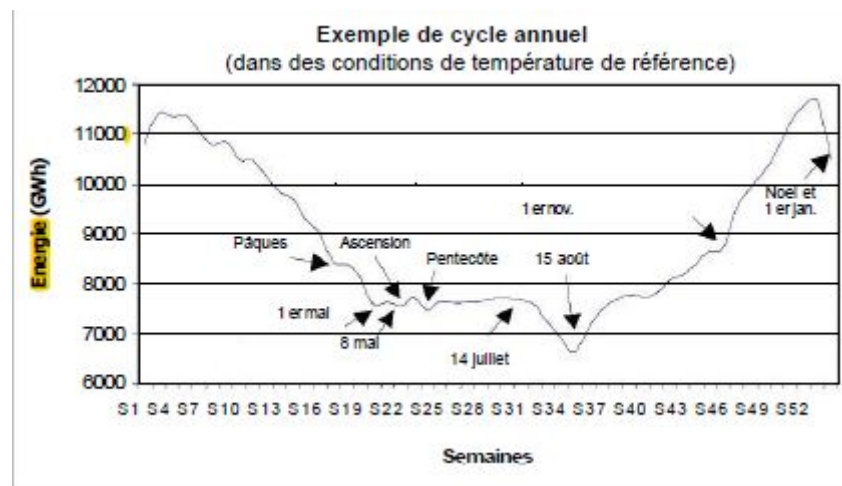


FIGURE 2.1 – Exemple de cycle annuel de consommation d'énergie électrique en 2011 [3].

On remarque ici un pic de consommation annuelle en décembre et un creux au 15 août qui correspond au jour férié fêtant l'Assomption. Avec les pics de température en été, on risque de voir apparaître aussi des augmentations de la consommation, pour les besoins de climatisation.

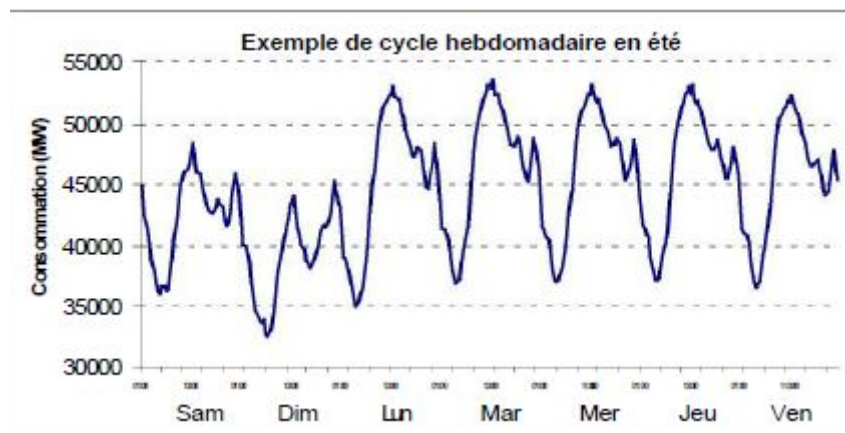


FIGURE 2.2 – Exemple de cycle hebdomadaire de consommation électrique en puissance (MW) [3].

On note ici une consommation stable pendant les 5 jours ouvrables et une baisse remarquable pendant le week-end.

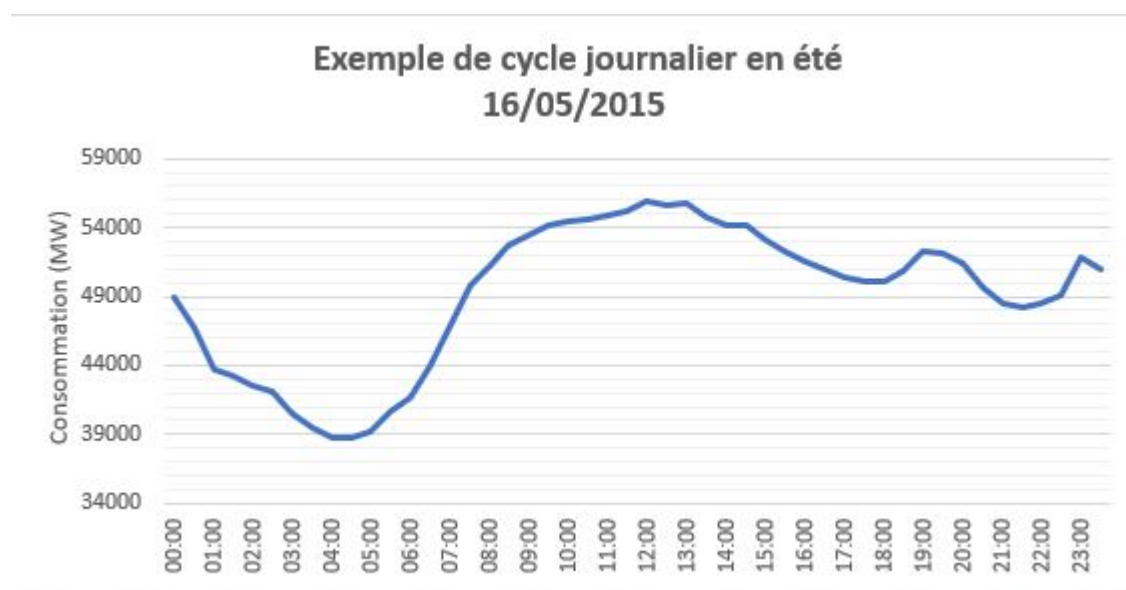


FIGURE 2.3 – Exemple d'une consommation électrique en été le 16/05/2015 en MWh *.



FIGURE 2.4 – Exemple d'une consommation électrique en hiver le 22/02/2015 en MWh *.

L'évolution de la consommation électrique sur une journée est caractérisé par 4 points : le creux de nuit, qui correspond à la consommation minimale sur les 24

heures de la journée, la pointe du matin, le creux d'après-midi et la pointe du soir. Le maximum de consommation est atteint à la pointe du matin vers 13h00 en été et à la pointe du soir vers 19h00 en hiver [3].

Les prévisions de la consommation et de la production éolienne sont établies par RTE. La production éolienne prévue dépend de conditions météorologiques et de la localisation du parc de générateurs éoliennes, ainsi que du type et du nombre d'aérogénérateurs. La prévision est basée sur l'historique et les prévisions météo.

2.2 Analyse des coûts

On a créé 4 nouvelles séries chronologiques représentant l'évolution de 4 nouvelles variables calculées pour chaque jour pendant une année (comme mentionné avant, on travaille toujours sur 2015, 2016 et 2017), les quatre nouvelles variables sont les suivantes : l'erreur de prévision de la consommation nette, le coût prévu de production, le coût d'ajustement de production et la variance de la demande résiduelle journalière. Chaque série de chaque année est stockée dans un fichier excel.

Il s'agit d'une analyse descriptive basée sur l'évolution temporelle de ces variables le long de l'année et ensuite, on a fait intervenir quelques études statistiques comme la corrélation entre ces variables puis la droite de régression linéaire par la méthode des moindres carrés ordinaire du coût d'ajustement de production en fonction de la variance de la demande journalière et finalement l'interprétation des données aberrantes (outliers). On a commencé premièrement par calculer l'erreur de prévision de la consommation nette X_h (2015, 2016 et 2017).

2.2.1 L'erreur de prévision de la consommation nette

La consommation nette est la différence entre la consommation totale de l'électricité et la production éolienne. L'erreur de prévision de la consommation nette est calculée comme la différence entre la prévision de la consommation nette et sa réalisation, cette dernière est calculée pour chaque heure du jour. En effet on a obtenu une série chronologique pour chaque année de 2015 à 2017 et aussi une pour toutes les années ensemble. On introduit la variable $X_{(t,h)}$ qui représente à chaque heure h de chaque jour t l'erreur de prévision de la consommation nette, on va se limiter sur un seul compteur h qui s'étend de la première heure du premier jour

jusqu'à la dernière heure du dernier jour. Par exemple, $h \in [1, 8760]$ pour l'année 2015 et $h \in [1, 26304]$ pour les trois années ensemble.

$$X_h = PD_h - PE_h - (RD_h - RE_h)$$

où PD_h est la prévision de consommation de la veille pour le lendemain, PE_h la prévision de production éolienne de la veille pour le lendemain, RD_h la réalisation de la consommation et RE_h la réalisation de la production éolienne.

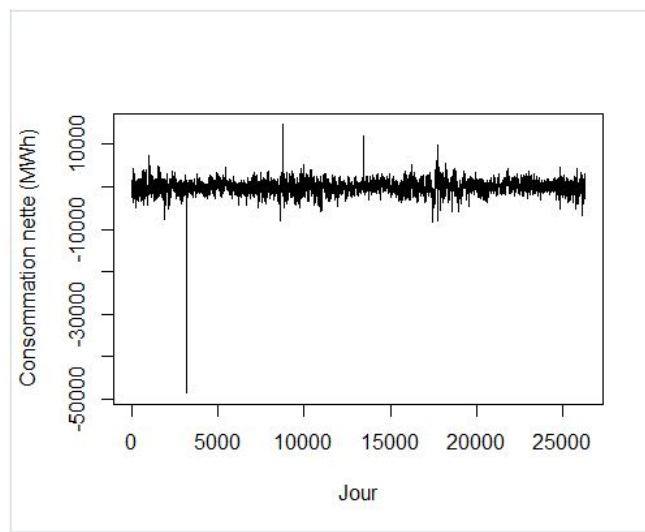


FIGURE 2.5 – L'évolution de l'erreur de prévision de la consommation nette au long des trois années (2015, 2016, 2017) *.

2.2.2 Coût prévu de production

Il s'agit d'une variable journalière qui représente la somme sur une journée de la différence entre la prévision de la consommation totale de la veille pour le lendemain et la prévision de la production éolienne pour le lendemain, multiplié par le prix spot day-ahead S_h au pas d'une heure.

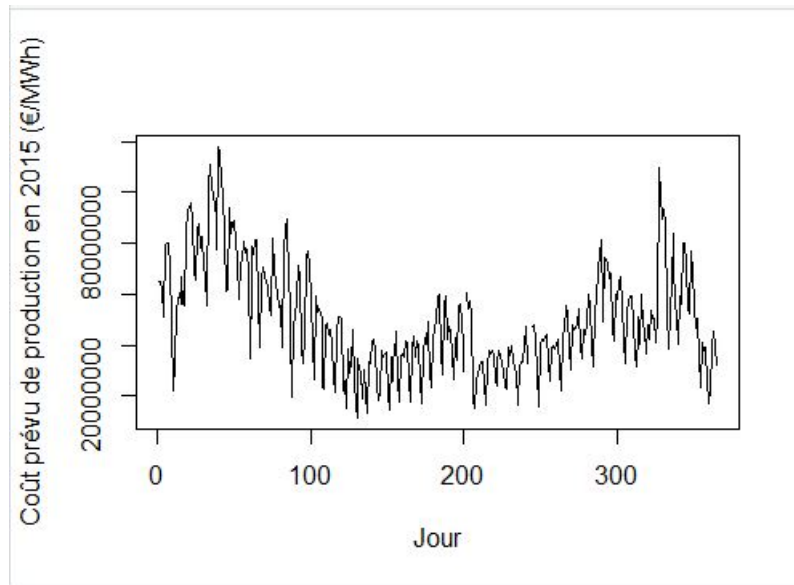


FIGURE 2.6 – Exemple du coût prévu de production pour l'année 2015 *.

