Analyse économétrique des prix d’électricité

—— Modélisation de la moyenne des log-rendements

Rapport de recherche

Août 2006

Nom de l’étudiant : Qingzhou YANG
Directeur de recherche : Nour MEDDAHI

Sommaire

Le but de ce projet est d'utiliser les modèles ARMA ou SARMA (ARMA avec saisonnalité) pour modéliser la moyenne conditionnelle des log-rendements du prix de l'électricité sur de différents marchés régionaux. Les études empiriques reconnaissent des saisonnalités sur les rendements du prix de l'électricité, donc il est nécessaire d'ajouter le facteur de saisonnalité dans la modélisation. Les log-rendements journaliers présentent une forte autocorrélation de 7 jours, cela signifie une saisonnalité hebdomadaire des rendements. Afin de justifier la saisonnalité exacte que chaque marché possède, on estime les modèles ARMA avec de différents ordres pour ces log-rendements journaliers, en incluant une saisonnalité, deux saisonnalités, ou aucune saisonnalité respectivement. Le niveau de saisonnalité choisi en estimant les modèles SARMA est toujours 7 ou son multiple, afin de correspondre à l'autocorrélation trouvée. Parfois on fait d'abord une régression MCO des log-rendements sur les variables binaires pour éliminer les impacts de mois ou de saisons, puis modéliser le résidu. À la fin, on choisit un modèle qui décrit mieux la dynamique de la moyenne conditionnelle de ces log-rendements. La méthode d'estimation pour les modèles ARMA et SARMA est le Maximum de vraisemblance (MLE). L'analyse se concentre d'abord sur un marché et présente en détail toutes les démarches à suivre, ensuite on donne simplement les résultats des autres marchés, et le cas échéant, des approches particulières, car la procédure est la même pour les autres.

Table des matières

1. Introduction

Dans ce rapport, on va étudier les prix de l'électricité sur six marchés régionaux, tels que le marché PJM (Pennsylvanie, New Jersey, Maryland), New England, New York – zone A, New York – zone G, New York – zone J et Ontario. L'objectif du projet est de trouver le meilleur modèle de la moyenne conditionnelle des log-rendements journaliers pour chacun de ces six marchés. Le prix de l'électricité est bien affecté par les conditions météorologiques, telles que la température, la sécheresse, la précipitation, etc. Par conséquent, le prix de l'électricité serait très volatile, ce qui motive la recherche d'un modèle pour la volatilité du prix. La demande de l'électricité dépend aussi de la saison. Par exemple, en été et en hiver, les prix connaissent des sauts à cause de l'augmentation de la consommation de l'électricité. Alors, un plan de périodicité parmi les mois ou les saisons serait aussi un sujet à l'étude. L'importance de modéliser la moyenne conditionnelle est de connaître la tendance du mouvement de la série temporelle, par ailleurs, ayant le modèle pour la moyenne, on peut chercher à modéliser la volatilité à partir du résidu. Les modèles de la moyenne à estimer sont ARMA et SARMA (Seasonal ARMA), et la méthode d'estimation est le Maximum de vraisemblance (MLE).

Les étapes suivies sont:

1) Revue des études antérieures sur les caractéristiques du prix de l'électricité et la modélisation;
2) Présentation des marchés à traiter ainsi que leurs données;
3) Modèles et méthode d'estimation;
4) Résultats empiriques.

Les études antérieures montrent de différentes saisonnalités sur les prix de l'électricité, notamment la saisonnalité hebdomadaire. Dans notre projet, tous les six marchés au sujet montrent une autocorrélation de 7 jours sur les log-rendements journaliers, cela confirme la saisonnalité hebdomadaire reconnue dans les littératures antérieures. Selon l'approche habituelle, on essaie d'estimer les modèles avec une saisonnalité et ceux avec deux saisonnalités dans la modélisation SARMA. Prioritairement on choisit toujours la saisonnalité de 7 jours, si l'on ne trouve pas de bons modèles, on change la deuxième saisonnalité en multiple de 7 jours, comme 14, 21, 28, etc., jusqu'à ce que l'on trouve de bons modèles. L'analyse empirique se concentre principalement sur le marché PJM. On fait premièrement la régression MCO des log-rendements journaliers sur les variables binaires. Ensuite on estime les modèles ARMA ou SARMA pour le résidu, afin d'éliminer la périodicité de saisons. Enfin on estime le modèle ARMA(6,6) avec une saisonnalité de 7 jours. Autrement dit, le prix du marché PJM comporte une périodicité parmi les saisons, ainsi qu'une saisonnalité hebdomadaire. Par ailleurs, le résultat du marché Ontario est un peu différent. On régresse aussi les log-rendements sur les variables binaires afin de supprimer la périodicité de saisons, pourtant, en plus d'une première saisonnalité de 7 jours, le meilleur modèle consiste en deuxième saisonnalité de 28 jours, celle-ci correspond à la saisonnalité mensuelle. Alors, le prix du marché Ontario a une périodicité parmi les saisons, une saisonnalité hebdomadaire, et aussi une saisonnalité mensuelle. Puisque l'on peut modéliser les autres marchés avec la même méthodologie, on va donner simplement leurs résultats de l'estimation et indiquer les meilleurs modèles choisis.

Toutes les procédures sont développées en Matlab, notamment la programmation du Maximum de vraisemblance.

2. Revues de la littérature

2.1. Werner (2002)

Werner (2002) constate l'existence de la saisonnalité sur les prix de l'électricité. Par un graphique, il montre un plan de saisonnalité pour les prix moyens de Nord Pool aux différentes heures pendant une journée, appelé '*intra-day*'. Nord Pool est un marché de l'électricité qui couvre la Norvège, la Suède, la Finlande et le Danemark. Il conclut que le prix minimal se trouve entre 3 heures et 4 heures du matin, et que le prix maximal se trouve entre 9 heures et 10 heures du matin. De la même manière, il présente le plus haut niveau du prix le mardi et une forte diminution le vendredi, le samedi ainsi que le dimanche pour les prix moyens de Nord Pool à chaque jour pendant une semaine, toutes les semaines suivent presque le même plan, on appelle cette saisonnalité '*intra-week*'. On observe le même phénomène pour EEX (ou bien LPX), les prix moyens du lundi au vendredi sont supérieurs aux prix au week-end. Aussi, en terme d'une année, les prix de Nord Pool en hiver sont un peu plus élevés que les prix en été, et les prix futures de COB (California-Oregon Border) en été sont plus élevés que ceux en hiver. On parle de la saisonnalité '*intra-year*'.

Werner (2002) dit que le plan de la saisonnalité est souvent associé avec la variation mensuelle en parlant de la saisonnalité '*intra-year*', donc en plus du plan de la saisonnalité parmi les saisons, '*intra-year*' pourrait inclure la périodicité parmi les mois. Alors on distingue la saisonnalité par la saisonnalité des heures (*intra-day*), des jours (*intra-week*), des mois et des saisons (*intra-year*). Dans ce projet, en utilisant les prix journaliers, on ne considère pas la saisonnalité des heures, mais on devrait prendre en considération la saisonnalité des jours, ainsi que la périodicité parmi les mois et les saisons.

2.2. Francq, Roy et Zakoïan (2005)

À la fin de l'estimation du modèle ARMA ou SARMA, il faut faire le test Box-Pierce pour examiner si les résidus sont les bruits blancs. La statistique de test est donnée par

$$Q = T \sum_{k=1}^{m} \rho_k^{2}$$

où ρ_k est l'estimateur de l'autocorrélation du résidu à l'ordre k, T est la taille de l'échantillon, m est l'ordre du test. Si le résidu est un bruit blanc, Q suit une loi asymptotique de χ^2_{m-h}, où h est le nombre de paramètres, la constante n'étant pas incluse. Comparons la valeur de Q avec la valeur critique de la distribution χ^2 à (m – h) degrés de liberté, si Q est plus grand que cette valeur critique, on dit que le résultat du test est significatif. A condition que les résultats du test pour tous les ordres soient non significatifs, le résidu est un bruit blanc, on accepte le modèle ; sinon, le résidu n'est pas un bruit blanc et l'on rejette le modèle.

Selon cet article, pour un modèle avec h paramètres, ne comprenant pas la constante, la distribution asymptotique de Q est une loi χ^2 avec m-h degrés de liberté, à condition que la taille de l'échantillon et l'ordre du test soient grands. Autrement dit, pour un bon modèle ARMA ou SARMA, il est possible que les résultats du test Box-Pierce de ce modèle soient significatifs aux ordres les plus petits, mais qu'ils ne soient significatifs à aucun des ordres plus élevés. Par exemple, en choisissant des multiples de 7 comme ordres du test, et 5% comme niveau de significativité, même si les valeurs de Q sont plus grandes que les valeurs critiques de la loi χ^2 aux ordre 7, 14, 21, voir 28, il est recommandé d'examiner les ordres 35, 42, 49, si les résultats sont moins grands que les valeurs critiques correspondantes, il est probable que le modèle passe le test.

Cet article constate aussi que la version standard du test Box-Pierce n'est pas pertinente pour le modèle ARMA avec le résidu non-corrélé mais pas indépendant. L'absence d'indépendance du résidu est causée par la non-linéarité des données, en faisant la représentation linéaire pour les données non-linéaires, les résidus devraient être dépendants l'un de l'autre, alors seraient des bruits blancs faibles au lieu des bruits blancs. Par contre, le test Box-Pierce est basé sur l'indépendance des données, il se sert à tester si les résidus sont des bruits blancs. Puisque les bruits blancs faibles ne sont pas indépendants, en donnant 5% comme niveau de significativité, la valeur critique de la loi χ^2 serait différent de 5%. Cet article propose aussi une version modifiée du test Box-Pierce, de la manière que l'on puisse pendre la valeur critique de 5% en faisant le test avec le niveau de significativité qui est égale à 5%. Dans notre projet, on utilise quand même la version standard du test et 5% comme valeur critique pour simplifier.

3. Présentation des marchés et des données

Les six marchés en question sont les marchés de l'électricité en Amérique du nord. Ce sont des marchés pour les fournisseurs où on achète et vend les produits en gros. L'électricité est échangée sur ces marchés comme titres financiers, il existe même des produits dérivés tels que les *futures*, les *forwards* et les options. Dans ce rapport, on se concentre sur les prix courants. Ces marchés sont régionaux du fait de la contrainte de transmission, il est impossible de trouver un modèle général pour tous, donc on doit les développer pour chacun. Les marchés de l'électricité sont ouverts vingt-quatre heures par jour et sept jours par semaine, ils ne sont pas fermés au week-end, alors dans la modélisation, il vaut mieux faire attention à la saisonnalité de sept jours car un plan de périodicité hebdomadaire est probablement présent. De plus, le prix de l'électricité pourrait être bien affecté par la saison ou le mois, car en été et en hiver il pourrait avoir une tendance de sauts sur la consommation de l'électricité, donc on devrait prendre en considération la saisonnalité parmi les mois et les saisons.

Les données originales sont les prix courants heure par heure à chaque jour. On distingue une journée par deux périodes de temps: l'heure de pointe et l'heure hors pointe. L'heure de pointe est la période où on utilise beaucoup l'électricité, qui est de 8:00 à 23:00. L'heure hors pointe est la période où la consommation d'électricité est relativement faible, qui est de 0:00 jusqu'à 7:00. Il y aurait des différences significatives entre le prix à l'heure de pointe et celui à l'heure hors pointe. Pour ce rapport, choisissons premièrement les prix midi de chaque jour, qui se trouve en période de pointe, notés par P, puis divisons le prix midi du jour (P_t) par le prix midi un jour avant (P_{t-1}) pour avoir le rendement brut journalier, noté par R_t , ensuite prenons le logarithme naturel de R_t comme log-rendement du jour, noté par r_t. Soit

$$R_t = P_t / P_{t-1}$$

$$r_t = \log(R_t)$$

On va alors essayer de modéliser r_t .

4. Modèles et méthode d'estimation

4.1. Modèles

4.1.1 ARMA

Le modèle ARMA (Autoregressive Moving Average) sert à décrire la moyenne conditionnelle. Un ARMA(p,q) de la série temporelle r_t s'écrit comme :

$$r_t = c + \sum_{i=1}^{p} a_i r_{t-i} + \sum_{j=1}^{q} b_j \eta_{t-j} + \eta_t$$

où η_t est un bruit blanc ;

$\{a_i\}$ sont les coefficients de l'ordre AR ;

$\{b_j\}$ sont les coefficients de l'ordre MA ;

c est la constante ;

r_{t-i} est la série r_t au retard i ;

η_{t-j} est la série η_t au retard j .

4.1.2. SARMA

Si la saisonnalité de s jours se présente, on pourrait estimer le modèle SARMA (Seasonal ARMA, ou ARMA avec saisonnalité) comme dynamique de la moyenne conditionnelle. Pour trouver un bon SARMA, il vaut mieux essayer d'éliminer une ou deux fois la saisonnalité avant l'estimation, puis choisir le meilleur parmi les modèles estimés.
Éliminer une fois la saisonnalité donne un ARMA(p,q) avec une saisonnalité :

$$y_t = r_t - x * r_{t-s}$$

$$y_t = c + \sum_{i=1}^{p} a_i y_{t-i} + \sum_{j=1}^{q} b_j \eta_{t-j} + \eta_t$$

Éliminer deux fois la saisonnalité donne un ARMA(p,q) avec deux saisonnalités :

$$y_t = r_t - x1 * r_{t-s}$$

$$z_t = y_t - x2 * y_{t-s}$$

$$z_t = c + \sum_{i=1}^{p} a_i z_{t-i} + \sum_{j=1}^{q} b_j \eta_{t-j} + \eta_t$$

où η_t est le bruit blanc ;

x , $x1$ et $x2$ sont les paramètres de saisonnalité ;

r_{t-s} est la série r_t au retard s ;

y_{t-i} , y_{t-s} est la série y_t au retard i ,s respectivement ;

z_{t-i} est la série η_t au retard i ;

η_{t-j} est la série η_t au retard j .

4.1.3. Régression MCO sur les variables binaires

Pour éliminer la saisonnalité parmi les mois, on introduit des variables binaires qui prennent la valeur de 1 ou 0. Si une certaine variable prend la valeur de 1, cela représente que l'on est en un certain mois correspondant ; sinon, on est en un autre mois. Il en est de même pour éliminer la périodicité parmi les saisons. L'approche est de faire d'abord la régression MCO des log-rendements (r_t) sur ces variables binaires, ensuite trouver le meilleur ARMA ou SARMA pour le résidu MCO. Voyons un résumé de ces deux régressions au-dessous, l'un est sur les mois, l'autre est sur les saisons :

1) Régresser r_t par MCO sur 11 variables binaires qui signifient respectivement janvier, février, … jusqu'à novembre.

$$r_t = a + b1 * jan + b2 * feb + b3 * mar + b4 * apr + b5 * may + b6 * jun + b7 * jul + b8 * aug + b9 * sep + b10 * oct + b11 * nov + u_t$$

où jan = 1 si on est en janvier, jan = 0 sinon ;

feb = 1 si on est en février, feb = 0 sinon ;

les même notations pour les autres variables.

u_t est le résidu.

Si toutes les variables prennent la valeur de 0, cela signifie qu'on est en décembre. On inclut seulement 11 variables ici pour éviter la multicolinéarité.

2) Pour la même raison, en faisant la régression MCO de r_t sur les saisons, on inclut seulement 3 variables binaires. Par exemple, on met l'hiver comme groupe de base, et fait la régression ci-dessous :

$$r_t = c + d1 * spring + d2 * summer + b3 * autumn + v_t$$

où spring = 1 si on est au printemps, spring = 0 sinon ;

summer = 1 si on est en été, summer = 0 sinon ;

autumn = 1 si on est en automne, autumn = 0 sinon.

Si spring = summer = autumn = 0, cela signifie qu'on est en hiver.

v_t est le résidu.

