

Projet Séries Temporelles

M1 MAS - 16/05/2021

Sommaire

Introduction	2
Partie 1: Analyse et traitement de la série	3
Partie 2: Prévion sur un horizon d'un an	11
Partie 3: Analyse plus poussée	19
Conclusion	23
Liste des tableaux et figures	24
Annexe	25
Bibliographie	28

Introduction

Le thème de notre étude est l'industrie extractive sur le marché intérieur de l'Union Européenne entre 1999 et 2020.

L'industrie extractive correspond au secteur de l'industrie travaillant dans l'extraction de produits minéraux présents à l'état naturel sous forme solide (houille et minerais), liquide (pétrole) ou gazeuse (gaz naturel).

La série temporelle étudiée représente les relevés mensuels de l'indice de chiffre d'affaires (CA) de l'industrie extractive française du 1er Janvier 1999 au 1er Janvier 2020.

Nous allons dans un premier temps analyser et traiter la série. En effet, toute série temporelle doit être traitée avant de pouvoir effectuer des prévisions. En particulier, la non-stationnarité de la série doit être traitée afin de ne pas tomber dans un risque de régression fallacieuse. Ensuite nous comparerons différents modèles de prévision, naïf, lissage exponentiel et arima-sarima. Enfin nous intégrerons une seconde série à notre modèle pour étudier une relation d'équilibre entre deux phénomènes.

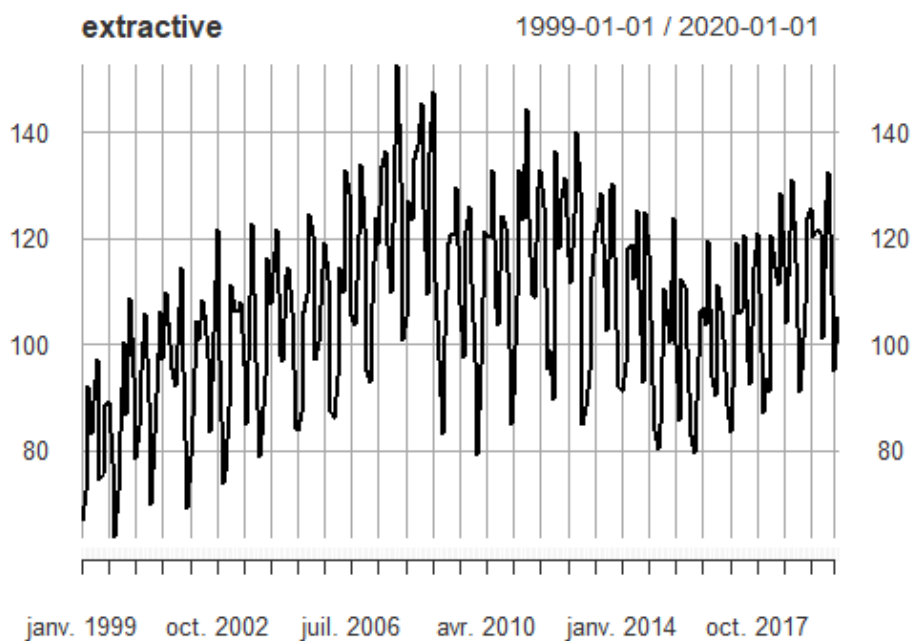
L'étude a été réalisée grâce au logiciel RStudio.

Partie 1: Analyse et traitement de la série

a) Analyse préliminaire de la série

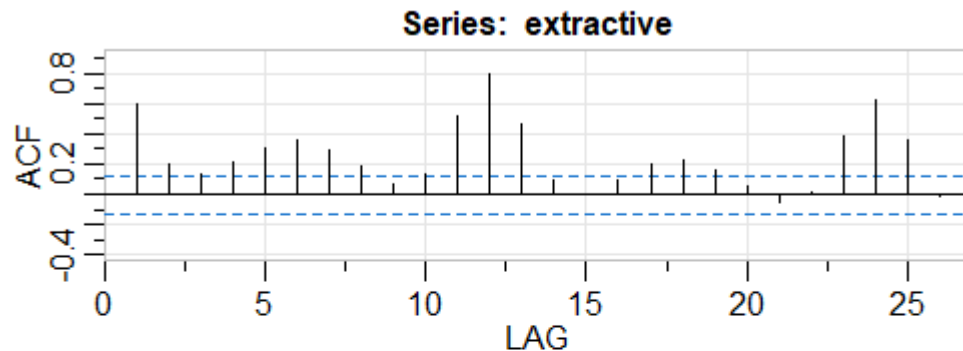
Avant toute chose, on doit prendre connaissance de notre série, que nous avons appelée **extractive**, telle quelle. Le premier point positif est que nous n'avons pas de valeurs manquantes dans notre série.

Figure 1: Evolution de l'industrie extractive



L'observation de la courbe nous montre une série qui semble non-stationnaire en moyenne, avec une tendance à la hausse. On remarque toutefois une légère rupture de niveau dans la courbe, qui semble apparaître aux alentours de 2008. Elle reste peu marquée et le résultat du test de rupture nous le confirme d'ailleurs (voir *Table 0*, Annexe, p25).