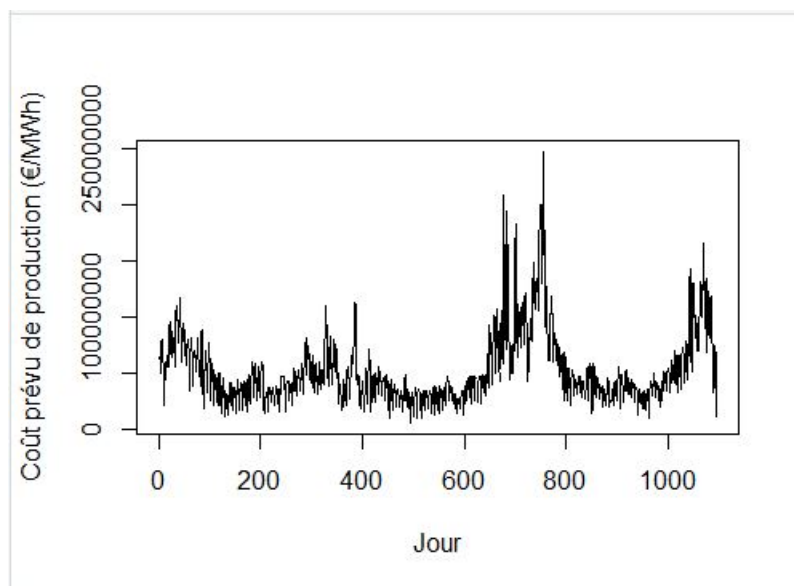


FIGURE 2.7 – L'évolution du coût prévu de production au long des trois années (2015, 2016 et 2017) *.

2.2.3 Coût d'ajustement de production

On va parler ici du prix règlement des écarts, il est lié au prix des offres d'ajustement posées par RTE pour assurer l'équilibrage du système électrique français. Les principes de calcul de ce prix permettent d'envoyer aux responsables d'équilibre¹, des opérateurs qui se sont engagés par contrat auprès de RTE à financer le coût des écarts constatés entre l'électricité introduite (production, achat en bourse, importation) et l'électricité soutirée (consommation, vente en bourse) (injections < soutirages) à l'intérieur d'un périmètre d'équilibre², une incitation financière sur leurs déséquilibres et reflètent le coût des actions d'équilibrage dirigées par RTE pour équilibrer entre l'offre et la demande.

Le coût d'ajustement de production est aussi une variable journalière qui est la somme sur une journée de la variable X_h multipliée par le PRE_h , prix règlement des écarts, au pas d'une heure. Donc, il faut appliquer le prix du règlement des écarts aux erreurs de prévision de la consommation nette afin d'ajuster la production.

Il faut mentionner qu'il existe deux valeurs pour le PRE_h , le PRE_h positif et le PRE_h négatif. Par ailleurs, si X_h est positif, cela signifie que la prévision était supérieure à la réalisation et que le système était donc en excès de production, il faut donc prendre le PRE_h positif. Dans le cas contraire, cela signifiait que le système était en déficit de production et il faut donc prendre le PRE_h négatif.

1. Les fournisseurs d'électricité (français ou étrangers), consommateurs (site d'un groupe, entreprise désignée par un groupe d'entreprises, etc...).

2. Au cas contraire, (injections > soutirages) ils doivent recevoir une compensation financière de RTE.

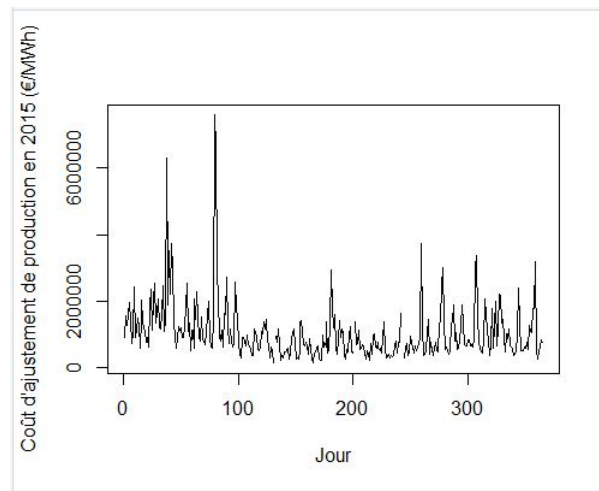


FIGURE 2.8 – Exemple du coût d'ajustement de production pour l'année 2015 *.

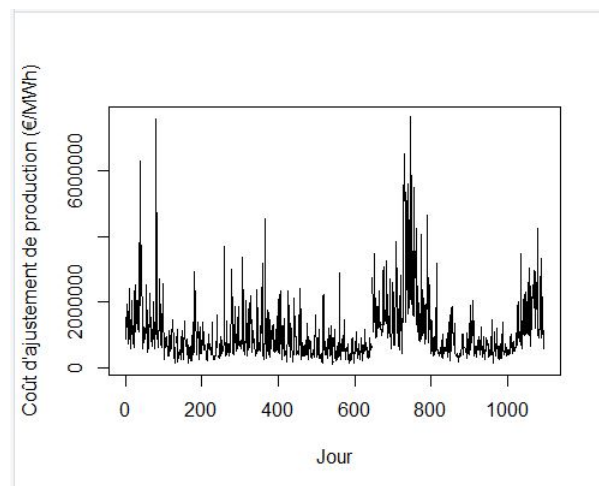


FIGURE 2.9 – L'évolution du coût d'ajustement de production au long des trois années (2015, 2016 et 2017) *.

D'après ce graphique on remarque qu'il y a une alternance de périodes tranquilles et de périodes de forte volatilité (Volatility clustering).

2.2.4 Variance de la demande résiduelle journalière

Il s'agit d'une variable journalière qui représente la somme sur une journée du carré de la différence entre 2 valeurs consécutifs de X_h au pas d'une heure.

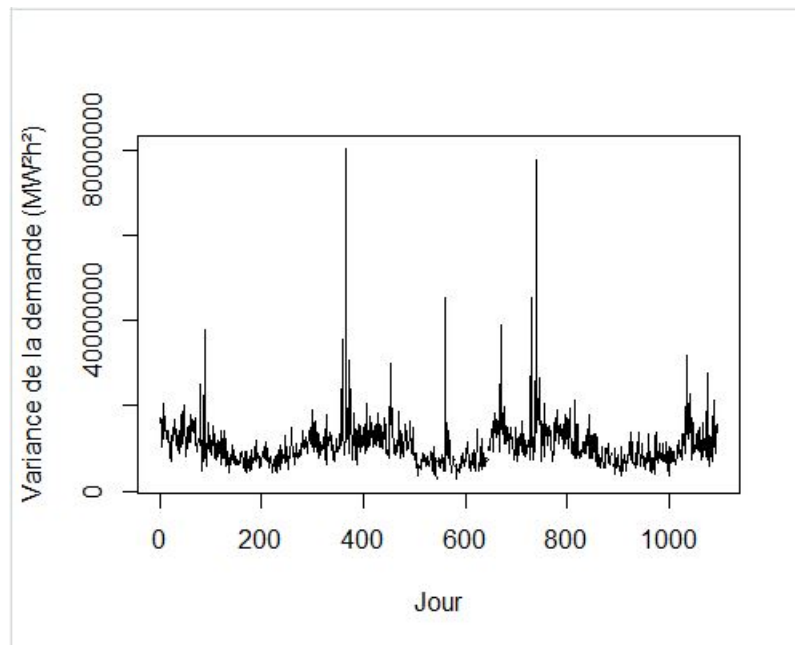


FIGURE 2.10 – L'évolution de la variance de la demande résiduelle journalière au long des trois années (2015, 2016 et 2017) *.

De même, on constate qu'il y a une alternance de périodes tranquilles et de périodes de fortes fluctuations.

On a tracé un graphique représentant le coût d'ajustement de production en fonction de la variance de la demande résiduelle journalière pour l'année 2017 accompagné de la droite de la régression linéaire et son équation et également un pour les trois années ensemble.

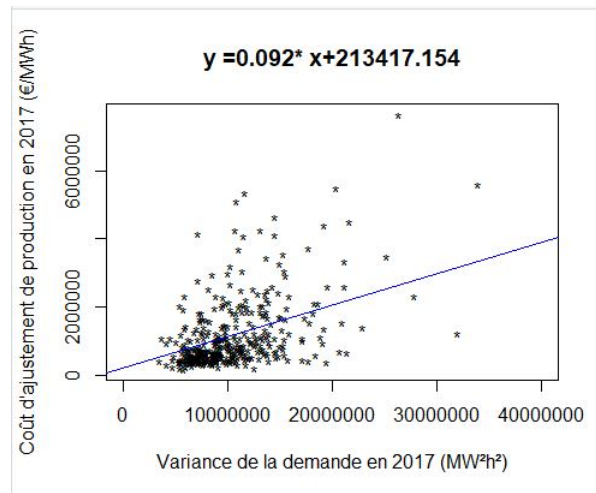


FIGURE 2.11 – Le coût d'ajustement de production en fonction de la variance de la demande résiduelle journalière en 2017 *.

```
call:
lm(formula = A2017 ~ v2017)

Residuals:
    Min       1Q   Median       3Q      Max
-2212451.8 -520224.0 -270199.6  310937.6 5015930.0

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 213417.153877631 105319.125797590  2.02639  0.043466 *
v2017        0.092000568    0.008716128 10.55521 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 925060.1 on 358 degrees of freedom
(5 observations deleted due to missingness)
Multiple R-squared:  0.2373445, Adjusted R-squared:  0.2352142
F-statistic: 111.4125 on 1 and 358 DF, p-value: < 0.00000000000000022204
```

FIGURE 2.12 – Les paramètres descriptifs de la régression linéaire *.