4.2. Méthode d'estimation

4.2.1 Bref aperçu

Pour trouver le meilleur ARMA ou SARMA comme description de la moyenne conditionnelle, il faut estimer plusieurs modèles et en choisir un. Essayons premièrement d'estimer les ARMA de l'ordre 1 jusqu'à l'ordre 5, soit ARMA(1,1), ARMA(2,2), ARMA(3,3), ARMA(4,4) et ARMA(5,5) pour les log-rendements sans considérer la saisonnalité de jours, puis estimons les ARMA des mêmes ordres pour les log-rendements mais avec une saisonnalité et avec deux saisonnalités. En dessinant le graphique de la fonction des autocorrélations (ACF) des log-rendements, si on voit des autocorrélations significatives à un retard qui est égale à un multiple de s, il pourrait y avoir une saisonnalité de s jours. Alors dans la modélisation SARMA, prenons la saisonnalité de s jours, autrement dit, estimons :

ARMA(1,1) avec une saisonnalité de s, ARMA(1,1) avec deux saisonnalités de s,
ARMA(2,2) avec une saisonnalité de s, ARMA(2,2) avec deux saisonnalités de s,
ARMA(3,3) avec une saisonnalité de s, ARMA(3,3) avec deux saisonnalités de s,
ARMA(4,4) avec une saisonnalité de s, ARMA(4,4) avec deux saisonnalités de s,
ARMA(5,5) avec une saisonnalité de s, ARMA(5,5) avec deux saisonnalités de s.

Avec les 5 modèles ARMA sans saisonnalité, on a enfin 15 modèles pour les log-rendements.

De la même manière, après avoir régressé par MCO les log-rendements sur les variables binaires de mois puis tiré les résidus, on estime les 15 modèles pour les résidus. Et on estime aussi les 15 modèles pour les résidus tirés de la régression MCO des log-rendements sur les variables binaires de saisons. Enfin, on a 45 modèles estimés pour chaque marché, ensuite on examine les statistiques et fait les tests afin de voir si l'ordre 5 est suffisant, sinon, on va augmenter l'ordre de ARMA ou SARMA pour trouver le meilleur modèle.

4.2.2 Hypothèses

On suppose la normalité et l'homoscédasticité dans la modélisation ARMA et SARMA, alors le résidu η_t suit une loi normale, et la log-vraisemblance est donnée par

$$\ell = -T/2 * \log(2\pi\sigma^2) - 1/(2\sigma^2) * \sum_{t=1}^{T} \eta^2_t$$

où σ^2 est la variance de η_t, suivi de l'hypothèse homoscédastique.

T est la taille de l'échantillon.

4.2.3 Fonction de vraisemblance

La méthode de maximum de vraisemblance (MLE) consiste à maximiser le logarithme de la fonction de vraisemblance du modèle.

Pour estimer un ARMA(p,q)

$$r_t = c + \sum_{i=1}^{p} a_i r_{t-i} + \sum_{j=1}^{q} b_j \eta_{t-j} + \eta_t$$

Premièrement exprimons η_t par

$$\eta_t = r_t - c - \sum_{i=1}^{p} a_i r_{t-i} - \sum_{j=1}^{q} b_j \eta_{t-j}$$

en mettant $\{a_i\}$, $\{b_i\}$ et c comme paramètres. Si r_{t-i} se trouve à l'extérieur de l'échantillon, on lui donne la valeur qui est égale à la moyenne inconditionnelle de r_t, en raison que l'espérance des log-rendements est estimée par leur moyenne inconditionnelle. Par contre, si η_{t-j} se trouve hors l'échantillon, on lui donne la valeur nulle, car l'espérance de résidu est égale à zéro.

À propos de SARMA, on programme les équations désaisonnalisées avant de donner l'expression de η_t.

Pour un ARMA(p,q) avec une saisonnalité :

$$y_t = r_t - x * r_{t-s}$$

$$\eta_t = y_t - c - \sum_{i=1}^{p} a_i y_{t-i} - \sum_{j=1}^{q} b_j \eta_{t-j}$$

Pour un ARMA(p,q) avec deux saisonnalités :

$$y_t = r_t - x1 * r_{t-s}$$

$$z_t = y_t - x2 * y_{t-s}$$

$$\eta_t = z_t - c - \sum_{i=1}^{p} a_i z_{t-i} - \sum_{j=1}^{q} b_j \eta_{t-j}$$

on prend aussi la moyenne inconditionnelle de r_t comme valeur de r_{t-s} qui se trouve hors l'échantillon, il en est de même pour y_{t-i}, y_{t-s} et z_{t-i}. En cas du modèle SARMA, on ajoute x ou bien $x1$ et $x2$ comme paramètres de la fonction de log-vraisemblance.

Si l'on fait la régression MCO des log-rendements et puis estime des ARMA ou SARMA du résidu (noté par U), on donne la valeur nulle aux retards de U qui sont à l'extérieur de l'échantillon, car l'espérance de U est zéro.

Ayant exprimé η_t dans la programmation, écrivons donc sa log-vraisemblance. Puisque Matlab ne donne que la fonction de minimisation, on devrait multiplier la formule de log-vraisemblance par –1, de la manière que l'on minimise

$$T/2 \; * \; \log(2\pi\sigma^2) \; + \; 1/(2\sigma^2) \; * \; \sum_{t=1}^{T} \eta^2_t$$

où T est le nombre d'observations du vecteur des log-rendements. La valeur de σ^2 est donnée par la variance de η_t, soit $\mathrm{var}(\eta_t)$, en raison de l'hypothèse d'homoscédasticité. Minimiser la formule ci-dessus est équivalent à maximiser la log-vraisemblance. Après avoir obtenu la valeur de minimisation, notée par **VAL**, on prend son nombre opposé, soit **–VAL**, comme valeur maximale de log-vraisemblance. Le vecteur de paramètres qui donne **VAL** est retenu comme l'estimateur des coefficients.

4.2.4 Choix des valeurs initiales

On a besoin des valeurs initiales des paramètres pour exécuter la routine de minimisation sous Matlab. Le choix des valeurs initiales est important. Étant donné les valeurs initiales des paramètres, le programme va calculer la valeur de la fonction objectif et chercher à donner le point du minimum local qui est le plus proche de cette valeur. Pourtant, la méthode MLE demande le maximum global, et donc, on devrait essayer plusieurs combinaisons de valeurs initiales et minimiser la fonction à partir de ces valeurs, de sorte d'obtenir quelques minimums locaux, à la fin on compare ces minimums locaux et choisit celui qui est le plus petit.

En estimant le modèle ARMA(1,1), utilisons zéro comme valeur initiale de tous les coefficients. À l'ordre plus élevé, prenons les coefficients estimés du modèle avec un ordre moins grand comme valeur initiale des coefficients de même ordre, zéro pour les coefficients des ordres ajoutés. Prenons un ARMA(3,3) comme exemple : pour les coefficients de AR(1), AR(2), MA(1), MA(2) et la constante, mettons leurs valeurs initiales qui sont égales à leurs estimateurs dans ARMA(2,2) ; pour les coefficients de AR(3) et MA(3), leurs valeurs initiales sont zéros. En estimant un ARMA(1,1) avec une saisonnalité, prenons les coefficients estimés du modèle ARMA(1,1) sans saisonnalité comme valeur initiale des coefficients de même ordre et zéro pour le paramètre de saisonnalité, soit x. En estimant un ARMA(3,3) avec une saisonnalité, on pourrait prendre les coefficients estimés du modèle ARMA(2,2) avec une saisonnalité comme valeur initiale des coefficients de même ordre ainsi que le paramètre de saisonnalité, zéro pour les autres, on pourrait aussi prendre les coefficients estimés du modèle ARMA(3,3) sans saisonnalité comme valeur initiale des coefficients de même ordre et zéro pour le paramètre de saisonnalité. Il est de même logique pour les modèles ARMA avec deux saisonnalités, à l'exception de prendre l'estimateur du paramètre de saisonnalité dans un ARMA avec une saisonnalité, soit x, comme valeur initiale du paramètre de la première saisonnalité dans un ARMA avec deux saisonnalités, soit x1 ; et de prendre zéro comme valeur initiale du paramètre de la deuxième saisonnalité, soit x2.

Puisqu'il vaut mieux essayer plusieurs combinaisons de valeurs initiales, on pourrait en choisir quelques unes à l'aide de Matlab et Stata. En Matlab, il y a une fonction qui peut modéliser la moyenne conditionnelle et la volatilité de la série temporelle. En Stata, on peut aussi estimer ARMA sans saisonnalité en utilisant le maximum de vraisemblance conditionnel et non-conditionnel. On exécute la routine de ces fonctions avec la série d'intérêt pour avoir de différents estimateurs des coefficients, puis essaie chaque estimateur comme valeurs initiales des coefficients dans un ARMA sans saisonnalité ainsi que celles des coefficients de même ordre dans un ARMA avec une saisonnalité (zéro pour le paramètre de saisonnalité). Comme ces valeurs initiales donnés par Matlab et Stata sont déjà les estimateurs des coefficients pour la moyenne conditionnelle et la volatilité, elles sont les valeurs les plus proches des vrais coefficients, alors à partir de ces estimateurs, on aurait une grande probabilité de trouver le minimum global en exécutant la routine de minimisation.

4.3. Statistiques

4.3.1. Log-likelihood

On utilise 'Log-likelihood' représentant la valeur de log-vraisemblance du modèle estimé. Le modèle à l'ordre plus élevé doit avoir une valeur de log-vraisemblance qui est plus grande que celle du modèle à l'ordre moins élevé. Le modèle à l'ordre moins élevé pourrait être considéré comme modèle avec plus de retards à condition que les coefficients de ces retards extra soient nuls, puisque l'on cherche le maximum de log-vraisemblance du modèle à l'ordre plus élevé, sa valeur doit dominer toutes les autres valeurs de log-vraisemblance du modèle de même ordre, incluant le modèle avec des retards ayant des coefficients nuls.

4.3.2. Test Box-Pierce

Une fois que l'on a estimé un ARMA ou un SARMA, on va examiner si le résidu η_t est un bruit blanc. Pour cela, on utilise le test Box-Pierce. La formule est donnée par

$$Q = T \sum_{k=1}^{m} \rho_k^2$$

Les marchés de l'électricité sont ouverts 7 jours par semaine, alors on devrait se concentrer particulièrement sur l'autocorrélation de 7 jours en faisant le test. Par conséquent, on choisit m qui est égale au multiplicateur de 7 dans la formule du Test Box-Pierce, et calcule la somme des autocorrélations au carré de l'ordre 1 jusqu'à l'ordre 7, 14, 21, 28, 35, 42 et 49 respectivement, ensuite multiplie chaque somme par le nombre d'observations des log-rendements, enfin on examine si ses résultats sont moins grands que la valeur critique de χ^2 avec le degré de liberté qui est 7-h, 14-h, 21-h, 28-h, 35-h, 42-h et 49-h respectivement, où h est la somme de l'ordre de AR, l'ordre de MA et le nombre de paramètres de saisonnalité s'il y en a. Théoriquement, si l'on ne rejette pas l'hypothèse nulle du test pour tous les m choisis, le modèle passe le test Box-Pierce, alors il s'agit d'un bon modèle, sinon, le modèle n'est pas bon.

Habituellement, on prend l'intervalle de confiance de 95%, si la valeur de Q est supérieure à la valeur critique de χ^2 avec (m – h) degrés de liberté au niveau 5%, on ne passe pas le test. Autrement dit, si p-value de Q avec (m – h) degrés de liberté est plus grand que 0.05, on passe le test.

4.3.3 AIC et BIC

S'il y a plusieurs bons modèles qui servent à décrire la moyenne d'une série, on va en choisir le meilleur. Alors, on va calculer le Critère d'information d'Akaike (AIC) et le Critère d'information de Bayes (BIC), le modèle qui minimise AIC et BIC sera le meilleur modèle. Leurs formules sont donnés par

$$AIC = \log\left(\sum_{t=1}^{T} \eta_t^2 / T \right) + (1 + h) * 2 / T$$

$$BIC = \log\left(\sum_{t=1}^{T} \eta_t^2 / T \right) + (1 + h) * \log T / T$$

où T est la taille de l'échantillon, h est la somme de l'ordre AR, l'ordre de MA et le nombre de paramètres de saisonnalité s'il y en a.

4.4. Tests des paramètres

Ayant l'estimateur des paramètres, il faut tester s'il est significatif ou non. Si certains paramètres ne sont pas significatifs, le modèle devrait être réduit à l'ordre moins élevé. Il y a deux façons pour faire le test : l'un est le test *t*, l'autre est le test pour le ratio de vraisemblance (Likelihood Ratio Test, ou test LR). Le test *t* est pour tester la significativité d'un paramètre unique, alors que le test LR est le test conjoint de tous les paramètres.

4.4.1. Test *t*

Soit θ_k l'estimateur d'un paramètre θ^0 et σ^2 la variance de θ_k, la formule du test *t* est

$$t = (\theta_k - \theta^0)/\sigma,$$

t suit la loi $N(0,1)$. Si σ est estimé, $(\theta_k - \theta^0)/\sigma$ suit la distribution *t* avec $(T - h - 1)$ degrés de liberté, où h est le nombre de paramètres, la constante n'est pas comprise. Lorsque la taille de l'échantillon tend vers l'infini, la valeur critique de la distribution *t* au niveau 5% est 1.96, qui est égale à la valeur critique au niveau 5% de la loi $N(0,1)$. Quant au test de significativité, $\theta^0 = 0$, alors on a

$$t = (\theta_k - \theta^0)/\sigma = \theta_k/\sigma.$$

Si $|\theta_k/\sigma| > 1.96$, on dit que θ_k est statistiquement différent de zéro, ou bien θ_k est significatif ; sinon, θ_k n'est pas significatif. Le meilleur modèle devrait avoir plus de paramètres estimés qui sont significatifs en comparaison des autres modèles. Ce qui nous reste est d'estimer la variance de l'estimateur des paramètres.

Gouriéroux (1997, 44) montre la méthode de calculer approximativement la matrice de covariance pour l'estimateur de MLE :

$$V_{as}[T^{1/2}(\hat{\theta} - \theta)] = J^{-1}\, I\, J^{-1} ,$$

$$\text{où } J = E_0[- \partial^2 \log \ell_t (Y;\theta) / \partial\theta\partial\theta']$$

$$I = E_0\{[\partial \log \ell_t (Y;\theta) / \partial\theta]\, [\partial \log \ell_t (Y;\theta) / \partial\theta']\}$$

ℓ_t représente la vraisemblance à une observation, θ est le vecteur de paramètres, $\hat{\theta}$ est l'estimateur de θ, Y est l'échantillon, T est la taille de l'échantillon, V_{as} signifie la variance asymptotique, et E_0 indique l'espérance sous la vraie distribution. Si la vraie distribution est normale, on a $I = J$, alors on pourrait simplifier la formule de variance asymptotique par :

$$V_{as}[T^{1/2}(\hat{\theta} - \theta)] = J^{-1} = I^{-1} .$$

En Matlab, il est plus facile de programmer la dérivé du premier ordre que la dérivé du second ordre, donc on préfère I^{-1} comme matrice de covariance pour la multiplication de la différence entre θ et son estimateur par la racine de la taille de l'échantillon, soit

$$V_{as}\,[T^{1/2}(\hat{\theta} - \theta)] = I^{-1}$$

$$\rightarrow T * V_{as}\,[(\hat{\theta} - \theta)] = I^{-1}$$

$$\rightarrow V_{as}\,[(\hat{\theta} - \theta)] = (\, T * I\,)^{-1}$$

$$\rightarrow V_{as}\,[\hat{\theta}] = (\, T * I\,)^{-1}$$

$$\rightarrow V_{as}\,[\hat{\theta}] = (\, \sum_{t=1}^{T}\{[\partial \log \ell_t (Y;\theta) / \partial\theta]\, [\partial \log \ell_t (Y;\theta) / \partial\theta']\}\,)^{-1}$$

On obtient l'expression analytique de la matrice de covariance pour l'estimateur des coefficients. Prenant les éléments de la diagonale principale sous la forme d'un vecteur, on obtient alors le vecteur de variance pour les paramètres estimés, dont chaque élément correspond à la variance de l'élément au même rang dans le vecteur de l'estimateur des paramètres.