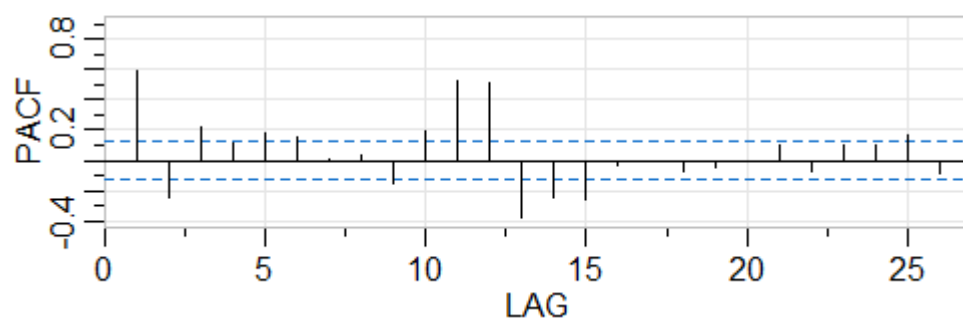
Figure 2: Fonction d'autocorrélation



La fonction d'autocorrélation est représentée dans ce que l'on appelle un corrélogramme et est visible ci-dessus (*Figure 2*). Elle mesure la corrélation, c'est-à-dire la relation, de la série avec elle-même, décalée de h périodes (ici on peut la mesurer de 1 à 25 périodes). On dit que l'autocorrélation est significative lorsqu'elle sort de l'intervalle en pointillé bleu. Par exemple, la pointe présente au décalage 1 nous indique une forte corrélation entre chaque valeur de série et la valeur précédente. Des pics sont également visibles autour du 12ème mois de chaque année, sous-entendant une saisonnalité que l'on tentera de définir plus tard.

On peut observer un motif qui s'atténue, avec une décroissance légère des termes. Nous devons toutefois attendre avant de donner des conclusions trop hâtives sur la saisonnalité et déterminer avant tout le type de la série.

Figure 3: Fonction d'autocorrélation partielle

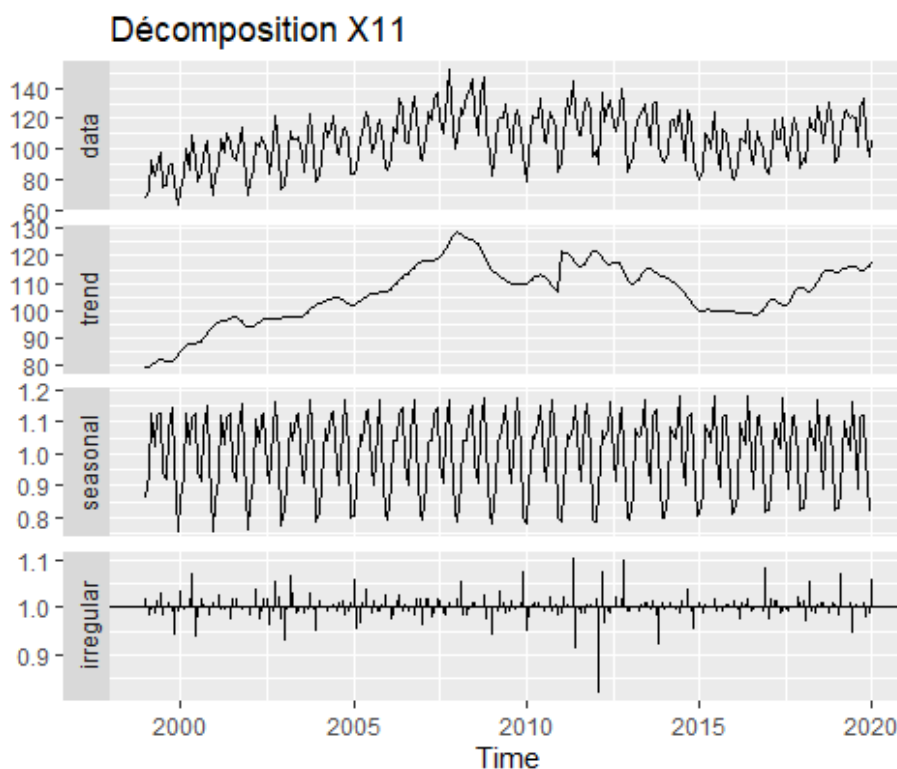


La fonction d'autocorrélation partielle, *Figure 3* ci-dessus, mesure la même chose que la fonction d'autocorrélation, exceptée que cette fois-ci l'influence des périodes

entre les deux décalages qui nous intéressent est retirée. On retrouve logiquement la pointe qui s'est déjà avérée significative au décalage 1. En ce qui concerne le reste de la fonction, on observe un motif, moins marqué que pour la FA, qui alterne corrélations positives et négatives. On remarque que la saisonnalité qui était clairement visible dans la FA se retrouve ici autour du 12ème décalage, mais pas pour le 24ème. La relation négative observée entre le 12ème et le 13ème décalage suggère une baisse de l'extraction des minéraux dans l'industrie extractive entre les mois de décembre et de janvier. Tandis que les relations positives à la fin de l'année suggèrent l'inverse. Autrement dit, la période "chaude" d'extraction de minéraux semble se dérouler à la fin de chaque année de façon saisonnière.

On note **ts_extr** la série *extractive* transformée en objet ts.

Figure 4: Schéma de décomposition



La figure 4 page précédente regroupe quatre graphiques décomposant notre série temporelle. Le premier correspond aux données telles qu'elles. Le second illustre la

tendance que suit l'industrie extractive depuis 1999, une tendance qui s'est avérée à la hausse au premier abord mais qui semble stagner depuis les années 2010. La saisonnalité de notre série est quant à elle confirmée dans le troisième graphique, dans lequel on peut voir un schéma saisonnier qui se répète chaque année. Le dernier graphique montre les éventuelles irrégularités de notre série, et nous confirme qu'il n'y a pas de valeurs aberrantes, exceptées peut-être aux alentours de 2013.

b) Test de Dickey-Fuller - Nature de la tendance

On veut confirmer la non-stationnarité de notre série *extractive* et déterminer la nature de la tendance, qui peut être stochastique (DS) ou déterministe (TS). Dans le premier cas, le temps intervient dans la moyenne, la variance et la covariance de la série temporelle. Cela signifie que si un choc vient bouleverser la tendance de la série, il va avoir des conséquences irréversibles sur celle-ci. Dans le cas d'une tendance déterministe, c'est l'inverse que l'on observe: le comportement de long terme de la série temporelle est alors entièrement prévisible. Pour ce faire, on utilise le **test ADF, ou test de Dickey-Fuller avec correction paramétrique** (pour tenir compte de l'autocorrélation) au niveau 5% qui pose les hypothèses suivantes:

H_0 : La série comporte une RU ¹.