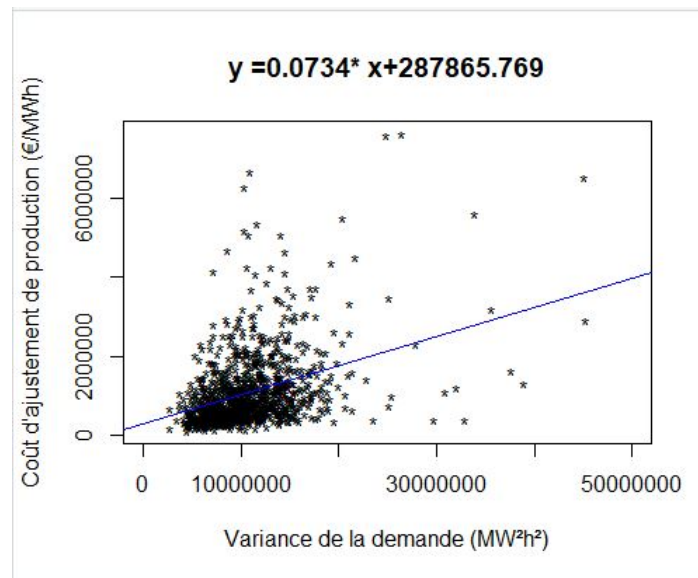


FIGURE 2.13 – Le coût d’ajustement de production en fonction de la variance de la demande résiduelle journalière pour les trois années (2015, 2016 et 2017) *.

```
call:
lm(formula = Ap ~ V)

Residuals:
    Min       1Q   Median       3Q      Max
-2320626.6 -462700.8 -225250.3  213880.5  5595818.2

Coefficients:
            Estimate      Std. Error t value      Pr(>|t|)
(Intercept) 287865.769309882    56362.903004161    5.10736  0.00000038609 ***
V             0.073427890       0.004708525   15.59467 < 0.000000000000000222 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 829526.6 on 1078 degrees of freedom
(16 observations deleted due to missingness)
Multiple R-squared:  0.1840712, Adjusted R-squared:  0.1833143
F-statistic: 243.1938 on 1 and 1078 DF, p-value: < 0.00000000000000022204
```

FIGURE 2.14 – Les paramètres descriptifs de la régression linéaire *.

On a constaté la significativité de résultats de la régression, les probabilités critiques sont inférieures aux risques de première espèce (1%, 5%, 10%) mais aussi le caractère un peu faible du coefficient de détermination R^2 . C’est le même cas pour la régression précédente.

Comme mentionné dans la première partie, dans la journée il y a une pointe le matin et autre le soir. On a tracé un graphique représentant le coût d’ajustement

de production en fonction de la variance de la demande résiduelle journalière pour la période entre 18h et 21h en 2017 accompagné de la droite de la régression linéaire et son équation et également un pour les trois années ensemble.

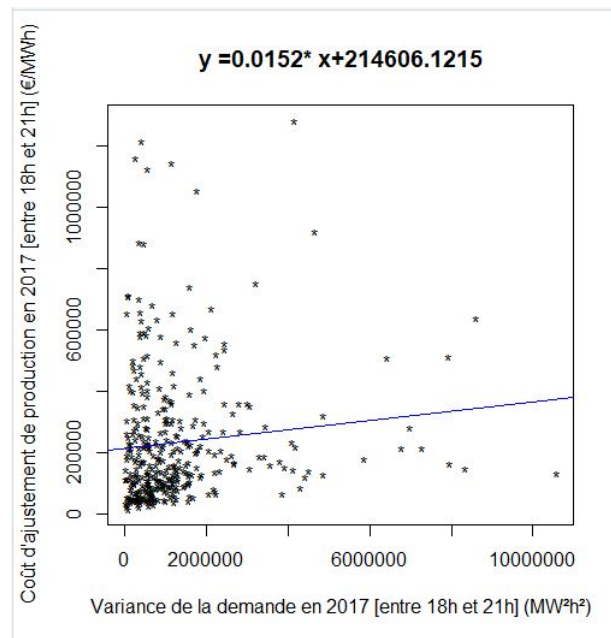


FIGURE 2.15 – Le coût d'ajustement de production en fonction de la variance de la demande résiduelle journalière (entre 18h et 21h) en 2017 *.

```
Call:
lm(formula = Ap2017 ~ vp2017)

Residuals:
    Min       1Q   Median       3Q      Max
-241197.83 -140100.54  -73487.27   60408.24 1003987.03

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 214606.12155693  14591.66276847  14.70745 < 0.0000000000000002 ***
vp2017       0.01516348   0.00764236   1.98414  0.048002 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 211280.9 on 359 degrees of freedom
(4 observations deleted due to missingness)
Multiple R-squared:  0.01084706,    Adjusted R-squared:  0.008091755
F-statistic: 3.936796 on 1 and 359 DF,  p-value: 0.04800182
```

FIGURE 2.16 – Les paramètres descriptifs de la régression linéaire *.

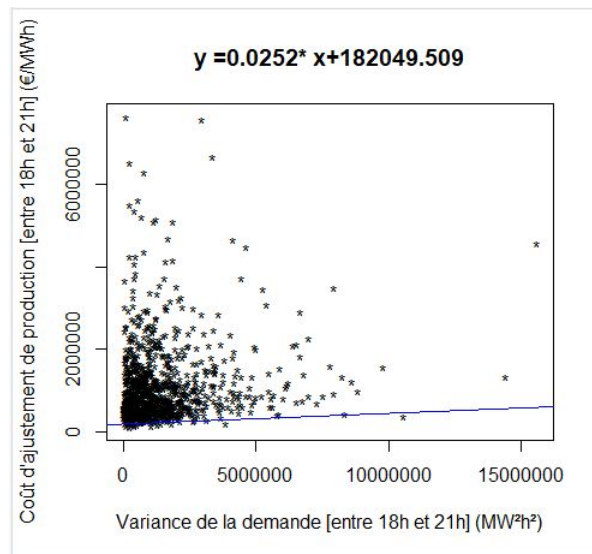


FIGURE 2.17 – Le coût d'ajustement de production en fonction de la variance de la demande résiduelle journalière (entre 18h et 21h) pour les trois années (2015, 2016 et 2017) *.

```
Call:
lm(formula = App ~ vp)

Residuals:
    Min       1Q   Median       3Q      Max
-314290.59 -129743.88  -67816.77   46756.81 1729402.31

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 182049.508580197  8496.017581420  21.42763 < 0.000000000000000222 ***
vp           0.025154893    0.004069056   6.18200  0.00000000089513 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 209829.4 on 1083 degrees of freedom
(11 observations deleted due to missingness)
Multiple R-squared:  0.03408537,    Adjusted R-squared:  0.03319348
F-statistic: 38.21709 on 1 and 1083 DF, p-value: 0.0000000008951343
```

FIGURE 2.18 – Les paramètres descriptifs de la régression linéaire *.

On note la significativité de résultats de la régression, les probabilités critiques sont inférieures aux risques de première espèce (1%, 5%, 10%) comme dans le cas du test statistique. C'est le même aussi pour la régression précédente au moins pour 10%. Mais, on remarque aussi que le coefficient de détermination R^2 est faible et très faible pour la régression précédente, qui peut s'opposer à l'établissement de la relation linéaire simple.

2.3 Explications des données aberrantes dans les coûts

Les figures précédentes sont déjà ajustées, c-à-d les graphiques sont tracés après le traitement des points extrêmes. Au début, on a vu des outliers dans les coûts et au niveau de la variance de la demande résiduelle journalière qui risquaient de biaiser les résultats des estimations.

On va présenter quelques points atypiques trouvés, par exemple et c'est le plus influant, l'absence des données de la prévision de consommation pour 00 :00 de certains jours, ce qui permet de prendre les données de la veille pour 23 :30. En outre, il y a d'autres explications comme le changement horaire soit celui d'été ou celui d'hiver, et aussi on peut les expliquer par les jours fériés comme la fête de nouvel an, les pâques etc...

Comme on a constaté que l'établissement d'une relation linéaire entre le coût de l'ajustement de production et la variance de la demande résiduelle est un peu difficile, donc il faut chercher à trouver d'autres méthodes de modélisation et de prévision pour mieux estimer finalement le coût de la volatilité.

Chapitre 3

Modélisation et estimation de la volatilité

On a utilisé des techniques de séries temporelles pour aborder notre étude sur la série représentant l'erreur de prévision de la consommation nette X_h . A cet effet, un modèle univarié dynamique peut être spécifié, puis estimé. Mais avant de passer au modèle, les propriétés des séries temporelles des données doivent être examinées afin d'éviter tout résultat d'estimation biaisé.

On va retourner un peu vers le graphique de X_h (voir la figure 2.5) avant d'aborder la partie suivante.

On a remarqué qu'au début de la série, précisément au début de l'année 2015, qu'il y a une donnée aberrante qui peut influencer sur le modèle de sorte qu'on peut obtenir des résultats fallacieux après. En effet, il faut soit éliminer cette donnée ou changer sa valeur, on a décidé de faire la moyenne entre la valeur de la donnée juste après et celle de juste avant puis l'affecter à cette donnée aberrante, ce qui nous donne après le graphique suivant :

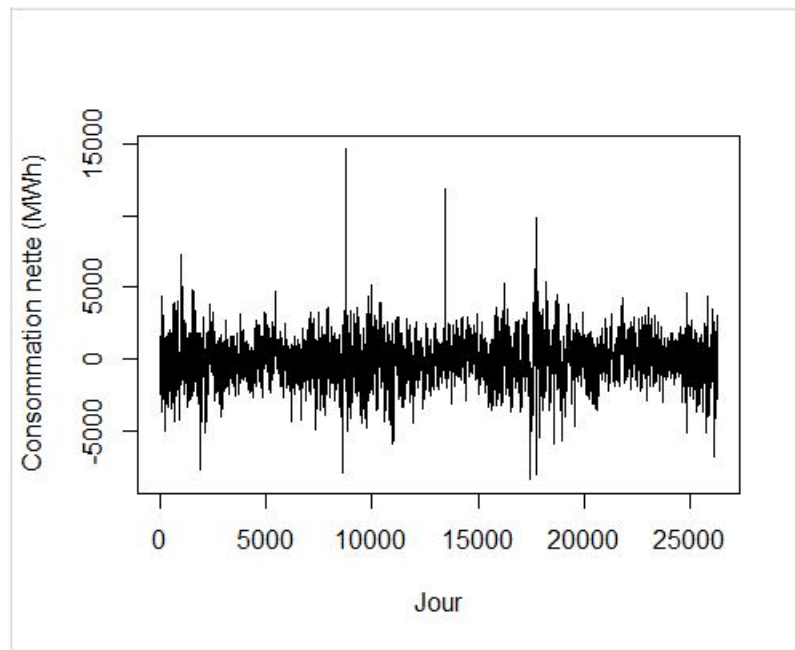


FIGURE 3.1 – L'évolution de l'erreur de prévision de la consommation nette au long des trois années (2015, 2016 et 2017) *.

Min.	Ecart-type	Median	Moyenne	Max.	NA's	Kurtosis	Skewness
-8398.00000	1404.46134	-95.00000	-99.13835	14603.00000	229	5.927676849	0.4559868354

FIGURE 3.2 – Les paramètres descriptifs de la série X_h *.