4.4.2. Test pour le ratio de vraisemblance (Test LR)

Le test LR est basé sur la différence de la valeur de log-vraisemblance entre le modèle contraint et le modèle non-contraint. Puisque MLE maximise le logarithme de la fonction de vraisemblance, le modèle contraint jette les variables en comparaison du modèle non-contraint, donc donne la valeur de log-vraisemblance plus petite que celle-ci du modèle non-contraint. Le test est dans le sens de voir si la diminution de cette valeur est assez grande pour que les variables jetées soient importantes.[1] Parmi les modèles ARMA et SARMA qui passent le test Box-Pierce, on va décider si le modèle à l'ordre plus élevé est meilleur que celui à l'ordre moins élevé. En prenant le modèle à l'ordre plus élevé comme modèle non-contraint, on pourrait considérer le modèle à l'ordre moins élevé comme modèle contraint, où la contrainte est que les paramètres ajoutés dans le modèle à l'ordre plus élevé sont nuls. Ces paramètres ajoutés peuvent être les coefficients de l'ordre AR et l'ordre MA, aussi ils peuvent être les paramètres de saisonnalité. La formule du test est donnée par

$$\text{LR} = 2(\mathcal{L}_{ur} - \mathcal{L}_r),$$

où $\mathcal{L}_{ur}$ est la valeur de log-vraisemblance du modèle non-contraint, et $\mathcal{L}_r$ est la valeur de log-vraisemblance du modèle contraint. Sous l'hypothèse nulle, soit les paramètres ajoutés dans le modèle non-contraint sont égales à zéro, le résultat de LR suit approximativement une distribution chi-carré avec q degrés de liberté, où q est le nombre de paramètres impliqués par l'hypothèse nulle. Alors, si LR > χ^2_q , où χ^2_q est la valeur critique au niveau 5% de la distribution chi-carré avec q degrés de liberté, on refuse l'hypothèse nulle, donc les paramètres jetés par le modèle contraint sont importants, et le modèle ARMA ou SARMA à l'ordre plus élevé est meilleur que le modèle à l'ordre moins élevé. En revanche, si LR < χ^2_q , les paramètres jetés par le modèle contraint n'ont pas d'importance, on choisit le modèle à l'ordre plus petit.

[1] Wooldridge (2003, 559).

5. Résultats empiriques

On donne plus de détails à l'analyse du marché PJM, tels que l'examen de la saisonnalité, des résultats du test Box-Pierce, le test *t* ainsi que le test LR. On va montrer les tableaux des statistiques pour tous les 45 modèles estimés, puis faire le test *t* seulement pour les bons modèles, ensuite on va augmenter l'ordre du modèle et faire le test LR à la fin de l'estimation pour voir s'il faut passer au modèle à l'ordre plus grand. Le modèle du marché Ontario devrait être plus compliqué à cause de la saisonnalité particulière, on va essayer de trouver le meilleur choix de la saisonnalité en modélisation, et enfin ne présenter que les modèles représentatifs. Pour les quatre autres marchés, on va présenter simplement les bons modèles et donner la conclusion.

5.1. PJM (Pennsylvanie, New Jersey, Maryland)

5.1.1 Examen des statistiques

En dessinant la graphique ACF des log-rendements journaliers, voir Graphique 1, on montre une autocorrélation significative tous les 7 jours, donc il pourrait y avoir une saisonnalité de 7 jours. Cette conclusion correspond à 7 jours d'ouverture par semaine du marché de l'électricité. Donc en modélisant SARMA, on essaye des ARMA avec une saisonnalité de 7 et des ARMA avec deux saisonnalités de 7. Tableau 1 donne les statistiques des 15 modèles estimés pour les log-rendements, sans faire la régression MCO. Il comprend trois petites tables : la première table intitulée 'ARMA pour r_t' veut dire ARMA sans saisonnalité, 'r_t' représente le log-rendement ; le deuxième table intitulée 'ARMA pour $y_t = r_t - x * r_{t-7}$' veut dire ARMA avec une saisonnalité, où 'r_{t-7}' est 'r_t' au retard 7, x est le paramètre de saisonnalité ; le troisième table intitulée 'ARMA pour $z_t = y_t - x2 * y_{t-7}$, où $y_t = r_t - x1 * r_{t-7}$' veut dire ARMA avec deux saisonnalités, où 'y_{t-7}' est 'y_t' au retard 7, x1 est le paramètre de la première saisonnalité, et x2 est le paramètre de la deuxième saisonnalité. Dans chaque petite table, le deuxième rang avec 'Log-likelihood' en avant donne les valeurs de log-vraisemblance, le deuxième dernier rang avec ‘AIC’ en avant donne AIC et le dernier rang avec ‘BIC’ en avant donne BIC. Au milieu ce sont les résultats du test Box-Pierce avec leur *p-value* entre parenthèses, dont la première colonne précise l’ordre du test, par exemple, ‘14 lags’ signifie que l’on additionne les autocorrélations au carré de l’ordre 1 jusqu'à l’ordre 14. Si le *p-value* est supérieur à 0.05, on écrit en gras ce *p-value* avec le résultat du test Box-Pierce correspondant. Cela montre que l’on passe le test avec l’ordre précisé au début de ce rang. Le premier rang de la table précise l’ordre du modèle, dont la colonne suivie montre les statistiques de ce modèle.

A l’examen de Tableau 1, on voit que seulement pour un ARMA(5,5) avec une saisonnalité et un ARMA(5,5) avec deux saisonnalités, on passe le test Box-Pierce aux ordres élevés, soit 35, 42 et 49. Graphique 2 donne ACF de l’ordre 0 à l’ordre 100 des résidus du modèle ARMA(5,5) avec une saisonnalité pour les log-rendements. Graphique 3 donne ACF de l’ordre 0 à l’ordre 100 des résidus du modèle ARMA(5,5) avec deux saisonnalités pour les log-rendements. On voit bien que les autocorrélations aux ordres élevés dans ces deux figures ne sont pas significatives, alors ces deux modèles passent le test. En comparant AIC et BIC de ces deux modèles, un ARMA(5,5) avec une saisonnalité a AIC minimal et BIC minimal, alors un ARMA(5,5) avec une saisonnalité est meilleur qu’un ARMA(5,5) avec deux saisonnalités.

Ensuite, on régresse par MCO les log-rendements sur les variables binaires de mois et de saisons, puis tire les résidus. On estime les 15 modèles pour les résidus de chaque régression respectivement et donne leurs statistiques par tableaux. Tableau 2 montre les statistique pour les modèle après avoir fait la régression sur les variables de mois, et Tableau 3 pour la régression sur les variables de saisons. Ces deux tableaux sont sous la même forme que Tableau 1, à l’exception que U signifie le résidu de la régression MCO. Examinons Tableau 2 et Tableau 3 de la même façon que Tableau 1 :

(1) Pour le résidu de la régression MCO des log-rendements sur les variables binaires de mois, seulement un ARMA(5,5) avec une saisonnalité et un ARMA(5,5) avec deux saisonnalités passent le test Box-Pierce. En comparant ces deux modèles, le modèle ARMA(5,5) avec une saisonnalité minimise AIC et BIC, donc est le meilleur.

(2) Pour le résidu de la régression MCO des log-rendements sur les variables de saisons, un ARMA(5,5) avec une saisonnalité et un ARMA(5,5) avec deux saisonnalités passe le test Box-Pierce. Le modèle ARMA(5,5) avec une saisonnalité minimise AIC et BIC, donc est le meilleur.

À la fin, comparons les meilleurs modèles dans les trois tableaux pour choisir le modèle qui décrit le mieux la moyenne de la série. Voyons bien que, selon les statistiques, un ARMA(5,5) avec une saisonnalité pour les résidus de la régression sur les variables de saisons a AIC et BIC minimaux parmi ces trois modèles.

5.1.2 Tests des paramètres

Tableau 4 donne les résultats du test *t* pour les paramètres des ARMA(5,5) avec une saisonnalité dans trois catégories : le modèle pour les log-rendements, le modèle pour les résidus MCO des log-rendements sur les variables de mois, et le modèle pour les résidus MCO des log-rendements sur les variables de saisons. Tableau 5 donne les résultats du test *t* pour les paramètres des ARMA(5,5) avec deux saisonnalités dans ces trois catégories. Dans ces deux tableaux, x, x1 et x2 sont les paramètres de saisonnalité, a1 à a5 signifient l'ordre AR(1) jusqu'à AR(5), b1 à b5 signifie l'ordre MA(1) jusqu'à MA(5), c est la constante. Tableau 4 montre que seulement la constante n'est pas significative pour les ARMA(5,5) avec une saisonnalité dans chaque catégorie, par contre, selon Tableau 5, outre la constante, les deux paramètres de saisonnalité ne sont pas significatifs non plus pour les ARMA(5,5) avec deux saisonnalités dans chaque catégorie. Puisque les deux paramètres de saisonnalité sont tous non-significatifs, le modèle ARMA avec deux saisonnalités doit être réduit en modèle ARMA sans saisonnalité, alors on n'accepte pas le modèle avec deux saisonnalités. Par contre, si seule la constante n'est pas significative, le modèle pourrait être un ARMA sans constante, donc on ne rejette pas le modèle. Cela justifie la préférence envers le modèle ARMA(5,5) avec une saisonnalité en comparaison du modèle ARMA(5,5) avec deux saisonnalités.

Pour le moment, le meilleur modèle choisi est dans la catégorie : modèles pour le résidu MCO sur les variables de saisons. Alors on va faire le test LR centré sur les modèles de cette catégorie. Premièrement comparons le modèle ARMA avec une saisonnalité et le modèle ARMA avec deux saisonnalités avec les mêmes ordres AR et MA, au sens du test LR, le modèle ARMA avec une saisonnalité est le modèle contraint, le modèle ARMA avec deux saisonnalités est le modèle non-contraint, le nombre de paramètres restreints est unité. Si le résultat du test LR est plus grand que la valeur critique au niveau 5% de chi-carré avec un degré de liberté, qui est égale à 3.8415, on rejette l'hypothèse nulle, donc le modèle ARMA avec deux saisonnalités est meilleur que le modèle ARMA avec une saisonnalité ; par contre, si le résultat du test est inférieur à 3.8415, on choisit le modèle ARMA avec une saisonnalité. Regardons Tableau 3:

pour ARMA(1,1), LR = 16 > 3.8415;
pour ARMA(2,2), LR = 14.6 > 3.8415;
pour ARMA(3,3), LR = 23.2 > 3.8415;
pour ARMA(4,4), LR = 3.4 < 3.8415;
pour ARMA(5,5), LR = 0.2 < 3.8415.

En un mot, de l'ordre 1 à l'ordre 3, le modèle ARMA avec deux saisonnalités est bien différent du modèle ARMA avec une saisonnalité. Pourtant, à l'ordre 4 et 5, le modèle ARMA avec deux saisonnalités n'est pas différent du modèle ARMA avec une saisonnalité, surtout pour un ARMA(5,5), la différence est très faible. Cela constitue aussi la raison que l'on préfère un ARMA(5,5) avec une saisonnalité à un ARMA(5,5) avec deux saisonnalités.

Ayant prouvé que le modèle ARMA(5,5) avec une saisonnalité est meilleur que les autres modèles estimés, on a l'intention de voir si le modèle à l'ordre plus élevé est encore meilleur que le modèle à l'ordre 5. Donc, on va estimer un ARMA(6,6) avec une saisonnalité, et faire le test LR en prenant un ARMA(5,5) avec une saisonnalité comme modèle contraint et un ARMA(6,6) avec une saisonnalité comme modèle non-contraint, dans ce cas, le nombre de paramètres restreints est deux, alors on compare le résultat LR avec la valeur critique au niveau 5% de chi-carré avec 2 degrés de liberté, qui est 5.9915. Si l'on constate que le modèle à l'ordre 6 est meilleur que celui à l'ordre 5, autrement dit, si l'on rejette l'hypothèse nulle du test LR entre ces deux modèles, on estime un ARMA(7,7) avec une saisonnalité et fait le test LR avec un ARMA(6,6) avec une saisonnalité en prenant celui-ci comme modèle contraint. On continue à augmenter un ordre du modèle et fait le test LR avec le modèle un ordre moins grand jusqu'à ce que l'on ne rejette pas l'hypothèse nulle de ce test, soit les deux modèles impliqués par le test ne sont pas différents. Enfin on prend le modèle contraint dans le dernier test que l'on fait comme meilleur un. Pour les modèles ajoutés ci-dessus, on estime aussi les ARMA avec deux saisonnalités aux mêmes ordres, puis donne leurs statistiques comme référence, notamment voir le changement da la valeur de log-vraisemblance.

Regardons Tableau 6, la valeur de log-vraisemblance pour un ARMA(6,6) avec une saisonnalité est –1014.1. En faisant le test LR, le modèle contraint est ARMA(5,5) avec une saisonnalité, dont la valeur de log-vraisemblance est –1019.9, on a donc

$$LR = 2(1019.9 - 1014.1) = 11.6 > 5.9915$$

Alors on rejette l'hypothèse nulle du test, cela signifie que les paramètres jetés par le modèle ARMA(5,5) avec une saisonnalité sont importants, par conséquent, le modèle ARMA(6,6) avec une saisonnalité est meilleur que le modèle à l'ordre 5. Ensuite on voit si l'on peut trouver un modèle encore meilleur en augmentant l'ordre du modèle SARMA, et estime un ARMA(7,7) avec une saisonnalité. Enfin on obtient sa valeur de log-vraisemblance qui est –1014.1 aussi. Donc au test LR sur l'ordre 7, en prenant ARMA(6,6) avec une saisonnalité comme modèle contraint, on a

$$LR = 2(1014.1 - 1014.1) = 0 < 5.9915$$

On ne rejette pas l'hypothèse nulle, donc les paramètres jetés par le modèle à l'ordre 6 ne sont pas importants et on s'arrête en ARMA(6,6) avec une saisonnalité. Voyons aussi les modèles à l'ordre 6 et à l'ordre 7 avec deux saisonnalités, ils ont les mêmes valeurs de log-vraisemblance que les modèles avec une saisonnalité aux mêmes ordres, soit –1014.1, donc on choisit ARMA(6,6) avec une saisonnalité comme meilleur modèle de la moyenne conditionnelle. Notons que ces modèles aux ordres 6 et 7 ont tous les résultats du test Box-Pierce qui ne sont pas significatifs de l'ordre 21 à l'ordre 49 du test, alors sont meilleurs que le modèle à l'ordre 5. De plus, le meilleur modèle choisi, soit

ARMA(6,6) avec une saisonnalité, minimise AIC. On peut aussi faire le test *t* sur les paramètres du modèle à l'ordre 6, Tableau 7 donne le test sur ARMA(6,6) avec une saisonnalité, en comparaison avec un ARMA(5,5) avec une saisonnalité, le coefficient de AR(5) n'est pas significatif, pour les ordres ajoutés, AR(6) et MA(6), leur coefficients ne sont pas significatifs non plus. Tableau 8 donne le test sur un ARMA(6,6) avec deux saisonnalités comme référence, aussi les coefficients de AR(5), AR(6) et MA(6) sont non significatifs, de plus, les deux paramètres de saisonnalité ne sont pas significatifs. Ce résultat est tout au contraire du résultat du test LR, selon le dernier, les coefficients de AR(6) et MA(6) sont significatifs. Cette contradiction serait causée par la propriété du test *t*, le test est basé sur l'échantillon grand de sorte que la distribution des données est normale, la variance asymptotique est calculée sur la base de la normalité. Pourtant, la taille de l'échantillon du marché PJM est 1958, qui n'est pas assez grande pour l'approximation de la normalité, cela donne un biais sur les statistiques *t*. En revanche, le test LR est le test conjoint de tous les paramètres estimés. La règle du test est de voir la diminution de la valeur de log-vraisemblance, cette diminution ne dépend pas beaucoup de la normalité. Dans ce sens le test LR est meilleur que le test du paramètre unique.

Par conséquent, on prend finalement le test LR en compte pour examiner si l'estimateur est significatif. La conclusion est que, faisant la régression MCO des log-rendements sur les variables binaires de saisons, un ARMA(6,6) avec une saisonnalité pour le résidu MCO est le meilleur modèle de la moyenne conditionnelle.

5.2. Ontario

À l’examen de l’autocorrélation des log-rendements journaliers, on prend aussi la saisonnalité de 7 pendant la modélisation, voir Graphique 4. Après avoir estimé les 45 modèles pour la moyenne, on trouve le modèle ARMA avec deux saisonnalités pour les résidus MCO sur les variables de saisons comme candidats du meilleur modèle. On donne les statistiques de ces modèles de l’ordre 1 à l’ordre 5 au Tableau 9. Un ARMA(5,5) avec deux saisonnalités semble passer le mieux le test Box-Pierce, à partir de l’ordre 21, les résultats du test ne sont pas significatifs, néanmoins la plupart des coefficients estimés de l’ordre AR et MA ne sont pas significatifs au sens du test *t*, voit Tableau 10. À part le modèle ARMA(5,5) avec deux saisonnalités, les modèles passant les mieux le test Box-Pierce sont ceux avec seul l’ordre 35 qui est significatif, et parmi eux, un ARMA(2,2) avec deux saisonnalités minimise AIC alors qu’un ARMA(1,1) avec deux saisonnalités minimise BIC. Faisons le test LR du modèle ARMA(5,5) avec deux saisonnalités en prenant ARMA(2,2) avec deux saisonnalités comme modèle contraint, on a

$$\text{LR} = 2(350.4283 - 345.5287) = 9.7992 < 12.5916,$$

où 12.5916 est la valeur critique au niveau 5% de chi-carré avec 6 degrés de liberté. Alors le modèle à l’ordre 5 n’est pas différent du modèle à l’ordre 2, donc on rejette le modèle ARMA(5,5) avec deux saisonnalités.