H_1 : La série ne comporte pas de RU. La série est stationnaire.

Les modèles à estimer sont:

Modèle 1 (sans constante ni tendance): $\Delta X_t = \phi X_{t-1} + \sum_{j=1}^p w_j \Delta X_{t-j} + \varepsilon_t$

Modèle 2 (constante seulement): $\Delta X_t = \phi X_{t-1} + \gamma + \sum_{j=1}^p w_j \Delta X_{t-j} + \varepsilon_t$

¹RU: Racine Unitaire

Modèle 3 (constante et tendance): $\Delta X_t = \phi X_{t-1} + \gamma + \delta t + \sum_{j=1}^p w_j \Delta X_{t-j} + \varepsilon_t$

Selon les conclusions des tests, on devra effectuer le **test de Racine Unitaire** (RU) qui pose lui des hypothèses inversées, à savoir:

H_0 : *La série est non-stationnaire (NS).*

H_1 : *La série est stationnaire.*

On adopte une stratégie séquentielle. Autrement dit, on commence par estimer le modèle général, soit le modèle 3 ci-dessus, puis le modèle 2, et enfin le modèle 1, en fonction de la significativité des coefficients testés.

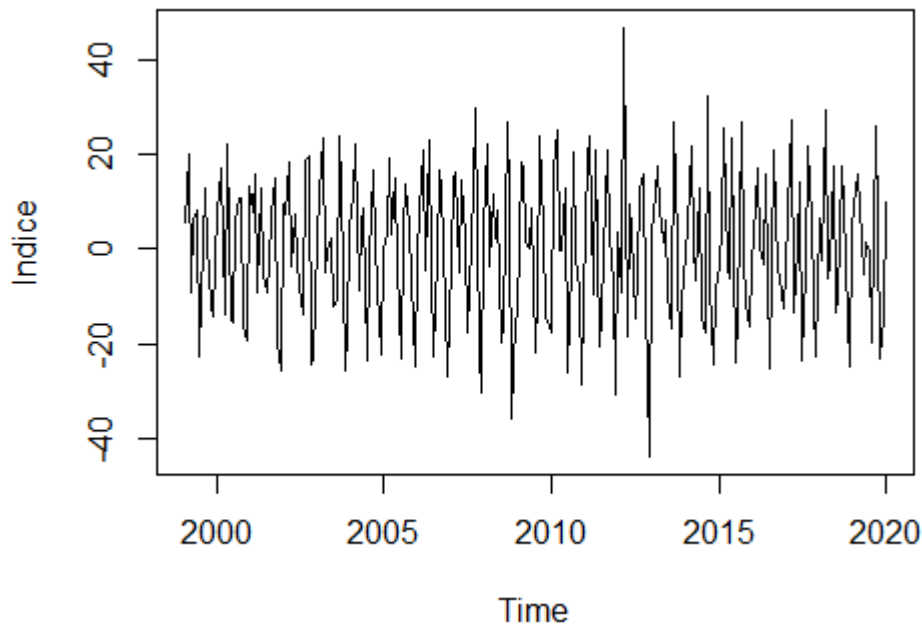
Dans notre cas, on commence par estimer le *modèle 3*.

On conclut que la tendance est non significative. On accepte donc H_0 et on passe au *modèle 2*. (voir *Table 1, Test a1*, de l'Annexe, p25)

En ce qui concerne le *modèle 2*, on constate que la constante est significative. On procède donc au test de RU qui nous fait accepter H_0 . Il y a présence de racine unité, la série est non-stationnaire.

Nous n'arrêtons pas la stratégie séquentielle tant que notre série n'est pas stationnaire. Ainsi, on différencie *extractive* pour enlever la racine unitaire et on applique de nouveau le test. On note **dext** la série *extractive* à laquelle on a appliqué le filtre de différence. Son évolution est visible à la page suivante, *Figure 5*.

Figure 5: Série temporelle de l'industrie extractive après première différenciation



On peut déjà observer une série qui semble stationnaire en moyenne. Il semble toutefois y avoir toujours les deux valeurs aberrantes vers 2013...

On recommence donc le test sur la série différenciée avec le *modèle 3*, et obtient les statistiques de test, concernant la tendance et la constante. On rejette l'hypothèse nulle de significativité de la tendance. On passe donc à l'*étape 2*. La constante est non-significative. On passe à l'*étape 3* et on arrive au *modèle 1 (sans constante ni tendance)*, pour lequel on effectue directement le test de RU. On rejette H_0 . La série est stationnaire.

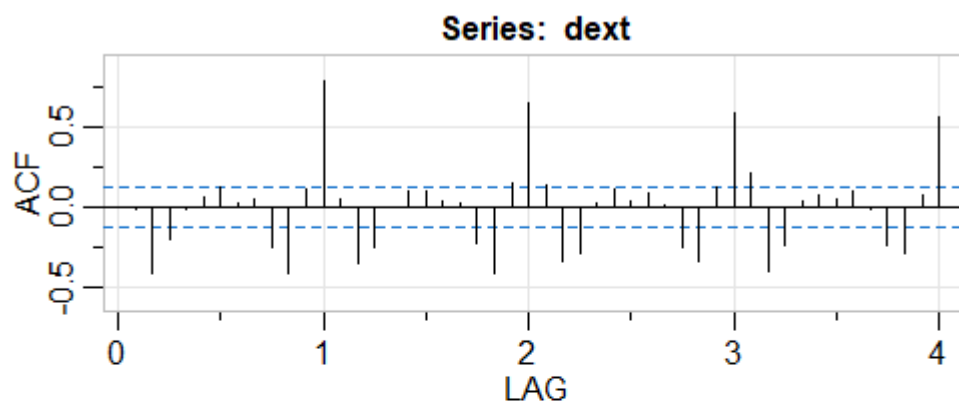
Les résultats des tests sont disponibles à la *Table 1, Test a2* de l'Annexe, p25.