Le graphique nous montre que la série est très certainement stationnaire autour de sa moyenne et qu'elle semble présenter des changements de variance au cours du temps. On a vérifié quand même sa stationnarité à l'aide des tests statistiques.

3.1 Tests de racine unitaire

Une série temporelle x_t ($t=1,2,\dots$) est dite stationnaire (au sens faible) si ses propriétés statistiques sont invariables dans le temps (espérance, variance, autocovariance). L'autocovariance (ou l'autocorrélation) d'une série fait mention que dans une série temporelle, la mesure d'un phénomène à un instant t peut être dépendant de mesures précédentes (aux instants $t-1$, $t-2$, $t-3,\dots$) ou de mesures suivantes (à $t+1$, $t+2$, $t+3,\dots$). Une série autocorrélée est ainsi corrélée à elle-même, avec un décalage (lag=retard) donné.

Pour une variable x_t de moyenne μ et d'écart-type σ .

Autocovariance de x pour un décalage de h :

$$\gamma_h = E[(x_t - \mu)(x_{t+h} - \mu)]$$

Autocorrélation de x pour un décalage de h :

$$\rho_h = \frac{\gamma_h}{\sigma^2}$$

Un exemple de série temporelle stationnaire est le bruit blanc, une suite de variables aléatoire ϵ_t , $t \in \{0,1,2,\dots\}$:

- de moyenne nulle : $E(\epsilon_t) = 0, \forall t \in \{0,1,2,\dots\}$
- de variance constante : $V(\epsilon_t) = \sigma_\epsilon^2, \forall t \in \{0,1,2,\dots\}$
- d'autocorrélation nulle : $\gamma(h) = \text{cov}(\epsilon_t, \epsilon_{t+h}) = 0, \forall t \in \{0,1,2,\dots\}, \forall h \in \{0,1,2,\dots\}$

Il faut noter qu'une série non-stationnaire ouvre la voie vers l'étude de quel type de non-stationnarité il s'agit. Une série non-stationnaire peut quand même être stationnaire en différence : x_t n'est pas stationnaire, mais la différence $x_t - x_{t-1}$ est stationnaire, c'est le cas de la marche aléatoire ($x_t = x_{t-1} + \epsilon_t$ tel que ϵ_t est un bruit blanc). Une série peut également être stationnaire en tendance, c-à-d elle présente une composante additive en fonction du temps que l'on annule pour ne conserver que la composante stationnaire de la série.

Les tests de stationnarité permettent de vérifier si une série est stationnaire ou non. Il y a deux types différents de test : les tests de racine unitaire comme le test de Dickey-Fuller, le test augmenté de Dickey-Fuller (ADF), le test de Phillips-Perron (PP) etc... pour lesquels l'hypothèse nulle est que la série a été générée par un processus présentant une racine unitaire d'où elle n'est pas stationnaire, et les tests de stationnarité, comme le test KPSS, pour lesquels l'hypothèse nulle H_0 est que la série est stationnaire, et ce type de tests est utilisé souvent pour vérifier la robustesse d'un test de racine unitaire.

Sur la série de l'erreur de prévision de la consommation nette, on a appliqué premièrement le test ADF de Dickey-Fuller (DF) augmenté, puis le test des moindres carrés généralisés-Dickey Fuller augmenté (DF-GLS) et enfin le test de racine unitaire de Philippe-Perron (appelé parfois le test M) (voir l'annexe A). Les résultats de ces tests sous R sont ensuite affichés :

Augmented Dickey-Fuller Test alternative: stationary								Type 3: with drift and trend			
Type 1: no drift no trend				Type 2: with drift no trend							
lag	ADF	p.value		lag	ADF	p.value		lag	ADF	p.value	
[1,]	0	-48.5	0.01	[1,]	0	-48.7	0.01	[1,]	0	-48.7	0.01
[2,]	1	-37.8	0.01	[2,]	1	-37.9	0.01	[2,]	1	-37.9	0.01
[3,]	2	-30.8	0.01	[3,]	2	-30.9	0.01	[3,]	2	-30.9	0.01
[4,]	3	-30.5	0.01	[4,]	3	-30.6	0.01	[4,]	3	-30.6	0.01
[5,]	4	-28.5	0.01	[5,]	4	-28.6	0.01	[5,]	4	-28.6	0.01
[6,]	5	-27.9	0.01	[6,]	5	-28.0	0.01	[6,]	5	-28.0	0.01
[7,]	6	-27.4	0.01	[7,]	6	-27.5	0.01	[7,]	6	-27.5	0.01
[8,]	7	-27.7	0.01	[8,]	7	-27.8	0.01	[8,]	7	-27.8	0.01
[9,]	8	-27.4	0.01	[9,]	8	-27.5	0.01	[9,]	8	-27.5	0.01
[10,]	9	-23.7	0.01	[10,]	9	-23.8	0.01	[10,]	9	-23.8	0.01
[11,]	10	-22.7	0.01	[11,]	10	-22.8	0.01	[11,]	10	-22.8	0.01
[12,]	11	-23.8	0.01	[12,]	11	-23.9	0.01	[12,]	11	-23.9	0.01
[13,]	12	-21.7	0.01	[13,]	12	-21.7	0.01	[13,]	12	-21.7	0.01
[14,]	13	-22.0	0.01	[14,]	13	-22.1	0.01	[14,]	13	-22.1	0.01
[15,]	14	-23.2	0.01	[15,]	14	-23.3	0.01	[15,]	14	-23.3	0.01
[16,]	15	-21.7	0.01	[16,]	15	-21.8	0.01	[16,]	15	-21.8	0.01
[17,]	16	-20.0	0.01	[17,]	16	-20.1	0.01	[17,]	16	-20.1	0.01
[18,]	17	-18.6	0.01	[18,]	17	-18.7	0.01	[18,]	17	-18.7	0.01
[19,]	18	-17.9	0.01	[19,]	18	-17.9	0.01	[19,]	18	-18.0	0.01
[20,]	19	-17.8	0.01	[20,]	19	-17.8	0.01	[20,]	19	-17.8	0.01
[21,]	20	-16.6	0.01	[21,]	20	-16.7	0.01	[21,]	20	-16.7	0.01
[22,]	21	-17.2	0.01	[22,]	21	-17.2	0.01	[22,]	21	-17.2	0.01
[23,]	22	-14.9	0.01	[23,]	22	-14.9	0.01	[23,]	22	-14.9	0.01
[24,]	23	-11.5	0.01	[24,]	23	-11.6	0.01	[24,]	23	-11.6	0.01
[25,]	24	-17.3	0.01	[25,]	24	-17.4	0.01	[25,]	24	-17.4	0.01
[26,]	25	-18.8	0.01	[26,]	25	-18.9	0.01	[26,]	25	-18.9	0.01
[27,]	26	-18.3	0.01	[27,]	26	-18.3	0.01	[27,]	26	-18.4	0.01
[28,]	27	-17.9	0.01	[28,]	27	-18.0	0.01	[28,]	27	-18.0	0.01
[29,]	28	-17.4	0.01	[29,]	28	-17.5	0.01	[29,]	28	-17.5	0.01

Note: in fact, p.value = 0.01 means p.value <= 0.01

FIGURE 3.3 – Le test ADF pour la série de l'erreur de prévision de la consommation nette *.

##### # Elliot, Rothenberg and Stock Unit Root Test # #####									
Test of type DF-GLS detrrending of series with intercept									
Coefficients:									
	Estimate	Std. Error	t value	Pr(> t)					
yd.lag	-0.005773792	0.001671523	-3.45421	0.00055456 ***					
yd.diff.lag1	-0.081938105	0.010785580	-7.59700	0.0000000000000334711 ***					
yd.diff.lag2	-0.046539511	0.010790757	-4.31291	0.0000162915261635545 ***					
yd.diff.lag3	-0.055012579	0.008748482	-6.28824	0.0000000003366643618 ***					
yd.diff.lag4	-0.092851699	0.008618447	-10.77360	< 0.000000000000000222 ***					
yd.diff.lag5	-0.050373793	0.008539089	-5.89920	0.0000000037902160043 ***					
yd.diff.lag6	-0.057356822	0.008531527	-6.72293	0.0000000000189415166 ***					
yd.diff.lag7	-0.057693283	0.008524863	-6.76765	0.0000000000139423755 ***					
yd.diff.lag8	-0.073650640	0.008538009	-8.62621	< 0.000000000000000222 ***					
yd.diff.lag9	-0.088868106	0.008558568	-10.38352	< 0.000000000000000222 ***					
yd.diff.lag10	-0.047263481	0.008589173	-5.50268	0.0000000384761959189 ***					
yd.diff.lag11	-0.009965317	0.008557595	-1.16450	0.24425371					
yd.diff.lag12	-0.056295241	0.008516969	-6.60977	0.000000000407721378 ***					
yd.diff.lag13	-0.027133480	0.008528139	-3.18164	0.00146962 **					
yd.diff.lag14	-0.034255293	0.008527133	-4.01721	0.0000593904487221469 ***					
yd.diff.lag15	-0.074362221	0.008507568	-8.74071	< 0.000000000000000222 ***					
yd.diff.lag16	-0.077519677	0.008543489	-9.07354	< 0.000000000000000222 ***					
yd.diff.lag17	-0.050877189	0.008565316	-5.93991	0.000000029621700998 ***					
yd.diff.lag18	-0.042544113	0.008527548	-4.98902	0.0000006187016191505 ***					
yd.diff.lag19	-0.032399711	0.008497117	-3.81302	0.00013823 ***					
yd.diff.lag20	-0.060865487	0.008476586	-7.18042	0.000000000007525451 ***					
yd.diff.lag21	-0.055198655	0.008475469	-6.51276	0.000000000779058784 ***					
yd.diff.lag22	-0.137339509	0.008474068	-16.20703	< 0.000000000000000222 ***					
yd.diff.lag23	-0.143506332	0.008537273	-16.80939	< 0.000000000000000222 ***					
yd.diff.lag24	0.594869351	0.008650872	68.76409	< 0.000000000000000222 ***					
yd.diff.lag25	0.069426751	0.010747778	-6.45964	0.000000001106269784 ***					
yd.diff.lag26	-0.083827129	0.010731038	-7.81165	0.0000000000000063058 ***					

					Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
					Residual standard error: 446.6105 on 8634 degrees of freedom				
					Multiple R-squared: 0.5863623, Adjusted R-squared: 0.5850688				
					F-statistic: 453.3089 on 27 and 8634 DF, p-value: < 0.000000000000000222				
					Value of test-statistic is: -3.4542				
					Critical values of DF-GLS are:				
					1pct 5pct 10pct				
					critical values -2.57 -1.94 -1.62				

FIGURE 3.4 – Le test DF-GLS pour la série de l'erreur de prévision de la consommation nette *.

```
#####
# Phillips-Perron Unit Root Test #
#####

Test regression with intercept

Call:
lm(formula = y ~ y.l1)

Residuals:
    Min       1Q   Median       3Q      Max
-6562.7255 -413.8701  21.2930  424.7232 4982.2444

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -12.664029056  4.267474622  -2.96757  0.0030044 **
y.l1         0.872062443  0.003030989 287.71548 < 0.000000000000000222 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 687.3766 on 26072 degrees of freedom
Multiple R-squared:  0.7604826, Adjusted R-squared:  0.7604734
F-statistic: 82780.2 on 1 and 26072 DF, p-value: < 0.00000000000000022204

Value of test-statistic, type: Z-alpha is: -3430.3403

      aux. Z statistics
Z-tau-mu      -3.0066
```

FIGURE 3.5 – Le test M pour la série de l'erreur de prévision de la consommation nette *.