Pour un ARMA(1,1) avec deux saisonnalités et un ARMA(2,2) avec deux saisonnalités, puisque chacun des deux a l’ordre 35 significatif au sens du test Box-Pierce , on ne peut pas dire si ces modèles sont bons. La saisonnalité de 7 représente la périodicité d’une semaine, il pourrait avoir une saisonnalité sur plus d’une semaine, donc on garde 7 comme la première saisonnalité, et essaye la deuxième saisonnalité de 14, 21, 28 et 35 dans l’estimation du modèle ARMA avec deux saisonnalités pour les résidus MCO des log-rendements sur les variables de saisons. Enfin on choisit le meilleur modèle comme dynamique de la moyenne conditionnelle. Regardons Tableau 11, les modèles avec l’ordre 35 non significatif au sens du test Box-Pierce sont les bons modèles, en choisissant la deuxième saisonnalité de 28, un ARMA(1,1) avec deux saisonnalités donne

BIC minimal et un ARMA(2,2) avec deux saisonnalités donne AIC minimal parmi ces bons modèles. Ensuite faisons le test LR sur le modèle ARMA(2,2) avec deux saisonnalités en prenant ARMA(1,1) avec deux saisonnalités comme modèle contraint, on a

$$LR = 2(350.2895 - 346.0373) = 8.5044 > 5.9915,$$

alors le modèle ARMA(2,2) avec deux saisonnalités est bien différent du modèle ARMA(1,1) avec deux saisonnalités. Faisons encore le test LR sur le modèle ARMA(3,3) avec deux saisonnalités en prenant ARMA(2,2) avec deux saisonnalités comme modèle contraint, on a

$$LR = 2(346.0373 - 343.9012) = 4.2722 < 5.9915,$$

alors le modèle ARMA(3,3) avec deux saisonnalités n'est pas différent du modèle ARMA(2,2) avec deux saisonnalités.

Par conséquent, le meilleur modèle de la moyenne conditionnelle du marché Ontario est le modèle ARMA(2,2) avec deux saisonnalités pour les résidus MCO des log-rendements sur les variables de saisons, en prenant la première saisonnalité de 7 et la deuxième saisonnalité de 28.

5.3. NE (New England)

Prenons aussi la saisonnalité de 7 dans la modélisation selon le graphique ACF des log-rendements, voir Graphique 5. À l'estimation des 45 modèles, on voit que les meilleurs sont obtenus en régressant d'abord les log-rendements par MCO sur les variables de mois, puis estimant un ARMA(5,5) pour les résidus et un ARMA(5,5) avec une saisonnalité pour les résidus, celui-ci minimise AIC et celui-là minimise BIC. Faisons le test LR sur le modèle ARMA(5,5) avec une saisonnalité en prenant ARMA(5,5) comme modèle contraint :

$$LR = 2(100.3247 - 98.3968) = 3.8558 > 3.8415,$$

on rejette l'hypothèse nulle du test, alors le modèle ARMA(5,5) avec une saisonnalité est statistiquement différent du modèle ARMA(5,5) et donc est le meilleur.

On a l'intention de voir si le modèle à l'ordre plus élevé est encore meilleur que le modèle à l'ordre 5. Donc, on va estimer un ARMA(6,6) avec une saisonnalité, et faire le test LR en prenant ARMA(5,5) avec une saisonnalité comme modèle contraint, Tableau 12 donne les statistiques de ces deux modèles, plus un ARMA(5,5) sans saisonnalité comme référence.

Pour tester la différence entre le modèle ARMA(6,6) avec une saisonnalité et le modèle ARMA(5,5) avec une saisonnalité :

$$LR = 2(98.3968 - 98.2256) = 0.3424 < 5.9915,$$

alors le modèle ARMA(6,6) avec une saisonnalité n'est pas différent du tout du modèle ARMA(5,5) avec une saisonnalité. Par conséquent, on prend le modèle ARMA(5,5) avec une saisonnalité pour le résidu MCO des log-rendements sur les variables de mois comme meilleur modèle.

5.4. NYA (New York, zone A)

On voit une forte autocorrélation tous les 7 jours par le graphique ACF des log-rendements, voir Graphique 6, donc on prend la saisonnalité de 7 dans la modélisation. À l'estimation des 45 modèles, aucun modèle ne passe le test Box-Pierce, parmi eux, le modèle ARMA(5,5) avec deux saisonnalités pour les résidus MCO des log-rendements sur les variables de mois minimise AIC et BIC, donc est le meilleur. On a l'intention de voir si le modèle à l'ordre plus élevé est encore meilleur que le modèle à l'ordre 5. Donc, on va estimer un ARMA(6,6) avec deux saisonnalité du résidu, et faire le test LR en prenant ARMA(5,5) avec deux saisonnalités comme modèle contraint. Tableau 13 donne les statistiques de ces deux modèles, ARMA(6,6) avec deux saisonnalité ne passe pas le test Box-Pierce non plus.

Pour tester la différence entre le modèle ARMA(6,6) avec deux saisonnalités et le modèle ARMA(5,5) avec deux saisonnalités :

$$LR = 2(251.7475 - 249.5282) = 4.4386 < 5.9915,$$

alors le modèle ARMA(6,6) avec deux saisonnalités n'est pas différent du modèle ARMA(5,5) avec deux saisonnalités. Par conséquent, ARMA(5,5) avec deux saisonnalités pour le résidu MCO des log-rendements sur les variables de mois est le meilleur modèle.

5.5. NYG (New York, zone G)

On voit une forte autocorrélation tous les 7 jours par le graphique ACF des log-rendements, voir Graphique 7, donc on prend la saisonnalité de 7 dans la modélisation. À l'estimation des 45 modèles, aucun modèle ne passe le test Box-Pierce, parmi eux, le modèle ARMA(5,5) avec deux saisonnalités pour les résidus MCO des log-rendements sur les variables de mois minimise AIC et BIC, donc est le meilleur. On a l'intention de voir si le modèle à l'ordre plus élevé est encore meilleur que le modèle à l'ordre 5, voir Tableau 14. En faisant le test LR, les paramètres pour un ARMA(6,6) avec deux saisonnalités et les paramètres pour un ARMA(7,7) avec deux saisonnalités sont tous conjointement significatifs, de sorte que l'on ne rejette pas ces modèles. Pourtant, on rejette le modèle ARMA(8,8) avec deux saisonnalités dont le résultat du test LR :

$$LR = 2(243.6720 - 240.8838) = 5.5764 < 5.9915,$$

en prenant ARMA(7,7) avec deux saisonnalités comme modèle contraint. Alors le modèle ARMA(8,8) avec deux saisonnalités n'est pas différent du modèle ARMA(7,7) avec deux saisonnalités. Par conséquent, on fait d'abord la régression MCO des log-rendements sur les variables de mois, ensuite prend le modèle ARMA(7,7) avec deux saisonnalités pour le résidu comme meilleur modèle de la moyenne.

5.6. NYJ (New York, zone J)

On voit une forte autocorrélation tous les 7 jours par le graphique ACF des log-rendements, voir Graphique 8, donc on prend la saisonnalité de 7 dans la modélisation. À l'estimation des 45 modèles, on voit que les meilleurs sont obtenus en régressant d'abord les log-rendements par MCO sur les variables de mois, puis estimant un ARMA(5,5) avec une saisonnalité pour les résidus et un ARMA(5,5) avec deux saisonnalités pour les résidus, celui-ci minimise AIC et celui-là minimise BIC. Faisons le test LR sur le modèle ARMA(5,5) avec deux saisonnalités en prenant ARMA(5,5) avec une saisonnalité comme modèle contraint :

$$LR = 2(176.1671 - 173.7169) = 4.9004 > 3.8415,$$

on rejette l'hypothèse nulle du test, alors un ARMA(5,5) avec deux saisonnalités est meilleur qu'un ARMA(5,5) avec une saisonnalité. Estimons les ARMA avec deux saisonnalités aux ordre plus élevés et faisons le test LR, voir Tableau 15. En faisant le test LR, les paramètres pour un ARMA(6,6) avec deux saisonnalités et un ARMA(7,7) avec deux saisonnalités sont tous conjointement significatifs, de sorte que l'on ne rejette pas ces modèles. Pourtant, on rejette le modèle ARMA(8,8) avec deux saisonnalités dont le résultat du test :

$$LR = 2(195.6893 - 194.6828) = 2.0130 < 5.9915,$$

en prenant ARMA(7,7) avec deux saisonnalités comme modèle contraint. Alors le modèle ARMA(8,8) avec deux saisonnalités n'est pas différent du modèle ARMA(7,7) avec deux saisonnalités. Par conséquent, on fait d'abord la régression MCO des log-rendements sur les variables de mois, ensuite prend le modèle ARMA(7,7) avec deux saisonnalités pour les résidus comme meilleur modèle de la moyenne.

6. Conclusion

Ce projet consiste à trouver les meilleurs modèles ARMA ou SARMA pour la moyenne conditionnelle des log-rendements journaliers des différents marchés de l'électricité. Puisque les rendements de tous les six marchés présentent une forte autocorrélation tous les sept jours, on a besoin de déterminer le niveau de saisonnalité dans la modélisation. Cette autocorrélation hebdomadaire donne l'idée de prendre 7 ou un multiple de 7 comme niveau de saisonnalité. À l'estimation du modèle SARMA, on cherche à modéliser des modèles avec une saisonnalité et ceux avec deux saisonnalités. Pour le modèle avec une saisonnalité, on prend toujours la saisonnalité de 7, cela représente une périodicité hebdomadaire. Pour le modèle avec deux saisonnalités, on prend toujours la première saisonnalité de 7, quant à la deuxième, on donne le niveau de saisonnalité comme 7 ou son multiple, soit 14, 21, ou 28. D'ailleurs, afin de vérifier l'existence d'un plan de périodicité parmi les mois ou les saisons sur les marchés, on cherche à estimer une autre catégorie de modèle. La méthode de la modélisation est de faire la régression MCO des log-rendements sur les variables binaires, puis chercher le meilleur modèle ARMA ou SARMA pour les résidus. Cette approche permet d'éliminer la périodicité de mois ou de saisons. En faisant la régression MCO, les résidus ne sont pas corrélés avec les variables indépendantes, alors on peut enlever l'impact des mois ou des saisons en modélisant les résidus. On essaie une variété de saisonnalités parce que l'on ne sait pas laquelle correspond le mieux à la vraie dynamique des données. Après avoir estimé tous ces modèles, on fait une comparaison parmi eux et on choisit le meilleur.

Les résultats de ces six marchés sont bien différents, comme on le dit, les marchés tendent à rester régionaux, de sorte qu'il est difficile de trouver un modèle général.

(1) Pour avoir le meilleur modèle du marché PJM, on fait premièrement la régression MCO des log-rendements sur les variables binaires de saisons, ensuite estime le modèle ARMA(6,6) avec une saisonnalité hebdomadaire pour les résidus MCO.

(2) Pour le marché Ontario, on fait la régression MCO des log-rendements sur les variables binaires de saisons, et puis estime le modèle ARMA(2,2) avec deux saisonnalités pour les résidus, en prenant la première saisonnalité de 7 jours et la

deuxième saisonnalité de 28 jours. La première saisonnalité de 7 représente la saisonnalité hebdomadaire, alors que la deuxième saisonnalité de 28 représente la saisonnalité mensuelle.

(3) Pour le marché NE, on fait la régression MCO des log-rendements sur les variables binaires de mois, puis estime le modèle ARMA(5,5) avec une saisonnalité hebdomadaire pour les résidus.

(4) Pour le marché NYA, on fait la régression MCO des log-rendements sur les variables binaires de mois, puis estime le modèle ARMA(5,5) avec deux saisonnalités hebdomadaires pour les résidus.

(5) Pour le marché NYG, on fait la régression MCO des log-rendements sur les variables binaires de mois, puis estime le modèle ARMA(7,7) avec deux saisonnalités hebdomadaires pour les résidus.

(6) Pour le marché NYJ, on fait la régression MCO des log-rendements sur les variables binaires de mois, puis estime le modèle ARMA(7,7) avec deux saisonnalités hebdomadaires pour les résidus.

Les modèles choisis sur tous les marchés consistent à faire la régression MCO sur les variables binaires et modéliser les résidus, cela conclut que tous ces marchés comportent les saisonnalités '*intra-year*', autrement dit, une périodicité sur les mois ou les saisons. Chacun des marchés NYA, NYG et NYJ comprend deux saisonnalités hebdomadaires, et en modélisant les prix de ces trois marchés, on fait la régression MCO des log-rendements sur les variables de mois. Comme tous ces trois marchés appartiennent au marché New York, ils pourraient partager les mêmes structures et alors les mêmes caractéristiques, d'où la similarité. En général, la modélisation de la moyenne conditionnelle est effectuée dans le but d'estimer la volatilité, parce que la volatilité présente le risque, donc est le sujet auquel on s'intéresse. Par contre, on se limite en modélisation de la moyenne dans ce projet, car celle-ci comporte déjà beaucoup de sujets à étudier.

Appendice : Tableaux et graphiques

Tableau 1 : Modèles pour les log-rendements du marché PJM

ARMA pour r_t:

	ARMA (1,1)	ARMA(2,2)	ARMA(3,3)	ARMA(4,4)	ARMA (5,5)
Log-likelihood	-1124.6	-1124.5	-1066.9	-1065.0	-1063.4
Test Box-Pierce					
7 lags	56.4371(0)	55.8710(0)	27.1742(0)		
14 lags	95.3732(0)	94.5523(0)	43.9784(0)	41.7259(0)	34.7146(0)
21 lags	155.7968(0)	154.5486(0)	69.9526(0)	69.0320(0)	59.5232(0)
28 lags	219.9389(0)	218.2133(0)	101.7638(0)	102.6761(0)	89.4299(0)
35 lags	267.9077(0)	265.6868(0)	116.8528(0)	118.9558(0)	103.4682(0)
42 lags	318.6207(0)	316.0533(0)	132.5743(0)	134.3036(0)	118.0137(0)
49 lags	363.1179(0)	360.1983(0)	148.4565(0)	150.6449(0)	132.5091(0)
AIC	-1.6861	-1.6841	-1.7409	-1.7409	-1.7404
BIC	-1.6776	-1.6698	-1.7209	-1.7152	-1.7091

ARMA pour $y_t = r_t - x * r_{t-7}$:

	ARMA (1,1)	ARMA(2,2)	ARMA(3,3)	ARMA(4,4)	ARMA (5,5)
Log-likelihood	-1098.8	-1095.6	-1056.8	-1053.8	-1026.8
Test Box-Pierce					
7 lags	**8.2329(0.0834)**	**5.3366(0.0694)**			
14 lags	24.4564(0.0109)	**16.7794(0.0523)**	17.5582(0.0141)	11.6103(0.0405)	15.7624(0.0013)
21 lags	58.5550(0)	51.6075(0)	36.0456(0.0010)	30.7680(0.0021)	21.5685(0.0175)
28 lags	97.1542(0)	85.9749(0)	61.0406(0)	56.1807(0)	31.6624(0.0166)
35 lags	120.7820(0)	111.1623(0)	70.4229(0)	66.0047(0)	**34.1368(0.0823)**
42 lags	148.9544(0)	138.3092(0)	82.0427(0)	77.1189(0)	**41.8177(0.0929)**
49 lags	176.4643(0)	165.6295(0)	94.9380(0)	89.9005(0)	**49.3304(0.1031)**
AIC	-1.7115	-1.7127	-1.7502	-1.7513	-1.7767
BIC	-1.7001	-1.6956	-1.7274	-1.7228	-1.7425

ARMA pour $z_t = y_t - x2 * y_{t-7}$, où $y_t = r_t - x1 * r_{t-7}$:

	ARMA (1,1)	ARMA(2,2)	ARMA(3,3)	ARMA(4,4)	ARMA (5,5)
Log-likelihood	-1090.8	-1088.2	-1055.2	-1052.0	-1026.6
Test Box-Pierce					
7 lags	8.1794(0.0424)	**3.7655(0.0523)**			
14 lags	**14.8499(0.1376)**	**8.9683(0.3450)**	14.1719(0.0278)	**8.2506(0.0828)**	15.2399(0.0005)
21 lags	39.5621(0.0015)	33.2737(0.0043)	30.4088(0.0041)	25.1266(0.0087)	20.7444(0.0138)
28 lags	72.4410(0)	67.1268(0)	54.3732(0.0001)	49.4371(0.0001)	31.3553(0.0121)
35 lags	89.2611(0)	84.2627(0)	62.4785(0.0001)	57.9906(0.0002)	**33.9163(0.0664)**
42 lags	110.8788(0)	104.4093(0)	72.8838(0.0001)	67.8033(0.0002)	**41.2866(0.0823)**
49 lags	133.1701(0)	126.7688(0)	85.3588(0.0001)	79.9664(0.0001)	**48.6750(0.0948)**
AIC	-1.7186	-1.7191	-1.7509	-1.7520	-1.7760
BIC	-1.7044	-1.6992	-1.7252	-1.7207	-1.7389

Tableau 2 : Modèles pour les résidus de la régression MCO des log-rendements du marché PJM sur les variables de mois

ARMA pour U:

	ARMA (1,1)	ARMA(2,2)	ARMA(3,3)	ARMA(4,4)	ARMA (5,5)
Log-likelihood	– 1119.0	–1102.1	–1101.4	–1058.4	– 1024.7
Test Box-Pierce					
7 lags	58.6972(0)	65.5952(0)	56.6038(0)		
14 lags	95.6615(0)	106.3605(0)	96.6577(0)	40.0573(0)	23.9400(0.0001)
21 lags	154.3991(0)	167.3118(0)	157.8068(0)	66.0374(0)	29.7498(0.0017)
28 lags	215.5068(0)	233.0175(0)	222.9220(0)	96.5409(0)	38.0158(0.0039)
35 lags	263.9487(0)	285.1751(0)	271.7680(0)	112.6506(0)	42.1140(0.0175)
42 lags	317.7150(0)	337.8071(0)	322.9703(0)	130.1438(0)	52.9494(0.0114)
49 lags	362.1596(0)	385.5616(0)	369.9680(0)	145.8140(0)	59.7202(0.0180)
AIC	-1.6918	-1.7070	-1.7057	-1.7476	-1.7799
BIC	-1.6832	-1.6928	-1.6858	-1.7219	-1.7486

ARMA pour $y_t = U_t - x * U_{t-7}$:

	ARMA (1,1)	ARMA(2,2)	ARMA(3,3)	ARMA(4,4)	ARMA (5,5)
Log-likelihood	–1091.8	– 1088.4	– 1057.2	–1046.7	–1021.2
Test Box-Pierce					
7 lags	**9.1637(0.0571)**	**5.9190(0.0518)**			
14 lags	24.4294(0.0110)	**16.4810(0.0575)**	27.2443(0.0003)	11.2185(0.0472)	16.4961(0.0009)
21 lags	57.2453(0)	50.1830(0)	51.2703(0)	29.4547(0.0034)	21.8665(0.0158)
28 lags	93.0225(0)	81.9667(0)	76.9845(0)	52.4963(0.0001)	30.1955(0.0250)
35 lags	116.1655(0)	106.5620(0)	90.0564(0)	62.1132(0.0001)	**33.4869(0.0942)**
42 lags	145.9241(0)	135.2301(0)	107.3571(0)	74.5843(0)	**44.0592(0.0603)**
49 lags	173.2119(0)	162.4670(0)	121.8540(0)	87.0959(0)	**50.8958(0.0788)**
AIC	-1.7186	-1.7200	-1.7499	-1.7585	-1.7825
BIC	-1.7072	-1.7029	-1.7271	-1.7300	-1.7483

ARMA pour $z_t = y_t - x2 * y_{t-7}$, où $y_t = U_t - x1 * U_{t-7}$:

	ARMA (1,1)	ARMA(2,2)	ARMA(3,3)	ARMA(4,4)	ARMA (5,5)
Log-likelihood	–1083.9	–1080.8	–1048.6	–1045.0	– 1021.1
Test Box-Pierce					
7 lags	9.4303(0.0241)	**3.7470(0.0529)**			
14 lags	**15.3411(0.1201)**	**8.7119(0.3672)**	14.6660(0.0230)	**8.1073(0.0877)**	16.5677(0.0003)
21 lags	39.1904(0.0017)	32.6166(0.0053)	30.0861(0.0046)	24.1531(0.0121)	21.7294(0.0098)
28 lags	69.3254(0)	64.3123(0)	51.4336(0.0001)	46.0445(0.0003)	30.2695(0.0167)
35 lags	85.8527(0)	81.1531(0)	59.4436(0.0003)	54.2174(0.0006)	**33.5517(0.0719)**
42 lags	109.0201(0)	102.6236(0)	71.6847(0.0002)	65.3840(0.0004)	**43.6076(0.0517)**
49 lags	131.3956(0)	125.4503(0)	83.9081(0.0001)	77.3424(0.0002)	**50.5242(0.0683)**
AIC	-1.7257	-1.7268	-1.7576	-1.7592	-1.7816
BIC	-1.7114	-1.7068	-1.7320	-1.7279	-1.7446

Tableau 3 : Modèles pour les résidus de la régression MCO des log-rendements du marché PJM sur les variables de saisons

ARMA pour U:

	ARMA (1,1)	ARMA(2,2)	ARMA(3,3)	ARMA(4,4)	ARMA (5,5)
Log-likelihood	–1117.7	–1108.1	–1106.1	–1057.3	–1023.2
Test Box-Pierce					
7 lags	57.6937(0)	54.8048(0)	54.2437(0)		
14 lags	95.3894(0)	100.6848(0)	98.8710(0)	39.9841(0)	23.7224(0.0001)
21 lags	154.7217(0)	165.4644(0)	162.8995(0)	66.3389(0)	29.3241(0.0020)
28 lags	217.2495(0)	230.0998(0)	226.8421(0)	97.9715(0)	38.8470(0.0030)
35 lags	265.2449(0)	280.1013(0)	276.5324(0)	113.7640(0)	42.2233(0.0170)
42 lags	317.0454(0)	333.9286(0)	330.4401(0)	130.0206(0)	51.0001(0.0178)
49 lags	361.1280(0)	379.9356(0)	375.9394(0)	145.5980(0)	57.8711(0.0263)
AIC	-1.6932	-1.7009	-1.7009	-1.7487	-1.7815
BIC	-1.6846	-1.6866	-1.6809	-1.7231	-1.7501

ARMA pour $y_t = U_t - x * U_{t-7}$:

	ARMA (1,1)	ARMA(2,2)	ARMA(3,3)	ARMA(4,4)	ARMA (5,5)
Log-likelihood	–1091.1	–1087.7	–1059.3	–1046.0	–1019.9
Test Box-Pierce					
7 lags	**8.7189(0.0685)**	**5.6109(0.0605)**			
14 lags	24.4392(0.0110)	**16.6269(0.0549)**	42.0717(0)	11.3591(0.0447)	16.1241(0.0011)
21 lags	57.8431(0)	50.8798(0)	67.4366(0)	30.0080(0.0028)	21.5785(0.0174)
28 lags	95.2595(0)	84.2200(0)	97.4767(0)	54.4052(0)	30.9834(0.0201)
35 lags	118.6135(0)	109.1522(0)	112.4095(0)	64.0745(0)	**33.7066(0.0900)**
42 lags	147.1670(0)	136.6795(0)	129.3814(0)	75.6805(0)	**42.5962(0.0802)**
49 lags	174.3517(0)	163.8058(0)	144.6834(0)	88.1488(0)	**49.4579(0.1009)**
AIC	-1.7193	-1.7207	-1.7477	-1.7592	-1.7839
BIC	-1.7079	-1.7036	-1.7249	-1.7307	-1.7497

ARMA pour $z_t = y_t - x2 * y_{t-7}$, où $y_t = U_t - x1 * U_{t-7}$:

	ARMA (1,1)	ARMA(2,2)	ARMA(3,3)	ARMA(4,4)	ARMA (5,5)
Log-likelihood	–1083.1	–1080.4	–1047.7	–1044.3	–1019.8
Test Box-Pierce					
7 lags	8.7819(0.0323)	**3.7649(0.0523)**			
14 lags	**15.1046(0.1283)**	**8.9878(0.3433)**	14.4581(0.0249)	**8.6496(0.0705)**	16.0307(0.0003)
21 lags	39.3869(0.0016)	33.1838(0.0044)	30.1190(0.0045)	25.3178(0.0082)	21.3190(0.0113)
28 lags	71.3086(0)	66.3267(0)	52.9819(0.0001)	49.7413(0.0001)	31.0543(0.0132)
35 lags	87.9993(0)	83.4839(0)	60.9678(0.0002)	58.3682(0.0002)	**33.8004(0.0681)**
42 lags	109.9332(0)	104.0766(0)	71.7097(0.0002)	68.3894(0.0002)	**42.4190(0.0659)**
49 lags	132.1832(0)	126.7410(0)	83.8649(0.0001)	80.9113(0.0001)	**49.2822(0.0853)**
AIC	-1.7265	-1.7272	-1.7585	-1.7596	-1.7829
BIC	-1.7122	-1.7072	-1.7329	-1.7282	-1.7459

Tableau 4 : Test de significativité pour ARMA(5,5) avec une saisonnalité sur le marché PJM

	Pour les log-rendements				Pour les résidus MCO des log-rendements sur les variables de mois				Pour les résidus MCO des log-rendements sur les variables de saisons			
	Estimateur (θ_k)	Ecart-type(σ)	θ_k/σ	Significatif	Estimateur (θ_k)	Ecart-type(σ)	θ_k/σ	Significatif	Estimateur (θ_k)	Ecart-type(σ)	θ_k/σ	Significatif
x	0.0533	0.0229	2.3275	Oui	0.0635	0.0224	2.8348	Oui	0.0597	0.0228	2.6184	Oui
a1	1.1137	0.0189	58.93	Oui	1.1285	0.0176	64.12	Oui	1.1268	0.0178	63.30	Oui
a2	-1.6945	0.0153	-110.75	Oui	-1.7066	0.0142	-120.18	Oui	-1.7052	0.0144	-118.42	Oui
a3	1.2514	0.0272	46.01	Oui	1.2728	0.0253	50.31	Oui	1.2703	0.0256	49.62	Oui
a4	-1.2493	0.0153	-81.65	Oui	-1.2611	0.0142	-88.81	Oui	-1.2598	0.0144	-87.49	Oui
a5	0.3115	0.0189	16.48	Oui	0.3260	0.0176	18.52	Oui	0.3244	0.0178	18.22	Oui
b1	-1.7415	0.0101	-172.43	Oui	-1.7632	0.0086	-205.02	Oui	-1.7610	0.0087	-202.41	Oui
b2	2.1916	0.0132	166.03	Oui	2.2116	0.0114	194	Oui	2.2090	0.0120	184.08	Oui
b3	-2.1423	0.0186	-115.18	Oui	-2.1758	0.0158	-137.71	Oui	-2.1716	0.0165	-131.61	Oui
b4	1.7291	0.0133	130.01	Oui	1.7486	0.0117	149.45	Oui	1.7458	0.0122	143.10	Oui
b5	-0.9242	0.0104	-88.87	Oui	-0.9451	0.0090	-105.01	Oui	-0.9428	0.0090	-104.76	Oui
c	0.0003	0.0011	0.2727	Non	0.0001	0.0007	0.1429	Non	0.0002	0.0008	0.25	Non

Tableau 5 : Test de significativité pour ARMA(5,5) avec deux saisonnalités sur le marché PJM

	Pour les log-rendements				Pour les résidus MCO des log-rendements sur les variables de mois				Pour les résidus MCO des log-rendements sur les variables de saisons			
	Estimateur (θ_k)	Ecart-type(σ)	θ_k/σ	Significatif	Estimateur (θ_k)	Ecart-type(σ)	θ_k/σ	Significatif	Estimateur (θ_k)	Ecart-type(σ)	θ_k/σ	Significatif
x1	-0.0003	0.1020	-0.0029	Non	0.0500	0.5779	0.0865	Non	0.0678	0.2841	0.2386	Non
x2	0.0582	0.1041	0.5591	Non	0.0151	0.5755	0.0262	Non	-0.0050	0.2817	-0.0177	Non
a1	1.1210	0.0188	59.63	Oui	1.1395	0.0181	62.96	Oui	1.1316	0.0182	62.18	Oui
a2	-1.7005	0.0152	-111.88	Oui	-1.7154	0.0146	-117.49	Oui	-1.7090	0.0147	-116.26	Oui
a3	1.2619	0.0271	46.56	Oui	1.2887	0.0261	49.38	Oui	1.2772	0.0262	48.75	Oui
a4	-1.2551	0.0152	-82.57	Oui	-1.2700	0.0146	-86.99	Oui	-1.2636	0.0147	-85.96	Oui
a5	0.3185	0.0188	16.94	Oui	0.3370	0.0181	18.62	Oui	0.3290	0.0182	18.08	Oui
b1	-1.7481	0.0095	-184.01	Oui	-1.7653	0.0086	-205.27	Oui	-1.7641	0.0084	-210.01	Oui
b2	2.1991	0.0123	178.79	Oui	2.2134	0.0115	192.47	Oui	2.2126	0.0115	192.4	Oui
b3	-2.1537	0.0173	-124.49	Oui	-2.1785	0.0160	-136.16	Oui	-2.1771	0.0160	-136.07	Oui
b4	1.7363	0.0124	140.02	Oui	1.7502	0.0120	145.85	Oui	1.7495	0.0118	148.26	Oui
b5	-0.9302	0.0098	-94.92	Oui	-0.9469	0.0090	-105.21	Oui	-0.9459	0.0088	-107.49	Oui
c	0.0003	0.0010	0.3	Non	0.0001	0.0007	0.1429	Non	0.0002	0.0007	0.2857	Non

Tableau 6 : ARMA avec une saisonnalité et deux saisonnalités aux ordres 5, 6 et 7 pour le résidu MCO des log-rendements du marché PJM sur les variables de saisons

ARMA pour $y_t = U_t - x * U_{t-7}$:

	ARMA (5,5)	ARMA (6,6)	ARMA (7,7)
Log-likelihood	–1019.9	–1014.1	–1014.1
Test Box-Pierce			
14 lags	16.1241(0.0011)	4.9947(0.0254)	
21 lags	21.5785(0.0174)	**9.3017(0.3175)**	**9.2265(0.1612)**
28 lags	30.9834(0.0201)	**18.9773(0.2148)**	**18.9210(0.1256)**
35 lags	**33.7066(0.0900)**	**21.1320(0.5126)**	**21.0681(0.3931)**
42 lags	**42.5962(0.0802)**	**27.5352(0.5429)**	**27.5923(0.4322)**
49 lags	**49.4579(0.1009)**	**34.0471(0.5618)**	**34.0796(0.4639)**
AIC	-1.7839	-1.7877	-1.7857
BIC	-1.7497	-1.7478	-1.7401

ARMA pour $z_t = y_t - x2 * y_{t-7}$, où $y_t = U_t - x1 * U_{t-7}$:

	ARMA (5,5)	ARMA (6,6)	ARMA (7,7)
Log-likelihood	–1019.8	–1014.1	–1014.1
Test Box-Pierce			
14 lags	16.0307(0.0003)		
21 lags	21.3190(0.0113)	**9.2125(0.2378)**	**9.2104(0.1010)**
28 lags	31.0543(0.0132)	**18.8866(0.1693)**	**18.9203(0.0905)**
35 lags	**33.8004(0.0681)**	**21.0343(0.4569)**	**21.0630(0.3333)**
42 lags	**42.4190(0.0659)**	**27.6494(0.4831)**	**27.5892(0.3790)**
49 lags	**49.2822(0.0853)**	**34.1190(0.5105)**	**34.0927(0.4149)**
AIC	-1.7829	-1.7867	-1.7847
BIC	-1.7459	-1.7440	-1.7362

Tableau 7 : Test *t* de ARMA(6,6) avec une saisonnalité pour le résidu MCO des log-rendements du marché PJM sur les variables de saisons