En conclusion, on a une série intégrée d'ordre 1 de type DS ("Difference Stationary"), qui a alors une tendance stochastique (aléatoire). On conserve alors la série *dext* qui se charge d'éliminer la tendance de notre série initiale en la différenciant.

c) Saisonnalité

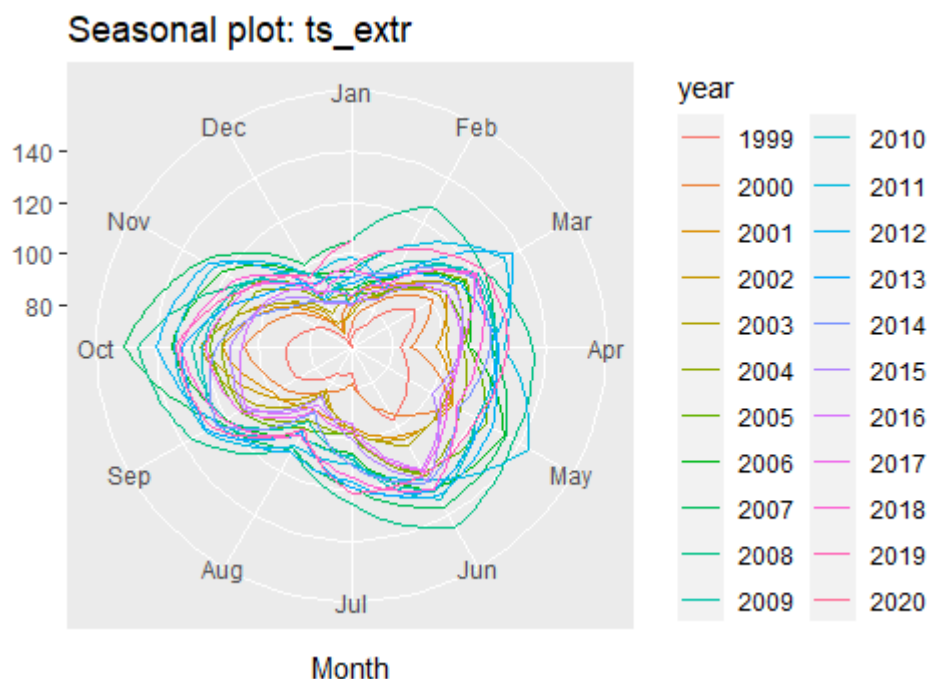
En présence de tendance, il faut tracer la FA de la série en différence première pour faire ressortir les saisonnalités.

Figure 6: Fonction d'autocorrélation de dext



On voit bien que l'autocorrélation est significativement très élevée pour chaque année.

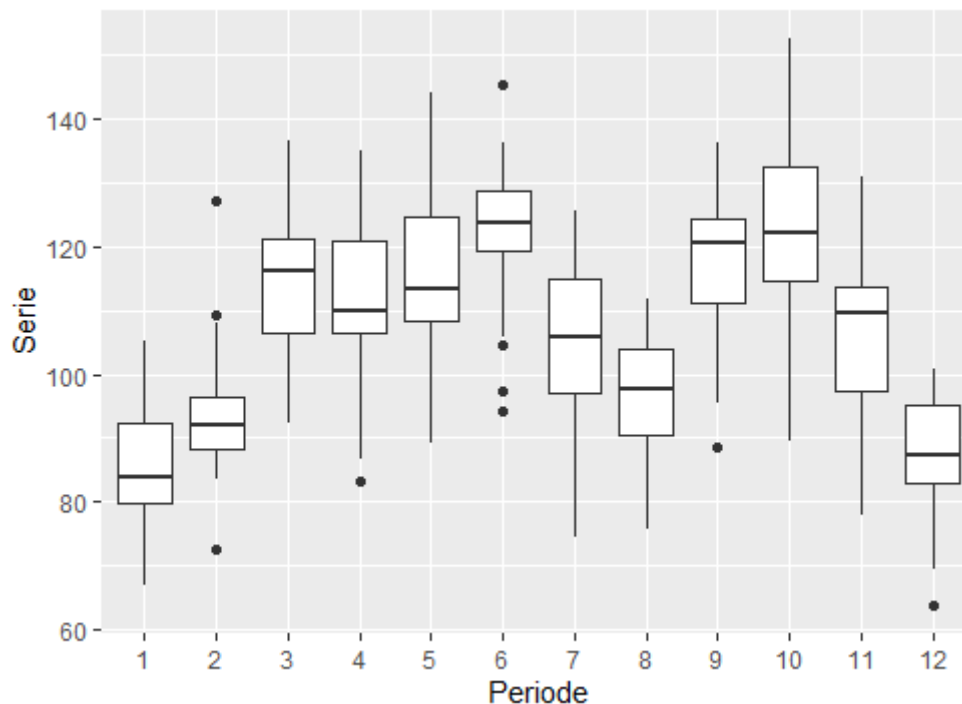
Figure 7: Graphique de saisonnalité annuelle



Le graphique ci-dessus fait ressortir un mixte d'une saisonnalité multiplicative et additive. En effet, le schéma, représenté par les cercles, croît fortement les dix

premières années, puis décroît et se stabilise. Cela est probablement dû à l'épuisement progressif des ressources naturelles, qui ne sont malheureusement pas disponibles en quantité infinie, et à l'exploitation croissante de nouvelles ressources, telles que les énergies renouvelables. L'innovation technologique est également à l'origine de ce phénomène.