Nous pouvons voir que les trois tests se comportent de manière similaire pour cette série. En effet, quel que soit les composantes déterministes et le nombre de retards choisis, la probabilité critique est toujours nulle ou presque ce qui nous conduit à rejeter l'hypothèse nulle de série non stationnaire.

```

KPSS Unit Root Test
alternative: nonstationary

Type 1: no drift no trend
lag stat p.value
37 2.94 0.01
-----
Type 2: with drift no trend
lag stat p.value
37 0.278 0.1
-----
Type 1: with drift and trend
lag stat p.value
37 0.116 0.1
-----
Note: p.value = 0.01 means p.value <= 0.01
      : p.value = 0.10 means p.value >= 0.10
    
```

FIGURE 3.6 – Le test KPSS pour la série de l'erreur de prévision de la consommation nette *.

Le test KPSS valide quant à lui l'hypothèse nulle selon laquelle la série est stationnaire sauf dans le cas avec ni tendance et ni constante mais la moyenne de X_h est différente de 0, alors cette spécification ne serait pas pertinente.

3.2 Le modèle ARMA

Le modèle ARMA (modèle autorégressif et moyenne mobile), appelé parfois modèle de Box-Jenkins, permet de représenter plusieurs processus aléatoires stationnaires. Il est très utilisé dans l'économétrie pour modéliser des séries économique. Avant d'aborder la structure de notre modèle, il faut noter qu'il faut quitter à chercher une représentation par le modèle ARMA dès que notre série présente l'absence de l'autocorrélation et l'autocorrélation partielle¹. Dans notre cas, cette condition que nous allons vérifier après est remplie.

1. Le coefficient d'autocorrélation partielle r_h^h mesure la liaison directe entre x_t et x_{t+h} une fois retirés les liens transitants par les variables intermédiaires : $x_{t+1}, x_{t+2}, \dots, x_{t+h-1}$.

3.2.1 Spécification, estimation et validation du modèle

Spécification

Étant donné une série temporelle x_t , le modèle ARMA est un outil pour comprendre et prédire, éventuellement, les valeurs futures de cette série. Le modèle est composé de deux parties : une partie autorégressive (AR) et une partie moyenne-mobile (MA). Le modèle est généralement noté ARMA(p,q), où p est l'ordre de la partie AR et q l'ordre de la partie MA. Pour expliquer les deux parties mentionnées au dessus, on va définir l'opérateur suivant.

Opérateur retard : Noté L pour lag, il est défini de la façon suivante :
On considère un processus aléatoire $(x_t, t \in \mathbb{N})$, l'opérateur retard, noté L, est défini comme suivant :

$$Lx_t = x_{t-1} \quad \forall t \in \mathbb{N}$$

Cet opérateur permet ainsi de définir une application qui à toute variable x_t associe la variable retardée x_{t-1} . Cet opérateur possède la propriété suivante :

$$L^j x_t = x_{t-j} \quad \forall t \in \mathbb{N}, \text{ particulièrement on a } L^0 x_t = x_t$$

Nous allons définir maintenant les deux parties AR et MA.

Le processus AR : Le processus stationnaire $(x_t, \forall t \in \mathbb{N})$ satisfait une représentation AR d'ordre p (p est le nombre de retards), notée AR(p), si et seulement si :

$$A(L)x_t = mu + \epsilon_t$$

avec $mu \in \mathbb{R}$, $A(L) = \sum_{j=0}^p ar_j L^j$ où $\forall j < p$, $ar_j \in \mathbb{R}$, $ar_0 = 1$ et $ar_p \in \mathbb{R}^*$, avec ϵ_t un bruit blanc. Si la constante est omise, le processus étant alors dit centré. On parle ici de représentation autorégressive, dans le sens où la variable x_t est déterminée par les valeurs passées $x_{t-1}, x_{t-2}, \dots, x_{t-p}$.

Des contraintes supplémentaires sur les paramètres sont nécessaires pour garantir la stationnarité. les processus où A admet au moins une racine inférieure à 1 en valeur absolue ne sont pas stationnaires.

Le processus MA : Le processus $(x_t, \forall t \in \mathbb{N})$ a une représentation MA d'ordre q (nombre de retards), notée $MA(q)$, si et seulement si :

$$x_t = B(L)\epsilon_t + mu$$

avec $E(x_t) = mu$. Le polynôme $B(L)$ étant défini par $B(L) = \sum_{j=0}^q ma_j L^j$ où $\forall j < q, ma_j \in \mathbb{R}, ma_0 = 1$ et $ma_q \in \mathbb{R}^*$, avec ϵ_t un bruit blanc.

Un processus $MA(q)$ est défini comme la somme des valeurs passées et présente d'un bruit blanc. Il est donc stationnaire par construction.

Remarque : Les processus $AR(p)$ et $MA(q)$ se distinguent par les propriétés sur les fonctions d'autocorrélation :

- x_t suit $MA(q)$, $\rho(h)$ (*autocorrélation*) = 0, $\forall h > q$.
- x_t suit $AR(p)$, r_h^h (*autocorrélation partielle*) = 0, $\forall h > p$.

Il s'agit des critères permettant de savoir si une série doit être modélisée par un $AR(p)$ ou $MA(q)$.

On revient maintenant à notre modèle $ARMA(p,q)$. Comme mentionné avant, $ARMA(p,q)$ est un modèle autorégressif et moyenne-mobile d'ordres (p,q) . C'est un processus temporel discret $(x_t, t \in \mathbb{N})$ vérifiant :

$$A(L)x_t = mu + B(L)\epsilon_t$$

avec $mu \in \mathbb{R}$, A et B les mêmes au dessus, $ar_0 = ma_0 = 1$ et $(ar_p, ma_p) \in \mathbb{R}^{*2}$, avec ϵ_t bruit blanc (terme d'erreur).

Comme pour le modèle AR , il y a une condition à vérifier pour assurer la stationnarité de la série x_t qui est similaire à celle du modèle AR (les racines de A doivent être supérieures à 1 en valeur absolue).

Remarque : Pour $ARMA(p,q)$ stationnaire, la fonction d'autocorrélation $\rho(h)$ et la fonction d'autocorrélation partielle r_h^h décroissent vers 0 à mesure que h augmente. D'où pas d'annulation de $\rho(h)$ et r_h^h à partir des rangs p ou q . Il y a aussi une autre condition qui sert à fournir le modèle le plus parcimonieux en termes de paramètres qui suppose que $A(L)$ et $B(L)$ n'ont pas des racines communes.

Jusqu'à maintenant, tout est bon avec la série représentant l'erreur de prévision de la consommation nette, elle vérifie la condition de stationnarité. On doit maintenant déterminer le nombre de retards p et q pour la modélisation de cette série. Du coup, l'identification de p et q est à l'aide de l'autocorrélogramme² de la série.

2. Représentation graphique des autocorrélations.

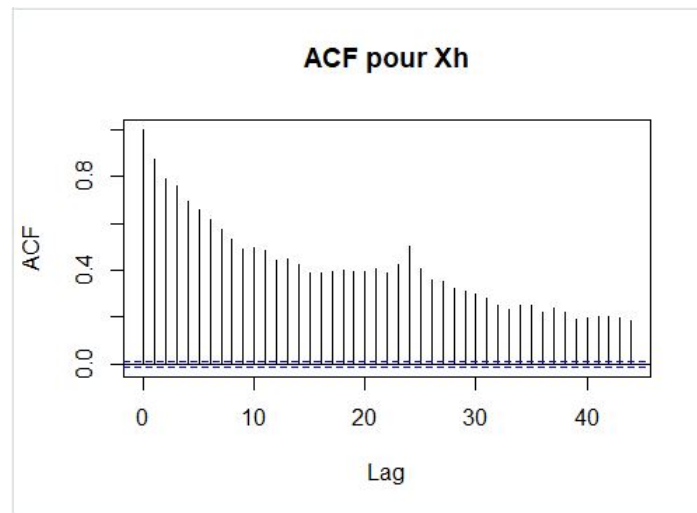


FIGURE 3.7 – La fonction d'autocorrélation pour l'erreur de prévision de la consommation nette *.

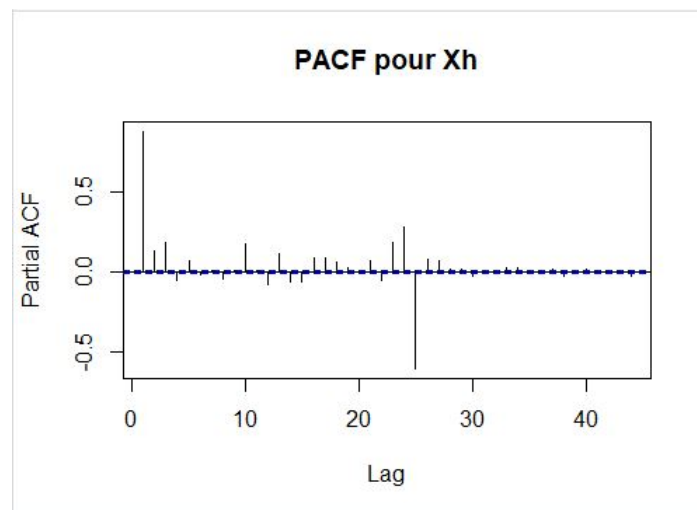


FIGURE 3.8 – L'autocorrélation partielle pour l'erreur de prévision de la consommation nette *.

Les deux outils déjà présentés, fonction d'autocorrélation et fonction d'autocorrélation partielle, jouent un rôle essentiel dans l'identification du processus générateur d'une variable. Les autocorrélations décroissent mais très lentement. Pour l'autocorrélation partielle, on observe des pics qui traduisent des saisonnalités. C'est très nette pour un retard de 24 h, mais aussi pour d'autres retards, comme 10 retards.