	Estimateur(θ_k)	Ecart-type(σ)	θ_k/σ	Significatif
Saisonnalité	0.0654	0.0227	2.8811	Oui
AR(1)	1.1710	0.2739	4.2753	Oui
AR(2)	-1.6755	0.3148	-5.3224	Oui
AR(3)	1.2821	0.4717	2.7180	Oui
AR(4)	-1.2016	0.3570	-3.3658	Oui
AR(5)	0.3165	0.3498	0.9048	Non
AR(6)	0.0648	0.0961	0.6743	Non
MA(1)	-1.8163	0.2768	-6.5618	Oui
MA(2)	2.2930	0.4894	4.6853	Oui
MA(3)	-2.2817	0.6134	-3.7198	Oui
MA(4)	1.8489	0.6047	3.0575	Oui
MA(5)	-1.0297	0.4861	-2.1183	Oui
MA(6)	0.0417	0.2632	0.1584	Non
Constante	0.0002	0.0006	0.3333	Non

Tableau 8 : Test *t* de ARMA(6,6) avec deux saisonnalités pour le résidu MCO des log-rendements du marché PJM sur les variables de saisons

	Estimateur(θ_k)	Ecart-type(σ)	θ_k/σ	Significatif
$1^{ère}$ saisonnalité	0.0776	0.2352	0.3299	Non
2^{e} saisonnalité	-0.0126	0.2332	-0.0540	Non
AR(1)	1.1678	0.2866	4.0747	Oui
AR(2)	-1.6759	0.3274	-5.1188	Oui
AR(3)	1.2798	0.4920	2.6012	Oui
AR(4)	-1.2032	0.3706	-3.2466	Oui
AR(5)	0.3156	0.3644	0.8661	Non
AR(6)	0.0619	0.0985	0.6284	Non
MA(1)	-1.8178	0.2891	-6.2878	Oui
MA(2)	2.2967	0.5110	4.4945	Oui
MA(3)	-2.2863	0.6406	-3.5690	Oui
MA(4)	1.8537	0.6314	2.9359	Oui
MA(5)	-1.0332	0.5075	-2.0359	Oui
MA(6)	-0.0436	0.2764	-0.1577	Non
Constante	0.0002	0.0006	0.3333	Non

Tableau 9 : ARMA avec deux saisonnalités pour les résidus MCO des log-rendements du marché Ontario sur les variables de saisons

	ARMA (1,1)	ARMA (2,2)	ARMA (3,3)	ARMA (4,4)	ARMA (5,5)
Log-likelihood	-355.3096	-350.4283	-347.3336	-347.0209	-345.5287
Test Box-Pierce					
7 lags	**2.9572(0.3983)**	**3.6616(0.0557)**			
14 lags	**15.3729(0.1190)**	**12.7247(0.1217)**	18.3193(0.0055)	17.8918(0.0013)	7.1367(0.0282)
21 lags	**19.7164(0.2890)**	**15.7350(0.3999)**	23.4578(0.0365)	23.1692(0.0167)	**8.9473(0.4422)**
28 lags	**25.3778(0.3855)**	**22.2187(0.4469)**	**30.2416(0.0660)**	30.2650(0.0349)	**14.4514(0.5651)**
35 lags	46.5168(0.0363)	42.7548(0.0480)	50.2803(0.0042)	49.5720(0.0024)	**34.0253(0.0648)**
42 lags	**48.9109(0.1106)**	**45.5958(0.1312)**	51.3233(0.0287)	50.7171(0.0190)	**36.9454(0.1787)**
49 lags	**54.5840(0.1549)**	**50.8210(0.1927)**	**56.6158(0.0531)**	55.8557(0.0392)	**42.3567(0.2509)**
AIC	-1.3642	-1.3760	-1.3805	-1.3736	-1.3715
BIC	-1.3214	-1.3160	-1.3033	-1.2793	-1.2601

Tableau 10: Test *t* de ARMA(5,5) avec deux saisonnalités pour les résidus MCO des log-rendements du marché Ontario sur les variables de saisons

	Estimateur(θ_k)	Ecart-type(σ)	θ_k/σ	Significatif
$1^{\text{ère}}$ saisonnalité	0.3521	0.0775	4.5432	Oui
2^{e} saisonnalité	-0.2936	0.0757	-3.8785	Oui
AR(1)	-2.4480	1.5402	-1.5894	Non
AR(2)	-2.1419	3.7771	-0.5671	Non
AR(3)	-0.5847	3.3308	-0.1755	Non
AR(4)	0.1196	0.9850	0.1214	Non
AR(5)	0.0222	0.1202	0.1847	Non
MA(1)	1.6926	1.5464	1.0945	Non
MA(2)	0.1278	2.6225	0.0487	Non
MA(3)	-1.4229	0.2179	-6.5301	Oui
MA(4)	-0.8631	2.1945	-0.3933	Non
MA(5)	-0.0016	1.3241	-0.0012	Non
Constante	0.0034	0.0138	0.2464	Non

Tableau 11 : ARMA avec deux saisonnalité avec la 2e saisonnalité de 14, 21, 28 et 35 pour les résidus MCO des log-rendements du marché Ontario sur les variables de saisons

La 2e saisonnalité de 14					
	ARMA (1,1)	ARMA (2,2)	ARMA (3,3)	ARMA (4,4)	ARMA (5,5)
Log likelihood	-353.7717	-349.2408	-346.6037	-346.3282	-343.1566
7 lags	**3.0188(0.2210)**				
14 lags	**15.7281(0.0728)**	**13.0331(0.0713)**	17.9508(0.0030)	17.3769(0.0006)	11.8472(0.0006)
21 lags	**18.9886(0.2693)**	**15.6172(0.3373)**	22.3588(0.0337)	21.9621(0.0153)	17.9953(0.0213)
28 lags	**24.2970(0.3875)**	**21.7783(0.4124)**	**28.7371(0.0702)**	28.6627(0.0378)	26.2274(0.0357)
35 lags	43.8079(0.0496)	**40.6544(0.0577)**	47.7977(0.0057)	47.0531(0.0033)	44.0591(0.0035)
42 lags	**46.3056(0.1404)**	**43.4584(0.1544)**	49.0614(0.0356)	48.4720(0.0237)	45.7532(0.0248)
49 lags	**51.1518(0.2134)**	**47.9647(0.2437)**	**53.8373(0.0707)**	**53.1235(0.0525)**	**49.9665(0.0609)**
AIC	-1.3664	-1.3768	-1.3794	-1.3723	-1.3771
BIC	-1.3150	-1.3082	-1.2936	-1.2694	-1.2571
La 2e saisonnalité de 21					
	ARMA (1,1)	ARMA (2,2)	ARMA (3,3)	ARMA (4,4)	ARMA (5,5)
Log likelihood	- 354.2372	-349.8799	-347.0806	- 346.7035	-343.1485
7 lags	**3.9549(0.1384)**				
14 lags	18.3150(0.0317)	15.9390(0.0257)	20.0131(0.0012)	19.0544(0.0003)	12.0164(0.0005)
21 lags	**20.6490(0.1924)**	**17.8708(0.2127)**	24.0741(0.0199)	23.1127(0.0103)	17.1373(0.0287)
28 lags	**24.9960(0.3505)**	**22.3325(0.3806)**	30.2292(0.0490)	29.8177(0.0277)	**24.6793(0.0544)**
35 lags	45.7199(0.0330)	42.6249(0.0378)	50.5978(0.0027)	49.2037(0.0018)	42.3036(0.0057)
42 lags	**48.1240(0.1042)**	**45.3896(0.1123)**	51.8099(0.0197)	50.5804(0.0146)	43.9926(0.0368)
49 lags	**54.6200(0.1310)**	**51.6734(0.1456)**	57.8286(0.0337)	56.3872(0.0277)	**49.1435(0.0709)**
AIC	-1.3645	-1.3742	-1.3774	-1.3708	-1.3771
BIC	-1.3131	-1.3056	-1.2917	-1.2679	-1.2571
La 2e saisonnalité de 28					
	ARMA (1,1)	ARMA (2,2)	ARMA (3,3)	ARMA (4,4)	ARMA (5,5)
Log likelihood	- 350.2895	-346.0373	-343.9012	-343.5782	- 341.4832
7 lags	**3.8274(0.1475)**				
14 lags	**16.7816(0.0522)**	14.8512(0.0380)	18.9551(0.0020)	18.3906(0.0004)	13.9378(0.0002)
21 lags	**20.0946(0.2160)**	**17.1671(0.2474)**	22.6494(0.0309)	22.0969(0.0146)	18.2130(0.0197)
28 lags	**24.2092(0.3923)**	**21.2375(0.4445)**	**28.3617(0.0767)**	28.3947(0.0405)	**24.7720(0.0531)**
35 lags	**36.9033(0.1800)**	**34.3549(0.1894)**	41.7639(0.0260)	41.0623(0.0164)	35.5766(0.0337)
42 lags	**39.2319(0.3701)**	**37.1648(0.3696)**	**43.3472(0.1073)**	**42.6824(0.0789)**	**37.3880(0.1366)**
49 lags	**44.4416(0.4530)**	**42.0383(0.4693)**	**47.9476(0.1816)**	**47.1091(0.1476)**	**41.6819(0.2373)**
AIC	-1.3807	-1.3899	-1.3904	-1.3836	-1.3840
BIC	-1.3292	-1.3213	-1.3047	-1.2807	-1.2639
La 2e saisonnalité de 35					
	ARMA (1,1)	ARMA (2,2)	ARMA (3,3)	ARMA (4,4)	ARMA (5,5)
Log likelihood	- 351.3015	-347.1594	-344.4318	-340.8676	- 340.7066
7 lags	**2.9363(0.2304)**				
14 lags	**16.6588(0.0543)**	14.8930(0.0374)	18.9552(0.0020)	18.6352(0.0003)	18.3171(0)
21 lags	**20.5502(0.1965)**	**17.6608(0.2227)**	23.3286(0.0251)	22.6227(0.0122)	22.0918(0.0047)
28 lags	**25.4626(0.3269)**	**22.9131(0.3486)**	**29.4431(0.0593)**	28.9680(0.0348)	28.1491(0.0207)
35 lags	**37.5596(0.1614)**	**35.3013(0.1613)**	42.5376(0.0216)	41.3057(0.0154)	40.3590(0.0098)
42 lags	**39.6766(0.3516)**	**37.5919(0.3513)**	**43.8683(0.0978)**	**42.7690(0.0776)**	**41.8267(0.0582)**
49 lags	**45.4073(0.4132)**	**42.9950(0.4284)**	**48.7869(0.1606)**	**47.7354(0.1338)**	**46.8152(0.1070)**
AIC	-1.3765	-1.3853	-1.3883	-1.3947	-1.3871
BIC	-1.3251	-1.3167	-1.3025	-1.2918	-1.2671

Tableau 12 : Les meilleurs modèles du marché NE

Faisons d'abord la régression MCO des log-rendements sur les variables de mois, puis estimons ARMA (5,5) sans saisonnalité, ARMA (5,5) avec une saisonnalité et ARMA (6,6) avec une saisonnalité pour les résidus :

	ARMA (5,5)	ARMA (5,5) avec une saisonnalité	ARMA (6,6) avec une saisonnalité
Log-likelihood	-100.3247	-98.3968	-98.2256
Test Box-Pierce			
14 lags	10.5221(0.0325)	8.7862(0.0323)	8.5744(0.0034)
21 lags	40.5345(0)	37.1496(0.0001)	36.0332(0)
28 lags	49.2841(0.0001)	45.8510(0.0002)	44.2877(0.0001)
35 lags	59.6535(0.0001)	56.6074(0.0002)	55.0244(0.0001)
42 lags	67.3892(0.0003)	63.7157(0.0005)	61.8707(0.0004)
49 lags	73.6618(0.0007)	69.8982(0.0012)	68.0074(0.0010)
AIC	-2.6836	-2.6849	-2.6823
BIC	-2.6434	-2.6410	-2.6312

Tableau 13 : Les meilleurs modèles du marché NYA

Faisons d'abord la régression MCO des log-rendements sur les variables de mois, puis estimons ARMA avec deux saisonnalités pour les résidus :

ARMA avec deux saisonnalités :

	ARMA (5,5)	ARMA (6,6)
Log-likelihood	249.5282	251.7475
Test Box-Pierce		
14 lags	**5.7509(0.0564)**	
21 lags	**15.7350(0.0726)**	18.1308(0.0114)
28 lags	**23.2606(0.1069)**	26.0534(0.0255)
35 lags	43.3947(0.0062)	45.5691(0.0015)
42 lags	50.7596(0.0103)	53.5661(0.0025)
49 lags	56.1039(0.0228)	58.7403(0.0072)
AIC	-3.1867	-3.1871
BIC	-3.1368	-3.1294

Tableau 14 : Les meilleurs modèles du marché NYG

Faisons d'abord la régression MCO des log-rendements sur les variables de mois, puis estimons ARMA avec deux saisonnalités pour les résidus :

ARMA avec deux saisonnalités :

	ARMA (5,5)	ARMA (6,6)	ARMA (7,7)	ARMA (8,8)
Log-likelihood	215.8331	220.1093	240.8838	243.6720
Test Box-Pierce				
14 lags	31.1257(0)			
21 lags	39.2542(0)	32.8795(0)	24.5546(0.0002)	23.7866(0)
28 lags	48.1902(0)	39.7561(0.0003)	35.3778(0.0004)	32.1757(0.0004)
35 lags	57.2613(0.0001)	47.1416(0.0009)	40.6195(0.0027)	36.5927(0.0038)
42 lags	75.3574(0)	62.9736(0.0002)	58.2708(0.0003)	55.6785(0.0003)
49 lags	109.5866(0)	94.1827(0)	73.0153(0.0001)	69.5951(0.0001)
AIC	-3.1370	-3.1404	-3.1681	-3.1693
BIC	-3.0871	-3.0827	-3.1027	-3.0962

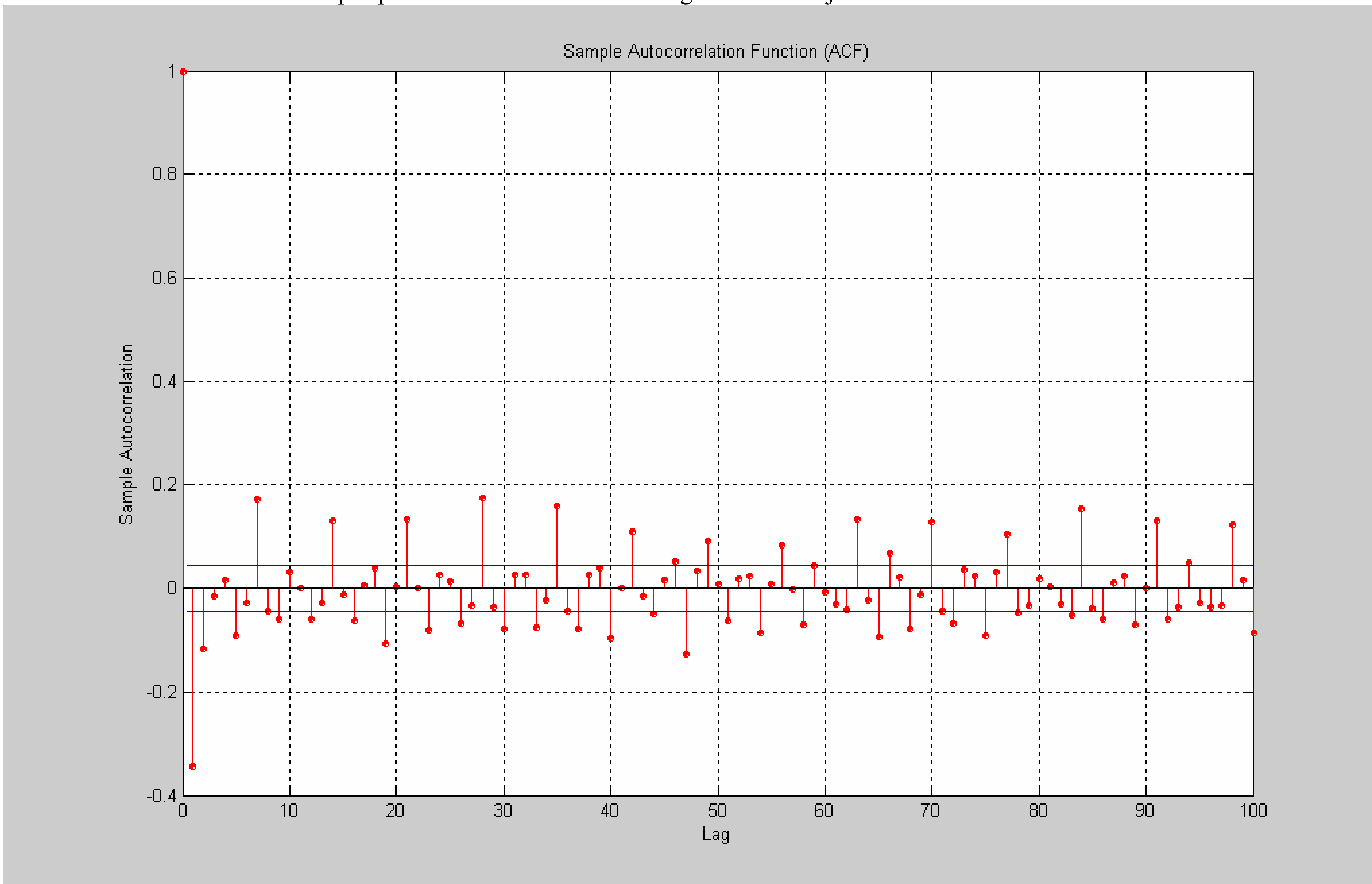
Tableau 15 : Les meilleurs modèles du marché NYJ