Figure 8: Répartition mensuelle de l'indice de CA de l'industrie extractive



En figure 8, nous pouvons observer la répartition des valeurs pour chaque mois. On peut en un coup d'œil détecter la présence de valeurs aberrantes. On note par exemple peu de variance pour le mois de Juin mais plusieurs valeurs aberrantes. De plus, ce graphique nous permet d'avoir un aspect plus visuel de l'autocorrélation et d'appuyer nos propos du début, à savoir que les mois de décembre et janvier sont les mois les moins fructueux, et que l'indice de CA suit la même évolution chaque année.

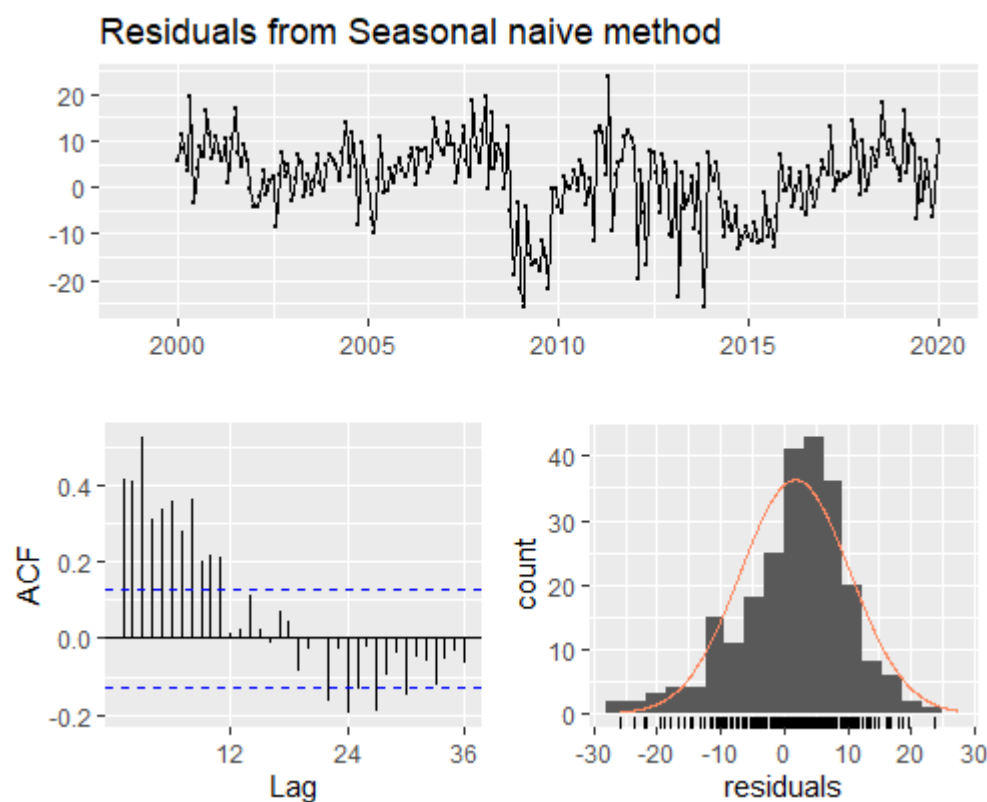
Partie 2: Prédiction sur un horizon d'un an

a) Modèle de référence (modèle naïf)

La série étudiée est saisonnière. On va utiliser la fonction *snaive()* du package *forecast*. Celle-ci utilise la méthode naïve qui consiste à définir la valeur suivante selon la valeur de la saison précédente.

Elle inclut “une composante aléatoire pour simuler des occurrences futures selon des marches aléatoires, où chaque valeur suivante est simulée aléatoirement, considérant la valeur précédente.”

Figure 9: Résidus provenant du modèle naïf saisonnier



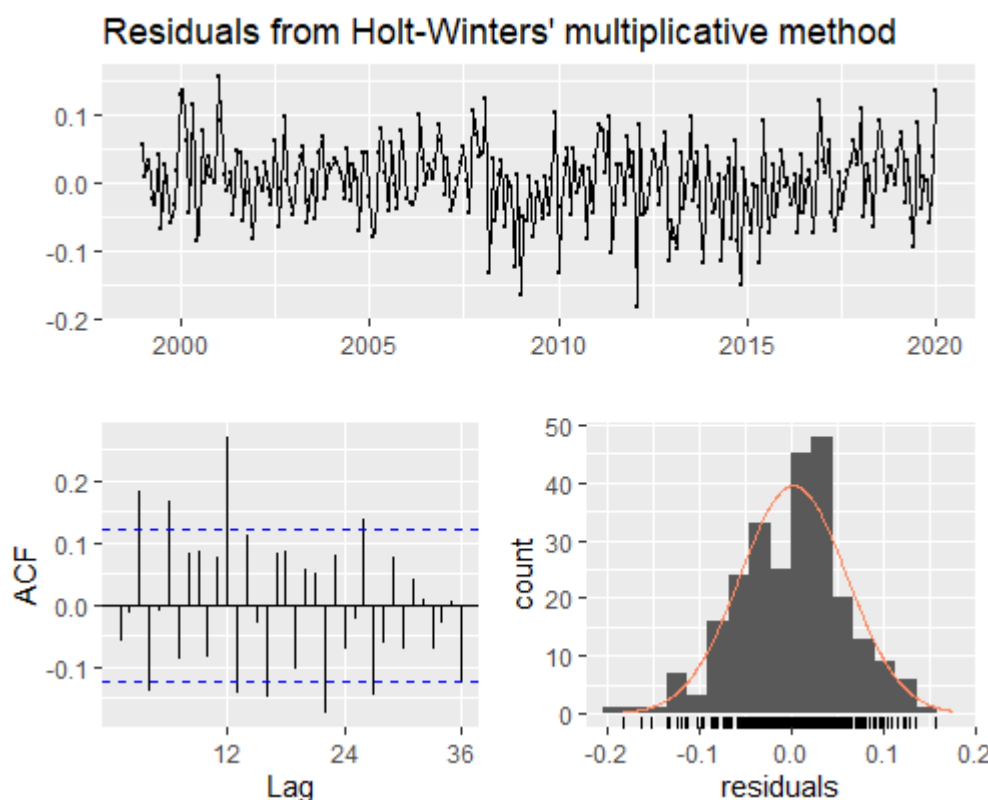
¹: Voir Bibliographie p28, *Les séries temporelles, analyse et modélisation d'agorécosystèmes*, S. Parent, 2020.