Estimation

D'après les deux graphiques précédents d'autocorrélation, on a attribué (25,1) à (p,q), également on a commencé par estimer un modèle ARMA(25,1) :

$$X_t = mu + \sum_{i=1}^{25} ar_i X_{t-i} + \epsilon_t + ma\epsilon_{t-1}$$

ar1	ar2	ar3	ar4	ar5
0.8153314917218	0.0556244069939	0.0577119860444	-0.0264189820434	0.0200224891785
ar6	ar7	ar8	ar9	ar10
-0.0066165242692	-0.0068819833165	-0.0201548663067	-0.0138817480936	0.0434692251775
ar11	ar12	ar13	ar14	ar15
0.0316341941089	-0.0443437514192	0.0305595278173	-0.0068050518994	-0.0448407270789
ar16	ar17	ar18	ar19	ar20
-0.0005753971272	0.0257866870923	0.0127621625535	0.0179424141953	-0.0292855672234
ar21	ar22	ar23	ar24	ar25
0.0099002463603	-0.0805541496747	-0.0254695042521	0.6983576937444	-0.5751085779359
mal	intercept			
0.0996452318541	-99.1368920757757			

FIGURE 3.9 – Les coefficients du modèle *.

Test de significativité des coefficients

On a vérifié la significativité des paramètres estimés avec la statistique de Student.

ar1	ar2	ar3
0.00000000000000000000000000000000	0.00000000000000009103828802	0.00000000000000000000000000000000
ar4	ar5	ar6
0.000004790007872346180307	0.000454815918435214072701	0.246482262339335900236392
ar7	ar8	ar9
0.227432407701303240799007	0.000407663301005944944677	0.014996319186891149399798
ar10	ar11	ar12
0.0000000000000022870594307	0.000000034065884335632290	0.00000000000000007771561172
ar13	ar14	ar15
0.000000098450699148600052	0.234819621597118333156118	0.0000000000000003774758284
ar16	ar17	ar18
0.920045409403046798857417	0.000006067363896411492874	0.025403902553136070707751
ar19	ar20	ar21
0.001646199996738539894636	0.000000279578139528169345	0.083160389318451377604902
ar22	ar23	ar24
0.00000000000000000000000000000000	0.000012132084778793128521	0.00000000000000000000000000000000
ar25	mal	inter cept
0.00000000000000000000000000000000	0.00000000000000000000000000000000	0.064040424839061804718199

FIGURE 3.10 – Les probabilités critiques associées aux coefficients $*$.

On remarque tout d'abord la non significativité de quelques paramètres donc on a décidé de modifier la structure de notre modèle en fixant la valeur de quelques coefficients dès le début. En effet, si on revient à la figure représentant la fonction de l'autocorrélation partielle, on trouve qu'il y a des autocorrélations intermédiaires non significatives où on a décidé de rendre la valeur des coefficients liés à ces retards nulle.

Estimation

Le nouveau modèle est le suivant :

$$X_t = mu + \sum_{i=1}^5 ar_i X_{t-i} + \sum_{i=21}^{25} ar_i X_{t-i} + \epsilon_t + ma\epsilon_{t-1}$$

ar1	ar2	ar3	ar4	ar5	ar6
0.82165448635	0.04137358364	0.06352801822	-0.03082902168	0.01678313763	0.00000000000
ar7	ar8	ar9	ar10	ar11	ar12
0.00000000000	0.00000000000	0.00000000000	0.00000000000	0.00000000000	0.00000000000
ar13	ar14	ar15	ar16	ar17	ar18
0.00000000000	0.00000000000	0.00000000000	0.00000000000	0.00000000000	0.00000000000
ar19	ar20	ar21	ar22	ar23	ar24
0.00000000000	0.00000000000	0.01076674773	-0.08599589039	-0.02607332171	0.72114541878
ar25	ma1	intercept			
-0.59514772618	0.09351419344	-99.13824192705			

FIGURE 3.11 – Les coefficients du modèle *.

Test de la significativité des coefficients

La significativité des paramètres estimés avec la statistique de Student.

z test of coefficients:					
	Estimate	Std. Error	z value	Pr(> z)	
ar1	0.8216544864	0.0075737363	108.48734	< 0.000000000000000222	***
ar2	0.0413735836	0.0067830453	6.09956	0.0000000010636	***
ar3	0.0635280182	0.0054488247	11.65903	< 0.000000000000000222	***
ar4	-0.0308290217	0.0055268776	-5.57802	0.0000000243275	***
ar5	0.0167831376	0.0045682340	3.67388	0.0002389	***
ar21	0.0107667477	0.0045651584	2.35846	0.0183509	*
ar22	-0.0859958904	0.0054453032	-15.79267	< 0.000000000000000222	***
ar23	-0.0260733217	0.0055695331	-4.68142	0.000028489501	***
ar24	0.7211454188	0.0055549894	129.81941	< 0.000000000000000222	***
ar25	-0.5951477262	0.0053635776	-110.96096	< 0.000000000000000222	***
ma1	0.0935141934	0.0090267254	10.35970	< 0.000000000000000222	***
intercept	-99.1382419270	52.7533641330	-1.87928	0.0602065	.

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

FIGURE 3.12 – Les probabilités critiques associées aux coefficients *.

On note la significativité des paramètres au moins pour 10% comme risque de première espèce. On va passer maintenant aux tests de validité de notre modèle sur les résidus.

3.2.2 Tests sur les résidus

On conserve le dernier modèle construit, les tests de cette partie servent à vérifier que les résidus $\hat{\epsilon}_t$ se comportent comme un bruit blanc.

Tests de Ljung-Box d'absence d'autocorrélation des résidus

On teste l'hypothèse d'absence d'autocorrélation des résidus jusqu'au rang k :

$$\begin{cases} H_0 : \rho_1 = \rho_2 = \dots = \rho_k = 0 \\ H_1 : \exists i \in \{1, 2, \dots, k\} \text{ t.q. } \rho_i \neq 0 \end{cases}$$

ρ_i est l'autocorrélation des résidus $\hat{\epsilon}_t$.

```
Box-Ljung test
data:  r
x-squared = 77.270555, df = 10, p-value = 0.000000000001716849
```

FIGURE 3.13 – Test d'absence d'autocorrélation jusqu'à l'ordre 10 *.

Dans ce cas, on rejette H_0 car la probabilité est inférieure au risque de première espèce. Par conséquent, on doit estimer un nouveau modèle, en augmentant le nombre de retards pour la partie AR ou la partie MA.

On a choisit d'augmenter celui de la partie MA jusqu'à obtenir l'absence d'autocorrélation des résidus, en effet on a obtenu le modèle ARMA (25,4) et on a refait les tests précédents affichés dans l'annexe B mais cette fois-ci, on a annulé aussi les coefficients liés aux retards d'ordre 2 et 4 et aussi la constante.

```
Box-Ljung test
data:  r
x-squared = 15.181656, df = 10, p-value = 0.125578
```

FIGURE 3.14 – Test d'absence d'autocorrélation jusqu'à l'ordre 10 *.

On ne rejette pas l'hypothèse nulle d'où l'absence de l'autocorrélation des résidus. On continue maintenant avec les autres tests sur les résidus du nouveau modèle.

Test de l'hypothèse de normalité des résidus

On doit vérifier si les résidus $\hat{\epsilon}_t$ suivent une loi normale. Ce test est basé sur le skewness S^3 et le kurtosis K^4 . On a utilisé le test appelé Jarque et Bera qui est fondé sur la comparaison entre le skewness et le kurtosis estimé à partir d'un échantillon et les valeurs de référence de la loi normale.

Les hypothèses du test sont :

$$\begin{cases} H_0 : S(X) = 0 \text{ et } K(X) = 3 \\ H_1 : S(X) \neq 0 \text{ ou } K(X) \neq 3 \end{cases}$$

```
Jarque Bera Test
data: r
x-squared = 53083.42, df = 2, p-value < 0.00000000000000022204
```

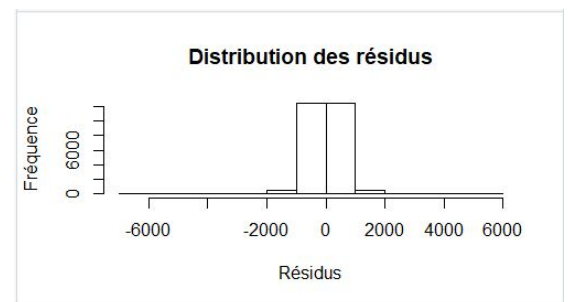


FIGURE 3.15 – Test de la normalité des résidus *.

On rejette la normalité de la série mais on a rien à faire, pas de modification. cette non-normalité des résidus s'explique par le fait que la variance conditionnelle change au cours du temps ce que nous allons affirmer dans la paragraphe suivante.

Tests d'homoscédasticité

On veut vérifier si la variance du résidu $\hat{\epsilon}_t$ ne dépend pas du temps. On utilise $\hat{\epsilon}_t^2$ comme une proxy de la volatilité :

3. Le coefficient d'asymétrie d'une distribution.
4. Le coefficient d'aplatissement.

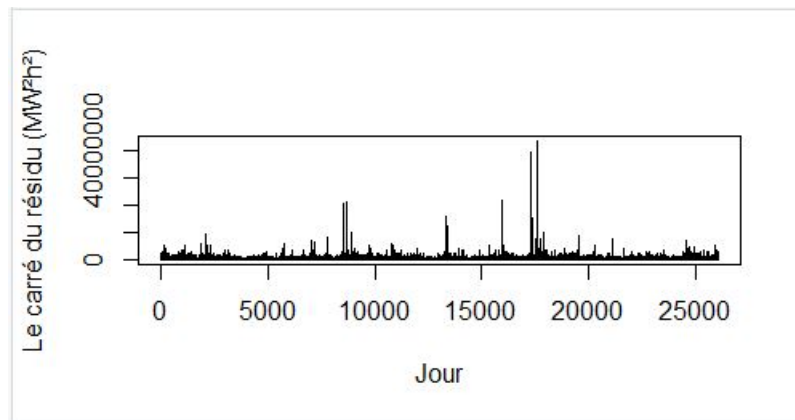


FIGURE 3.16 – Cluster de volatilité *.

D'après ce graphique on remarque qu'il y a une alternance de périodes tranquilles et de périodes de forte volatilité (Volatility clustering).
On applique le test d'un effet ARCH qui est fondé sur l'estimation de la régression :

$$\hat{\epsilon}_t^2 = a_0 + a_1 \hat{\epsilon}_{t-1}^2 + \dots + a_q \hat{\epsilon}_{t-q}^2 + u_t$$

Les hypothèses sont :

$$\begin{cases} H_0 : a_1 = a_2 = \dots = a_q = 0 \implies \text{pas d'effet ARCH} \\ H_1 : a_1 \neq 0 \text{ ou } a_2 \neq 0 \dots \text{ou } a_q \neq 0 \implies \text{effet ARCH} \end{cases}$$

	order	PQ	p.value	LM	p.value
[1,]	4	2221.813670	0	29330.250243	0
[2,]	8	3066.675948	0	13435.123678	0
[3,]	12	3408.009085	0	8837.413230	0
[4,]	16	3723.656209	0	6556.139574	0
[5,]	20	4187.391303	0	5097.225335	0
[6,]	24	6278.913262	0	4173.886942	0

FIGURE 3.17 – Test d'un effet ARCH pour les résidus *.