Faisons d'abord la régression MCO des log-rendements sur les variables de mois, puis estimons ARMA avec deux saisonnalités pour les résidus :

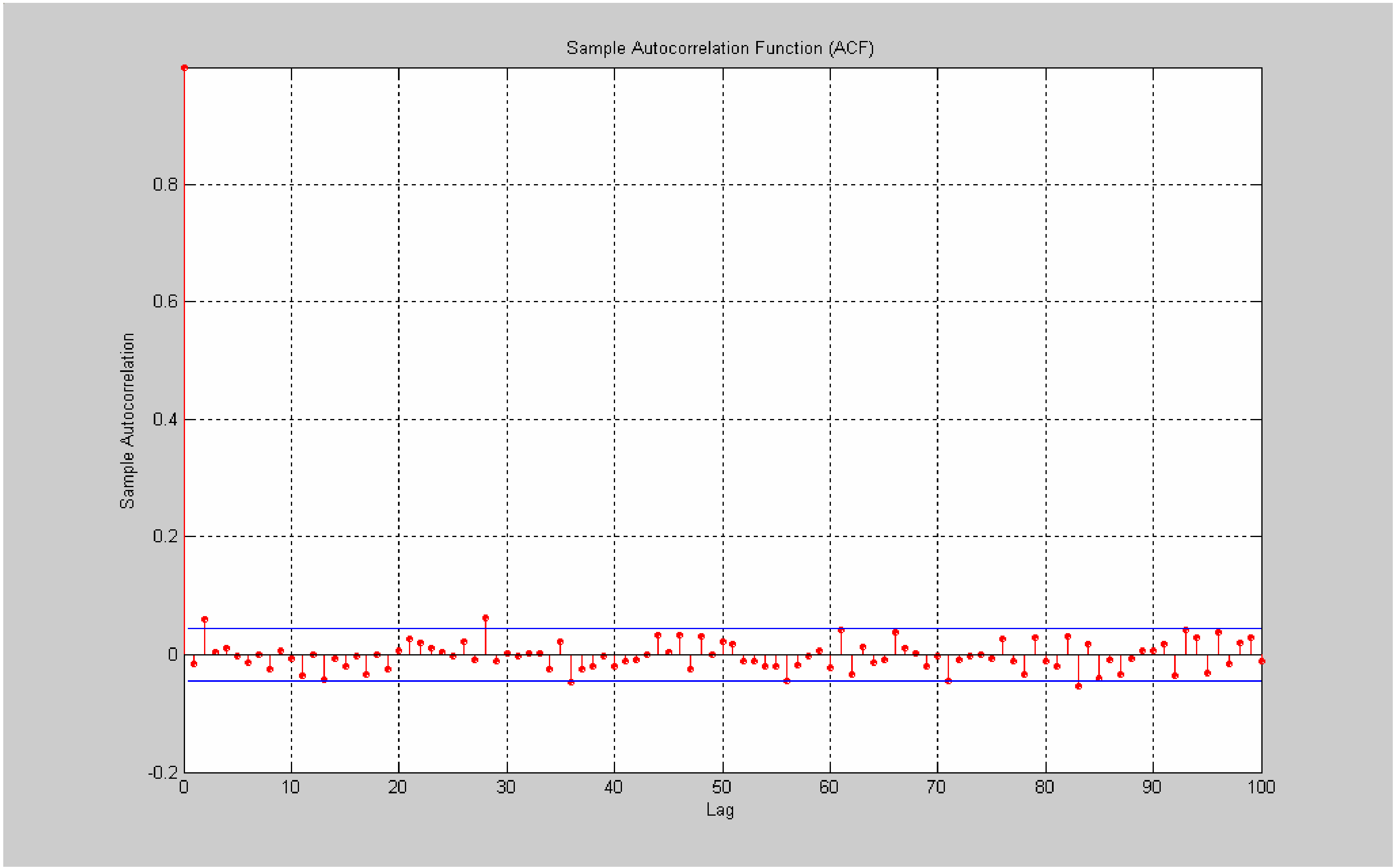
ARMA avec deux saisonnalités :

	ARMA (5,5)	ARMA (6,6)	ARMA (7,7)	ARMA (8,8)
Log-likelihood	176.1671	179.7282	194.6828	195.6893
Test Box-Pierce				
14 lags	18.4372(0.0001)			
21 lags	27.8623(0.0010)	22.1703(0.0024)	15.4728(0.0085)	14.4674(0.0023)
28 lags	37.5099(0.0018)	29.0444(0.0103)	21.8469(0.0393)	20.0978(0.0283)
35 lags	56.7384(0.0001)	45.2418(0.0016)	34.0835(0.0180)	31.9030(0.0155)
42 lags	65.1484(0.0002)	54.0027(0.0022)	46.9785(0.0071)	44.3423(0.0070)
49 lags	73.6675(0.0003)	62.1410(0.0032)	52.3319(0.0175)	49.7723(0.0177)
AIC	-3.0785	-3.0808	-3.0999	-3.0985
BIC	-3.0286	-3.0232	-3.0346	-3.0254

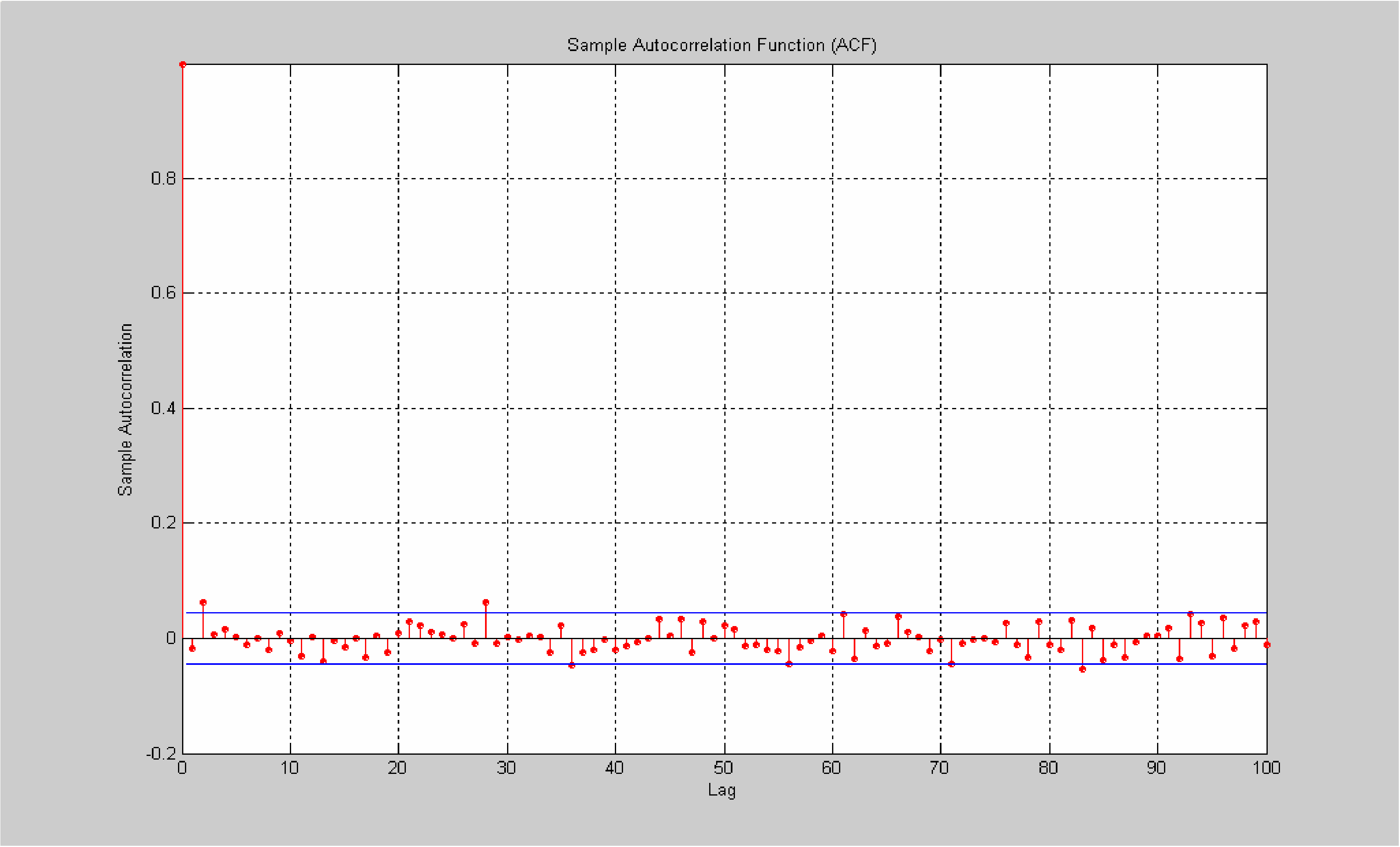
Graphique 1 : Autocorrélation des log-rendements journaliers du marché PJM



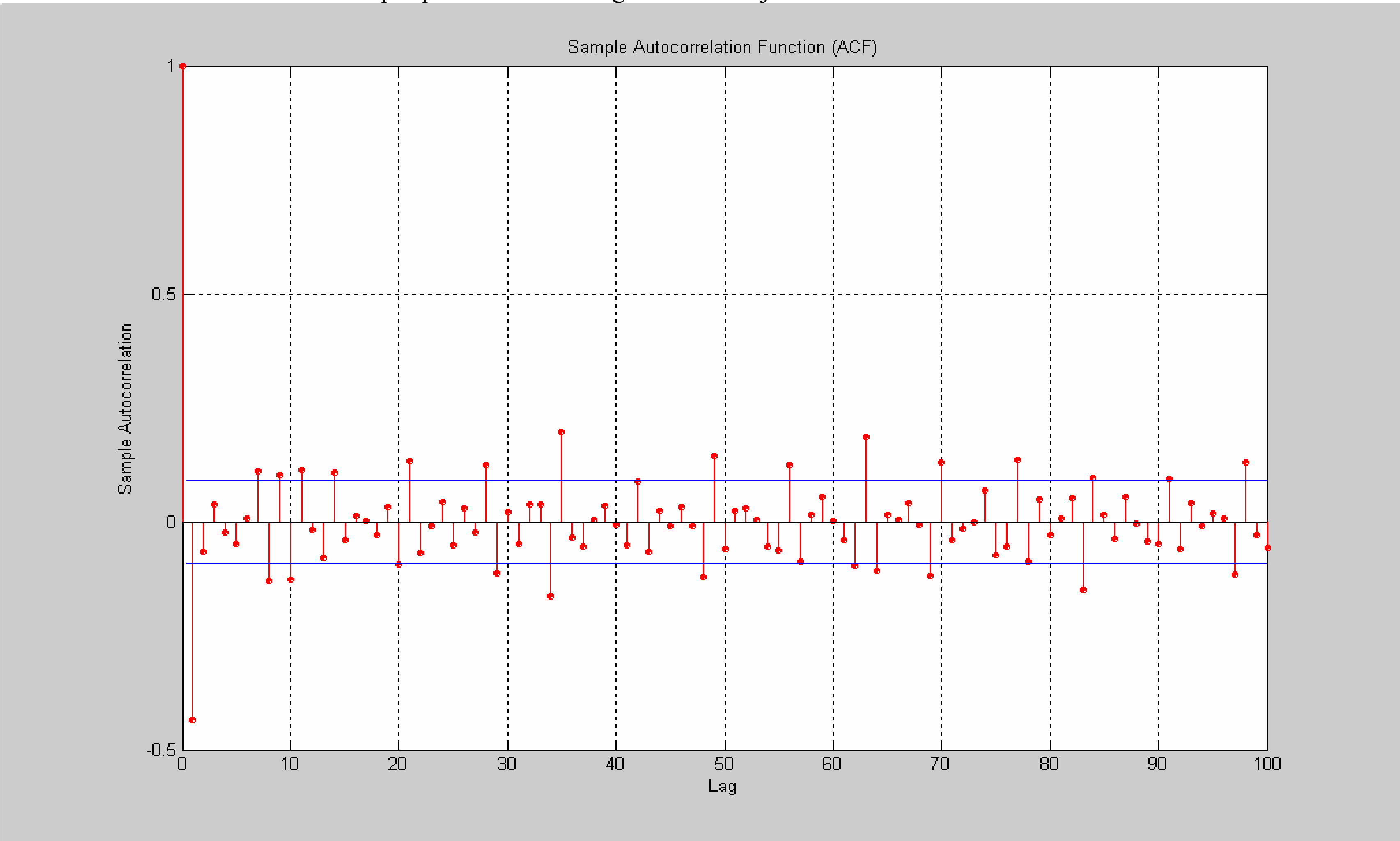
Graphique 2 : ACF du résidu de ARMA(5,5) avec une saisonnalité pour les log-rendements du marché PJM



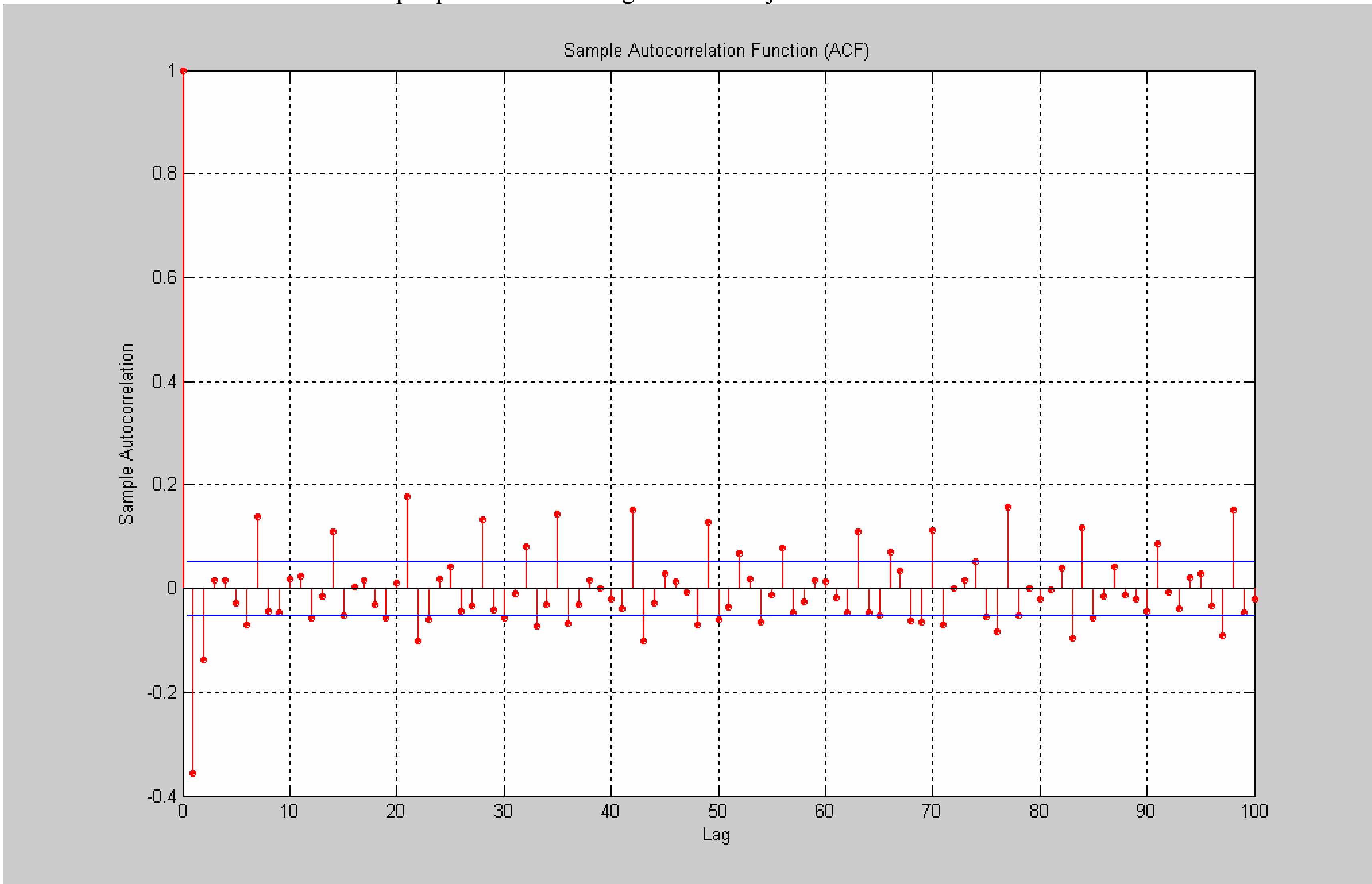
Graphique 3 : ACF du résidu de ARMA(5,5) avec deux saisonnalités pour les log-rendements du marché PJM