b) Modèle de lissage exponentiel

Le lissage exponentiel consiste à modéliser une série en fonction de ses informations passées, en accordant plus d'importance aux valeurs récentes.

Dans notre cas de série saisonnière avec tendance, c'est le lissage exponentiel de Holt-Winters qu'on utilise. La fonction d'autocorrélation, la distribution ainsi que la série des résidus du modèle sont visibles dans la Figure 10 ci-dessous. Ces graphiques suggèrent que les résidus suivent une loi normale, et semblent néanmoins autocorrélés. Nous définirons dans la section 2.d) plus précisément la signification de la fonction d'autocorrélation, ainsi que les hypothèses que devraient suivre les résidus d'un modèle pour que celui-ci soit considéré comme un "bon" modèle.

Figure 10: Résidus provenant du modèle Holt-Winters multiplicatif



c) Comparaison Holt-Winters et Naïf

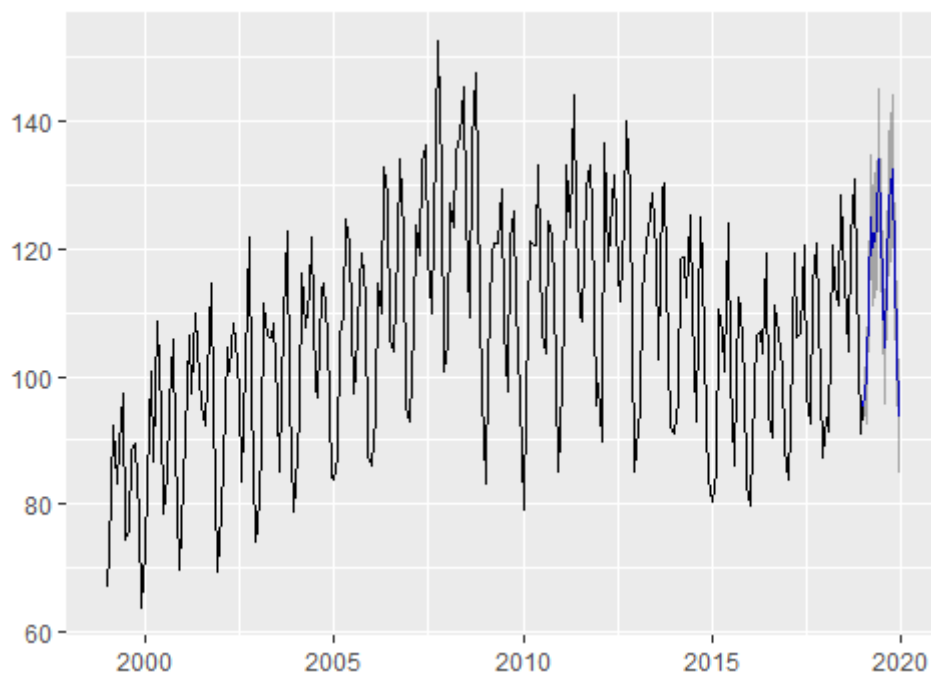
On compare nos deux modèles définis en a) et b) en évaluant la qualité de leur prévision temporelle.

Pour cela on crée un échantillon d'apprentissage et notre critère d'évaluation de la prévision sera le RMSE ("root-mean-square error" ou erreur quadratique moyenne), qui nous donne un indicateur sur la variabilité de la qualité de la prédiction.

$$RMSE(hw) = 6.51008 < RMSE(S-naïf) = 8.023571$$

On préfère le modèle de lissage exponentiel de Holt-Winters au modèle naïf. La prévision faite avec ce modèle donne:

Figure 11: Prévisions sur un an de l'indice de chiffre d'affaire de l'industrie extractive avec le modèle Holt-Winters multiplicatif



d) Modèle ARIMA (et SARIMA)

Dans cette section, nous allons utiliser la procédure Box-Jenkins pour attribuer un modèle de type ARIMA (ou SARIMA) à notre série temporelle.

En effet, notre série étant non-stationnaire, le processus ARMA (processus “AutoRegressive Moving Average”) ne suffit pas, et il faut utiliser son homologue le modèle ARIMA ou ARMA intégré qui fonctionne pour les séries non-stationnaires en moyenne. Appliquer un processus ARIMA revient à appliquer un processus ARMA(p,q) sur notre série différenciée $dext$.