La probabilité critique est nulle ce qui entraîne le rejet de l'hypothèse nulle d'absence d'effet ARCH. Par conséquent, on pourra estimer un modèle de volatilité conditionnelle (Modèle GARCH). En effet, on s'est dirigé vers la construction d'un modèle ARMA (25,4)-GARCH.

3.3 Le modèle ARMA-GARCH

Puisque la série temporelle comporte une volatilité variable c-à-d des périodes de fortes fluctuations suivies par des périodes de faibles fluctuations, la variance conditionnelle au temps t est variable. Elle dépend par exemple du carré des réalisations précédentes du processus ou du carré des innovations (variance hétéroscédastique). On fait intervenir le modèle GARCH maintenant pour estimer la volatilité conditionnelle de la série.

Ces modèles sont constitués de deux équations :

- Une equation de la moyenne décrivant la dynamique de l'erreur de prévision de la consommation nette
- Une equation de la variance décrivant l'évolution de la variance de l'erreur de prévision de la consommation nette au cours du temps

On conserve l'ancien modèle ARMA (25,4) construit avant pour décrire la dynamique de l'erreur de prévision de la consommation nette. Dans lequel ϵ_t représente les résidus du modèle et $\epsilon_t = \sigma_t U_t$ et U_t , une normale centrée réduite. Un modèle GARCH(p,q) peut par exemple être défini de la manière suivante :

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^q \beta_i \sigma_{t-i}^2$$

finalement, notre modèle est ARMA (25,4)-GARCH(p,q) où p et q à préciser dans la partie suivante.

3.3.1 Estimation du modèle

On va premièrement aborder le modèle ARMA(25,4)-GARCH(1,1) puis le modèle ARMA(25,4)-GARCH(1,2).

Le modèle ARMA(25,4)-GARCH(1,1) :

On a estimé le modèle ARMA(25,4)-GARCH(1,1) pour la série qui nous a donné les paramètres suivant :

Optimal Parameters				
	Estimate	Std. Error	t value	Pr(> t)
mu	0.000000	NA	NA	NA
ar1	0.732361	0.010607	69.0455	0.000000
ar2	0.000000	NA	NA	NA
ar3	0.084173	0.005595	15.0439	0.000000
ar4	0.000000	NA	NA	NA
ar5	0.049564	0.006042	8.2039	0.000000
ar6	0.000000	NA	NA	NA
ar7	0.000000	NA	NA	NA
ar8	0.000000	NA	NA	NA
ar9	0.000000	NA	NA	NA
ar10	0.000000	NA	NA	NA
ar11	0.000000	NA	NA	NA
ar12	0.000000	NA	NA	NA
ar13	0.000000	NA	NA	NA
ar14	0.000000	NA	NA	NA
ar15	0.000000	NA	NA	NA
ar16	0.000000	NA	NA	NA
ar17	0.000000	NA	NA	NA
ar18	0.000000	NA	NA	NA
ar19	0.000000	NA	NA	NA
ar20	0.000000	NA	NA	NA
ar21	0.017503	0.004311	4.0604	0.000049
ar22	-0.086652	0.004884	-17.7414	0.000000
ar23	-0.030771	0.004918	-6.2574	0.000000
ar24	0.736044	0.005188	141.8738	0.000000
ar25	-0.578518	0.007571	-76.4142	0.000000
ma1	0.144765	0.012517	11.5657	0.000000
ma2	0.148706	0.011226	13.2468	0.000000
ma3	0.068517	0.010499	6.5262	0.000000
ma4	0.041265	0.009892	4.1714	0.000030
omega	1975.722853	220.586869	8.9567	0.000000
alpha1	0.044928	0.002072	21.6872	0.000000
beta1	0.947624	0.002231	424.7622	0.000000

FIGURE 3.18 – Les coefficients du modèle *.

Le modèle ARMA(25,4)-GARCH(1,2) :

On a estimé le modèle ARMA(25,4)-GARCH(1,2) pour la série qui nous a donné les paramètres suivant :

Optimal Parameters				
	Estimate	Std. Error	t value	Pr(> t)
mu	0.000000	NA	NA	NA
ar1	0.731733	0.010604	69.0042	0.000000
ar2	0.000000	NA	NA	NA
ar3	0.084759	0.005587	15.1714	0.000000
ar4	0.000000	NA	NA	NA
ar5	0.049972	0.006037	8.2780	0.000000
ar6	0.000000	NA	NA	NA
ar7	0.000000	NA	NA	NA
ar8	0.000000	NA	NA	NA
ar9	0.000000	NA	NA	NA
ar10	0.000000	NA	NA	NA
ar11	0.000000	NA	NA	NA
ar12	0.000000	NA	NA	NA
ar13	0.000000	NA	NA	NA
ar14	0.000000	NA	NA	NA
ar15	0.000000	NA	NA	NA
ar16	0.000000	NA	NA	NA
ar17	0.000000	NA	NA	NA
ar18	0.000000	NA	NA	NA
ar19	0.000000	NA	NA	NA
ar20	0.000000	NA	NA	NA
ar21	0.017096	0.004309	3.9678	0.000073
ar22	-0.086011	0.004877	-17.6343	0.000000
ar23	-0.031307	0.004913	-6.3727	0.000000
ar24	0.736117	0.005175	142.2458	0.000000
ar25	-0.578210	0.007565	-76.4289	0.000000
ma1	0.144930	0.012563	11.5365	0.000000
ma2	0.147564	0.011174	13.2066	0.000000
ma3	0.068859	0.010460	6.5832	0.000000
ma4	0.040813	0.009845	4.1457	0.000034
omega	1971.811048	236.445354	8.3394	0.000000
alpha1	0.063948	0.006415	9.9680	0.000000
beta1	0.454622	0.003556	127.8467	0.000000
beta2	0.474800	0.003447	137.7270	0.000000

FIGURE 3.19 – Les coefficients du modèle *.

On a estimé d'autres modèles en modifiant p et q mais pas de significativité, d'où on s'est concentré sur les deux modèles précédents seulement.

3.3.2 Tests de validation

Le modèle ARMA(25,4)-GARCH(1,1) :

On a noté que les contraintes de non négativité sur les paramètres du modèle GARCH sont satisfaites. En outre, la condition $\alpha_1 + \beta_1 < 1$ est satisfaite aussi. Si le modèle est bien spécifié, alors $V(U_t) = 1$ tel que $U_t = \epsilon_t/\sigma_t$, le résidu standardisé, qui ne doit pas présenter d'effet de volatilité. Dans notre cas, $V(U_t) = 1.003509519$

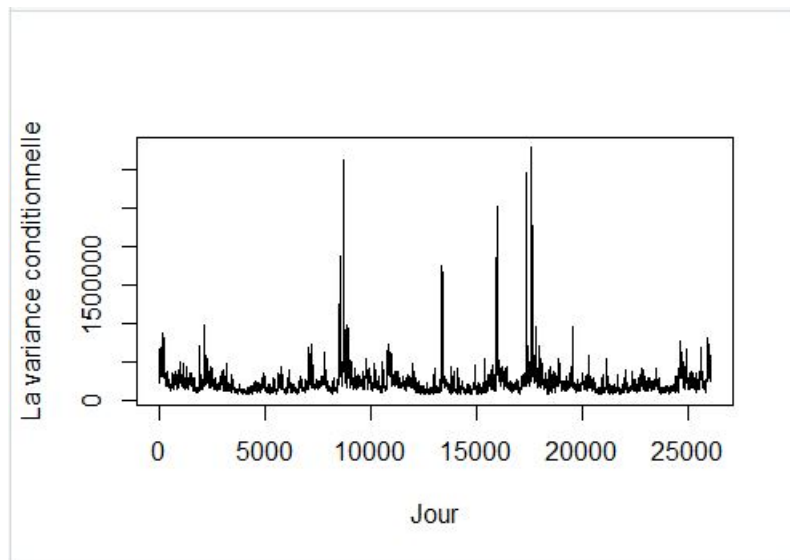


FIGURE 3.20 – L'évolution de la variance conditionnelle *.

Le modèle ARMA(25,4)-GARCH(1,2) :

On a remarqué que les contraintes de non négativité sur les coefficients du modèle GARCH sont satisfaites. Dans notre cas, $V(U_t) = 1.003738337$

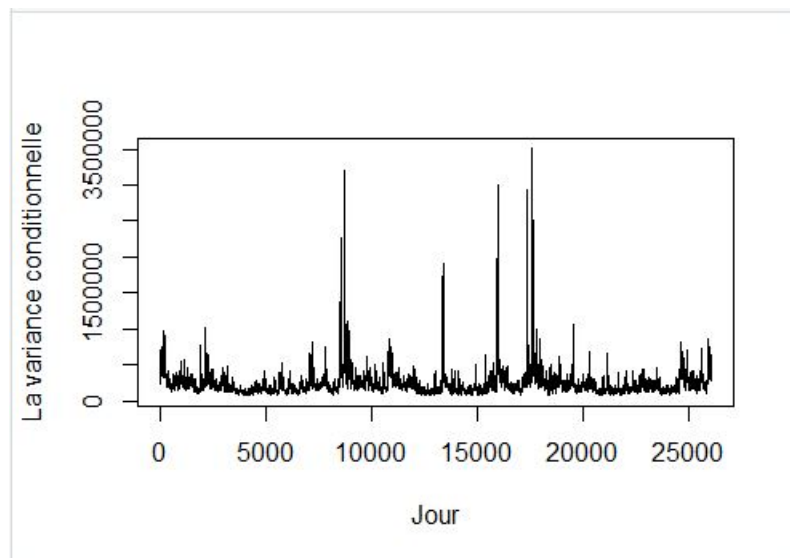


FIGURE 3.21 – L'évolution de la variance conditionnelle *.

On a conclu finalement que les deux modèles sont valides.

3.3.3 Choix du modèle

On a trouvé deux modèles valides mais il faut choisir un qui sera bien sûr le meilleur, ainsi quels sont les critères de sélection ?