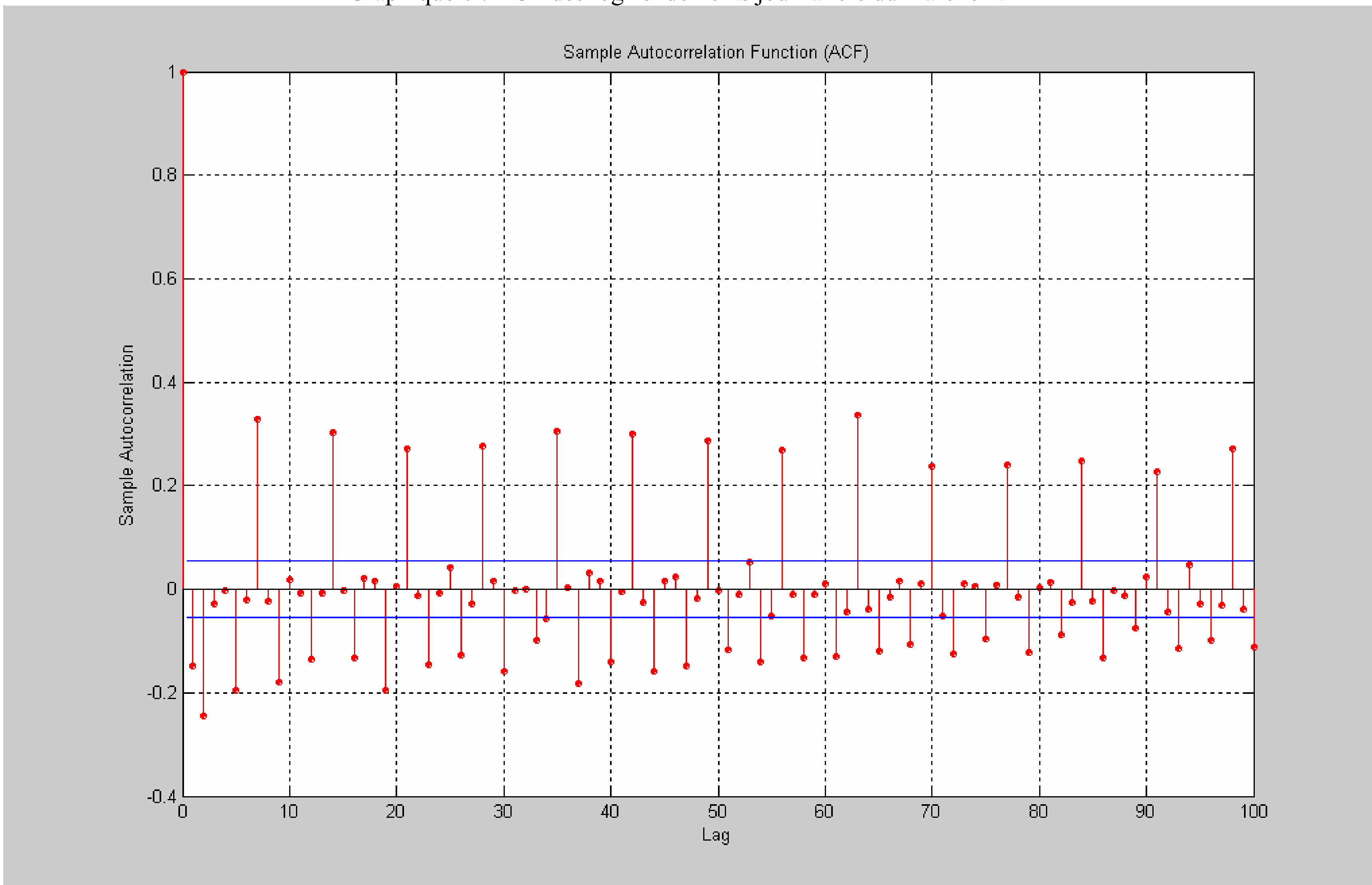
Graphique 4 : ACF des log-rendements journaliers du marché Ontario



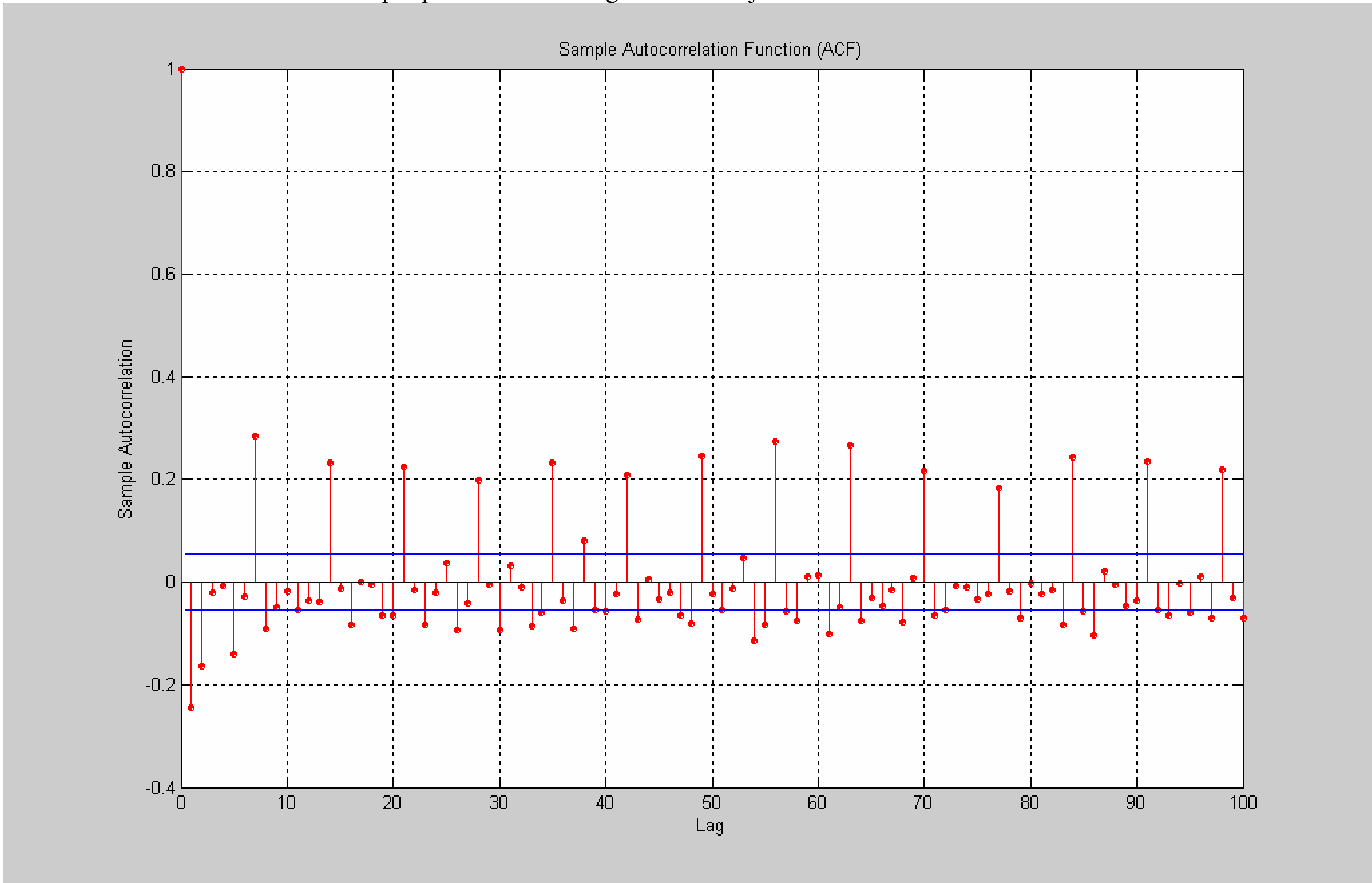
Graphique 5 : ACF des log-rendements journaliers du marché NE



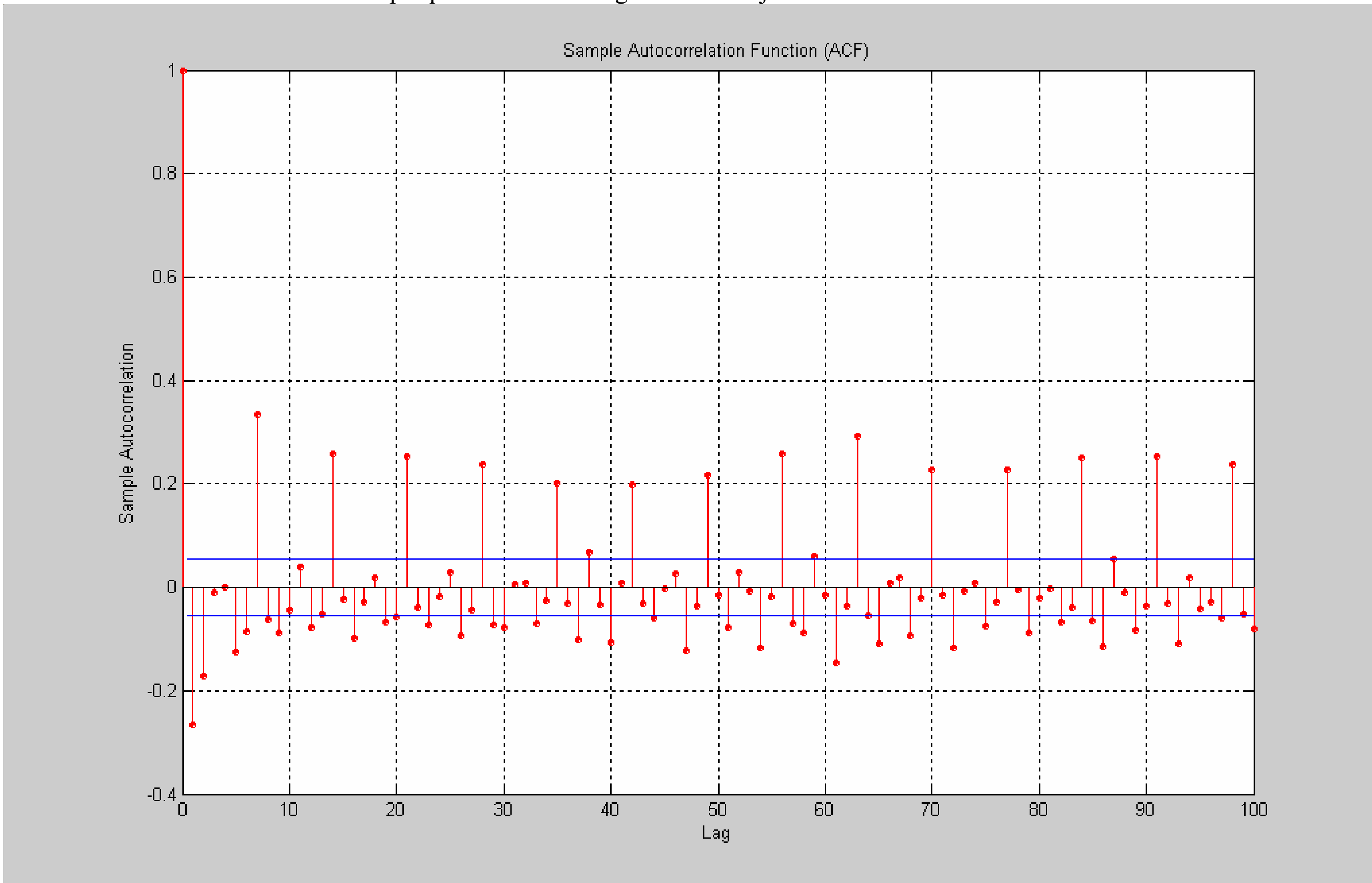
Graphique 6 : ACF des log-rendements journaliers du marché NYA



Graphique 7 : ACF des log-rendements journaliers du marché NYG



Graphique 8 : ACF des log-rendements journaliers du marché NYJ



Bibliographie

Francq, Chistian, Roy, Roch et Zakoïan, Jean-Michel, “Diagnostic Checking in ARMA Models With Uncorrelated Errors”, *Journal of the American Statistical Association*, vol.100, 2005, 532-544.

Gouriéroux, Christian, *ARCH Models and Financial Applications*, 1st ed., Springer Verlag, 1997.

Werner, Andreas, “Risk Measurement in the Electricity Market”, A thesis submitted in partial fulfillment for the M.Sc. in Mathematical Finance, University of Oxford, 2002.

Wooldridge, Jeffrey M., *Introductory Econometrics: A Modern Approach*, 2nd ed., South-Western College Pub., 2003.

Annexe sur les données

Les données sont les prix horaires de chaque jour, utilisées par Mihaela Capra – l'étudiante de Maîtrise en finance mathématique et computationnelle à l'Université de Montréal – pendant son stage à Hydro-Québec du 5 mai au 29 août 2003. On prend les prix midi de chaque jour comme prix journaliers.

Le marché PJM, ouvert en 1997, couvre une partie ou toute la région de Delaware, Illinois, Indiana, Kentucky, Maryland, Michigan, New Jersey, Caroline du Nord, Ohio, Pennsylvanie, Tennessee, Virginie, Virginie-Occidentale et District de Columbia. Les données sont pour la période du 1 avril 1998 au 11 août 2003, comprenant 1959 prix journaliers. En convertissant ces prix aux rendements journaliers, on compte 1958 observations dans l'analyse.

Les marchés de l'État de New York, depuis leur opération en 1999, sont divisés en 11 zones, telles que Long Island, ville New York, Dunwoodie, Millwood, vallée de l'Hudson, district capital, vallée du Mohawk, North Country, Central, Genesee, et West. Le West est désigné par zone A (NYA), la vallée de l'Hudson est désignée par zone G (NYG), et la ville New York est désignée par zone J (NYJ). L'échantillon de tous les marchés NYA, NYG et NYJ couvre la période du 18 novembre 1999 au 5 août 2003, comprenant 1357 prix journaliers, et donc 1356 rendements journalier dans l'analyse.

Le marché New England a commencé son opération depuis 1999, donnant son service aux États de Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island et Vermont. L'échantillon couvre la période du 18 mars 1999 au 28 février 2003, incluant 1444 prix journaliers, et alors 1443 rendements journaliers dans l'analyse.

Le marché Ontario n'est responsable que de la province Ontario, dont l'échantillon couvre la période du 1 avril 2002 au 3 août 2003, où il y a 490 prix journaliers et donc 489 rendements journaliers dans l'analyse.

Référence:

1. Capra, Mihaela, *Prévision pour la volatilité du marché de l'électricité*, Rapport de stage de l'Université de Montréal, septembre 2003.

2. Marché PJM: www.pjm.com

3. Marché New York: www.nyiso.com

4. Marché New England: www.iso-ne.com

5. Marché Ontario: www.ieso.ca

6. New York State Department of Public Service, *Staff Report on the State of Competitive Energy Markets: Progress to Date and Future Opportunities*, March 2, 2006.