Un modèle ARIMA est une combinaison de trois parties p,d,q où:

p : ordre du processus autorégressif, correspondant à la régression de la série sur elle-même par rapport aux périodes précédentes

q : ordre du processus de moyenne mobile

d : ordre d'intégration avant de procéder à l'autorégression et au calcul de la moyenne mobile

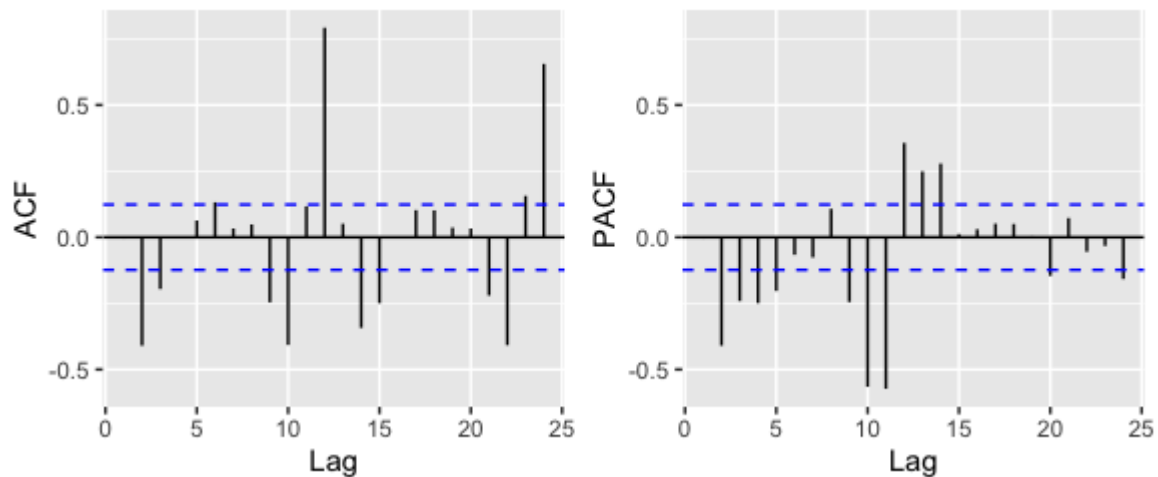
Dans notre cas, il suffit de différencier notre série une fois pour qu'elle devienne stationnaire, d'où $d=1$. Les deux autres paramètres sont déterminés ci-dessous.

A la différence du modèle ARIMA, le modèle SARIMA prend en compte la saisonnalité de la série.

La procédure Box-Jenkins se déroule en 4 étapes. L'identification du modèle est la première étape que nous allons commencer maintenant.

i. Identification

Figure 12: Corrélogramme de la série dext



Tenter de modéliser notre série temporelle avec un modèle ARIMA, d'après le corrélogramme de la série différenciée ci-dessus, reviendrait à choisir les deux modèles suivant: une moyenne mobile d'ordre 2 (MA(2)) ainsi qu'une partie autorégressive (AR) d'ordre 3 ou 4. En effet, les deux premières pointes significatives de la fonction d'autocorrélation suggèrent un MA(2) tandis que les trois (ou quatre) premiers pics significatifs de la FA partielle suggèrent un AR(3), voire un AR(4). Un *ARIMA* (2, 1, 4) est donc un premier modèle à envisager qui, même s'il ne s'avèrera pas idéal par rapport à la saisonnalité de la série, nous servira en guise de comparaison de performance de prévision des différents modèles identifiés à la fin de la procédure !

Un autre point de vue serait d'inclure une partie saisonnière au modèle ARIMA. Le corrélogramme suggère ici une saisonnalité décroissante marquée par des pics tous les 12 mois du côté de la FA. Du côté de la FAP, un pic apparaît autour du décalage 11-12, le reste de la fonction étant non-significatif. On pourrait donc modéliser cette partie en incluant un ordre des processus AR et MA saisonniers d'au moins 2 du côté de la moyenne mobile et de 1 du côté autorégressif, c'est-à-dire un $SARIMA_{(p,d,q)}(1, 1, 2)$.

Passons à la deuxième et troisième étape de la procédure Box-Jenkins: l'estimation des coefficients et la validation du modèle.

ii. Estimation et validation

Après avoir identifié les ordres p et q des modèles (S)ARIMA, la méthode de maximisation de vraisemblance est utilisée pour obtenir les estimateurs des coefficients des parties AR(p), MA(q), SAR(P) et SMA(Q) du modèle, ainsi que pour l'estimateur σ_{ϵ}^2 de la variance des résidus. Ensuite, il est primordial de vérifier, en appliquant des tests sur les paramètres et les résidus, si le modèle estimé respecte les critères suivants pour valider le modèle:

- ses coefficients sont significatifs
- ses résidus forment un bruit blanc

Tout d'abord, nous avons donc utilisé la méthode de maximum de vraisemblance pour le modèle ARIMA défini en \hat{I} , ainsi que pour définir le meilleur modèle SARIMA. Ce dernier étant plus difficilement identifiable avec le corrélogramme, nous avons décidé d'ajouter petit à petit des coefficients selon leur significativité (ou respectivement de les retirer).

Par conséquent, les étapes de validation, en partant du modèle $SARIMA_{(1,1,1)}(1, 1, 2)$ auquel on a associé une saisonnalité annuelle (décalage de 12), nous ont amené à sélectionner le modèle suivant: $SARIMA_{(0,1,1)}(2, 1, 3)$. En effet, il s'est révélé être le seul respectant les trois critères suivants:

- significativité de tous les coefficients au niveau 5% (voir *Table 3*, Annexe, p26)
- non-autocorrélation des résidus au niveau 10% (voir *Table 4*, Annexe, p26)
- normalité des résidus au niveau 5% (voir *Table 5*, Annexe, p26)

Les résidus de ce modèle sont donc très proches du comportement d'un bruit blanc.