On va baser la sélection sur les critères d'information qui représentent un arbitrage entre la précision du modèle et le nombre de paramètres à estimer. En effet, Le meilleur modèle est celui qui minimise le critère d'information retenu. Par exemple, le critère d'information Akaike (1969) ($AIC(p, q) = \ln(\hat{\sigma}_t^2) + \frac{2(p+q)}{T}$), le critère d'information BIC ou de Schwarz (1978) ($BIC(p, q) = SC(p, q) = \ln(\hat{\sigma}_t^2) + (p + q)\frac{\ln(T)}{T}$), le critère d'information de Hannan et Quinn (1979) ($HQ(p, q) = \ln(\hat{\sigma}_t^2) + (p + q)\frac{2\ln(\ln(T))}{T}$) etc...

Information Criteria

Akaike	15.104
Bayes	15.109
Shibata	15.104
Hannan-Quinn	15.106

Information Criteria

Akaike	15.102
Bayes	15.107
Shibata	15.102
Hannan-Quinn	15.104

FIGURE 3.22 – Les critères d'information pour ARMA(25,4)-GARCH(1,1) *. FIGURE 3.23 – Les critères d'information pour ARMA(25,4)-GARCH(1,2) *.

On a conclut de ce graphique que le meilleur modèle est ARMA(25,4)-GARCH(1,2), on va adopter ce modèle pour la partie suivante.

3.4 Le coût de la volatilité

Dans cette partie on a traité la relation entre la volatilité σ_t extraite du modèle et le coût d'ajustement. On a construit la régression linéaire du coût d'ajustement sur la volatilité, ce qui nous a donné le graphique suivant accompagné de la droite de la régression et son équation :

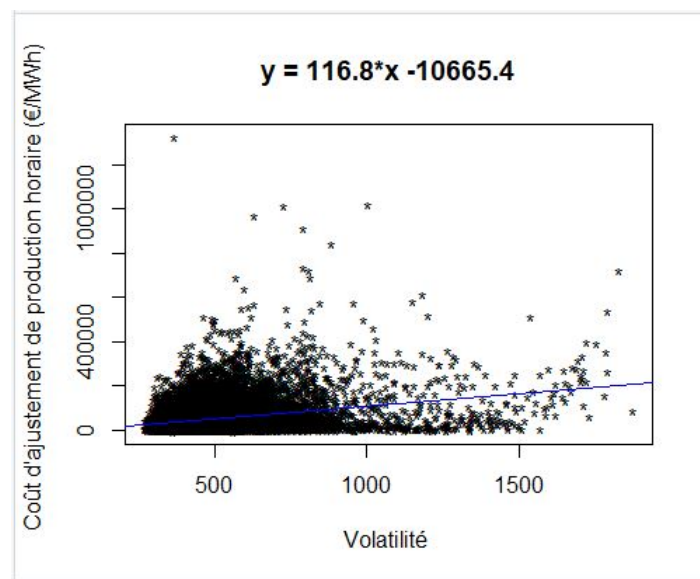


FIGURE 3.24 – Le coût d'ajustement de production en fonction de la volatilité *.

```
call:
lm(formula = Ah_new ~ garch@fit$sigma)

Residuals:
    Min       1Q   Median       3Q      Max
-167100.56 -28823.22 -13379.06  13338.74 1297096.09

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -10665.431226  1162.152184  -9.17731 < 0.00000000000000022 ***
garch@fit$sigma  116.814295    2.351592  49.67456 < 0.00000000000000022 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 53417.33 on 26073 degrees of freedom
Multiple R-squared:  0.08645806,    Adjusted R-squared:  0.08642302
F-statistic: 2467.561 on 1 and 26073 DF,  p-value: < 0.00000000000000022204
```

FIGURE 3.25 – Les paramètres descriptifs de la régression linéaire *.

On a constaté la significativité de résultats de la régression, les probabilités critiques sont inférieures aux risques de première espèce (1%, 5%, 10%) comme dans le cas du test statistique. Mais, on remarque que le coefficient de détermination R^2 est faible, qui peut refléter le pouvoir un peu faible d'explication de distribution des points de sorte que ça peut s'opposer à l'établissement de la relation linéaire simple mais elle reste la relation la plus proche de la réalité qu'on a achevé.

Conclusion

Mon travail au cours de ce stage d'été à PSL Dauphine a apporté de nouvelles contributions au Coût de la volatilité de la production intermittente sur les marchés day-ahead de l'électricité. Une grande recherche est faite pour la modélisation de l'erreur de prévision de la consommation nette où on a traité des modèles assez sophistiqués afin d'estimer après la volatilité attendue. Il s'agit normalement d'un stage de six mois, mais malgré la réduction de la période du stage pour l'adapter à ma disponibilité (2 mois et demi), j'ai obtenu des bons résultats grâce à ma détermination qui ouvrent la voie vers une autre étude sur les marchés intraday espérant être encore le réalisateur de cette étude dans le futur pour une nouvelle expérience.

Sur le plan individuel, j'ai eu l'opportunité que j'ai bien exploité, de me développer dans la programmation avec le langage R qui est dédié particulièrement aux études statistiques. Ce qui m'ouvre la voie vers le domaine de séries temporelles de sorte que je me suis familiarisé avec leurs techniques, surtout les séries représentant des observations économétriques. Ces séries sont caractérisées souvent par leur stationnarité qui oriente la recherche a posteriori vers la modélisation. Parmi les modèles rencontrés : le modèle autorégressif, moyenne mobile,... et le plus important leur combinaison, le modèle ARMA, qui sert souvent à modéliser les séries économétriques où on a eu l'occasion de me familiariser aussi avec les tests statistiques et leur mise en oeuvre avec R.

Personnellement, ce projet m'a permis d'apprendre beaucoup sur le domaine fascinant de la modélisation de séries temporelles dans l'économétrie et la finance. Je suis heureux d'avoir pu contribuer à la construction d'un modèle assez sophistiqué comme l'a décrit mon tuteur, le modèle ARMA-GARCH. Ce projet m'a également aidé à progresser dans le domaine de statistiques en particulier avec la modélisation et les tests statistiques et aussi à m'améliorer dans l'analyse, le raisonnement et surtout l'interprétation. Cet été à PSL Dauphine a été l'un des meilleurs expériences de ma vie et je suis reconnaissant à l'équipe qui m'a bien accueilli.

Bibliographie

- [1] Réseau de transport d'électricité. <https://data.rte-france.com>.
- [2] F.Karanfil et Y.Li. The role of continuous intraday electricity markets : The integration of large-share wind power generation in denmark. *The Energy Journal*, pages 107–130, 2017.
- [3] RTE. Consommation française d'électricité : Caractéristiques et méthode de prévision, Mai 2011.
- [4] Gui Wang Anupama Kowli et Ehsan Shaeepoorfard Sean Meyn, Matias Negrete-Pincetic. The value of volatile resources in electricity markets. *IEEE Decision and Control*, 2010.
- [5] Epex spot. <https://www.epexspot.com>.

Annexe A

Annexe A

La procédure du test pour le test ADF est la même que pour le test de Dickey-Fuller mais elle est appliquée au modèle :

$$\Delta x_t = \alpha + \beta t + \gamma x_{t-1} + \sigma_1 \Delta x_{t-1} + \dots + \sigma_{p-1} \Delta x_{t-p+1} + \epsilon_t$$

où α est une constante, β le coefficient sur une tendance temporelle et p l'ordre de décalage du processus autorégressif. Imposer les contraintes $\alpha = 0$ et $\beta = 0$ correspond à la modélisation d'une marche aléatoire et à l'aide de la contrainte $\beta = 0$ correspond à la modélisation d'une marche aléatoire avec une dérive. Par conséquent, il existe trois versions principales du test, analogues à celles examinées lors du test de Dickey-Fuller (test avec ni tendance et ni constante, test avec constante, test avec tendance temporelle et constante).

Egalement, le test DF-GLS est basé sur le standard dickey-fuller (1979) et il a la forme suivante :

$$\Delta x_t = \gamma x_{t-1} + \sigma_1 \Delta x_{t-1} + \dots + \sigma_{p-1} \Delta x_{t-p+1} + \epsilon_t$$

Ces deux tests sont ensuite effectués sous l'hypothèse nulle $H_0 : \gamma = 0$ qui suppose la non stationnarité de la série contre l'hypothèse alternative H_1 de $\gamma < 0$ qui suppose la stationnarité de la série. La statistique de test est la statistique de Student de $\hat{\gamma}$:

$$t_{\hat{\gamma}} = \frac{\hat{\gamma}}{\hat{\sigma}_{\hat{\gamma}}}$$

On rejette H_0 quand la statistique de test est plus petite que le seuil critique ou quand la probabilité critique est inférieure au risque de première espèce choisi (1%, 5% ou 10%).

Annexe B

Annexe B

Le modèle ARMA choisi est le suivant :

$$X_t = ar_1X_{t-1} + ar_3X_{t-3} + ar_5X_{t-5} + \sum_{i=21}^{25} ar_iX_{t-i} + \epsilon_t + \sum_{i=1}^4 ma_i\epsilon_{t-i}$$

ar1	ar2	ar3	ar4	ar5	ar6
0.72178632734	0.00000000000	0.09014965053	0.00000000000	0.06096481257	0.00000000000
ar7	ar8	ar9	ar10	ar11	ar12
0.00000000000	0.00000000000	0.00000000000	0.00000000000	0.00000000000	0.00000000000
ar13	ar14	ar15	ar16	ar17	ar18
0.00000000000	0.00000000000	0.00000000000	0.00000000000	0.00000000000	0.00000000000
ar19	ar20	ar21	ar22	ar23	ar24
0.00000000000	0.00000000000	0.01150978990	-0.08659363488	-0.03344665665	0.70156512979
ar25	ma1	ma2	ma3	ma4	intercept
-0.54409866680	0.19502672747	0.17569021020	0.09928678843	0.05555485774	0.00000000000

FIGURE B.1 – Les coefficients du modèle *.

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)
ar1	0.7217863273	0.0112740398	64.02198	< 0.000000000000000222 ***
ar3	0.0901496505	0.0061753655	14.59827	< 0.000000000000000222 ***
ar5	0.0609648126	0.0068580395	8.88954	< 0.000000000000000222 ***
ar21	0.0115097899	0.0045057105	2.55449	0.010634 *
ar22	-0.0865936349	0.0051536805	-16.80229	< 0.000000000000000222 ***
ar23	-0.0334466567	0.0052246690	-6.40168	0.00000000015368 ***
ar24	0.7015651298	0.0054491746	128.74705	< 0.000000000000000222 ***
ar25	-0.5440986668	0.0076922711	-70.73316	< 0.000000000000000222 ***
ma1	0.1950267275	0.0128786896	15.14337	< 0.000000000000000222 ***
ma2	0.1756902102	0.0119030857	14.76006	< 0.000000000000000222 ***
ma3	0.0992867884	0.0107478604	9.23782	< 0.000000000000000222 ***
ma4	0.0555548577	0.0103126477	5.38706	0.00000007161930 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

FIGURE B.2 – Les probabilités critiques associées aux coefficients *.