Nous avons également jugé pertinent d'utiliser la fonction *auto.arima* de R pour construire un troisième modèle. Cette dernière sélectionne automatiquement le

modèle qu'elle juge le plus pertinent par rapport aux données. Elle nous recommande un $SARIMA_{(0,1,2)}(2, 1, 2)$ qui est un processus pour lequel, appliqué à notre série, les coefficients sont significatifs et les résidus sont autocorrélés mais se rapprochent d'une loi normale.

Finalement, nous avons choisi d'inclure une nouvelle fois le modèle naïf défini en a) pour la comparaison.

Nous nous retrouvons finalement avec les modèles suivants, à comparer dans le cadre de prévisions:

- $ARIMA (2, 1, 4)$
- $SARIMA_{(0,1,1)}(2, 1, 3)$
- $SARIMA_{(0,1,2)}(2, 1, 2)$
- modèle naïf saisonnier

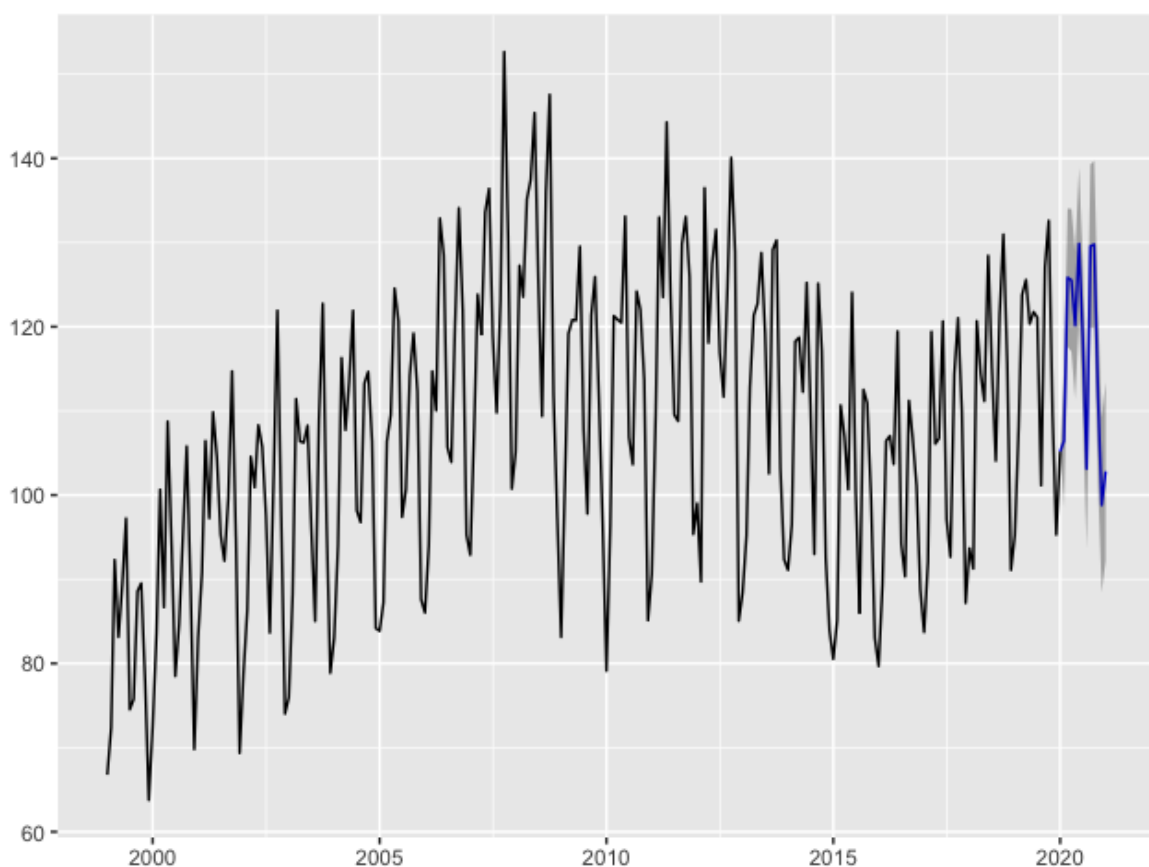
iv. Prévisions

Attribuer un modèle à une série temporelle a avant tout un but prédictif. Afin de mesurer la qualité d'une prévision, nous avons utilisé nos données de départ pour former deux échantillons: l'échantillon d'apprentissage et l'échantillon test. Les données d'apprentissage sont celles qui nous permettront d'effectuer des prévisions avec les différents modèles, tandis que les données test vont nous servir de vérification pour ces prévisions. Nous sommes intéressés par des prévisions sur une année, autrement dit par un horizon de $h=12$ mois, taille de l'échantillon test. Encore une fois nous utilisons le critère d'évaluation RMSE pour comparer les prédictions, qui sont visibles dans la table 6 de l'annexe, p27. Le modèle généré par *auto.arima* est celui qui minimise le RMSE, et est donc le "meilleur" ! Ce résultat est néanmoins très serré avec celui de $SARIMA_{(0,1,1)}(2, 1, 3)$, donc nous pouvons conclure que ces deux modèles se valent. Les modèles $ARIMA (2, 1, 4)$ et naïf sont quant à eux loin derrière.

Lorsque l'on vérifie les résidus des différentes prévisions, les résultats des tests de Ljung Box indiquent une forte autocorrélation (cf *Table 7*, Annexe, p25). Il est malheureusement courant en prévision, voire inévitable, d'obtenir un modèle dont les résidus ne forment pas un bruit blanc.

Au final, visuellement, les prévisions en utilisant le modèle sélectionné par le critère RMSE pour les mois de l'année 2020 sont représentées en bleu, dans la figure 13 qui suit.

Figure 13: Prévisions de l'indice de chiffre d'affaire de l'industrie extractive sur an avec le modèle $SARIMA_{(0,1,1)}(2,1,3)$



L'intervalle de confiance à 5% est visible en gris, autour de la ligne bleue.

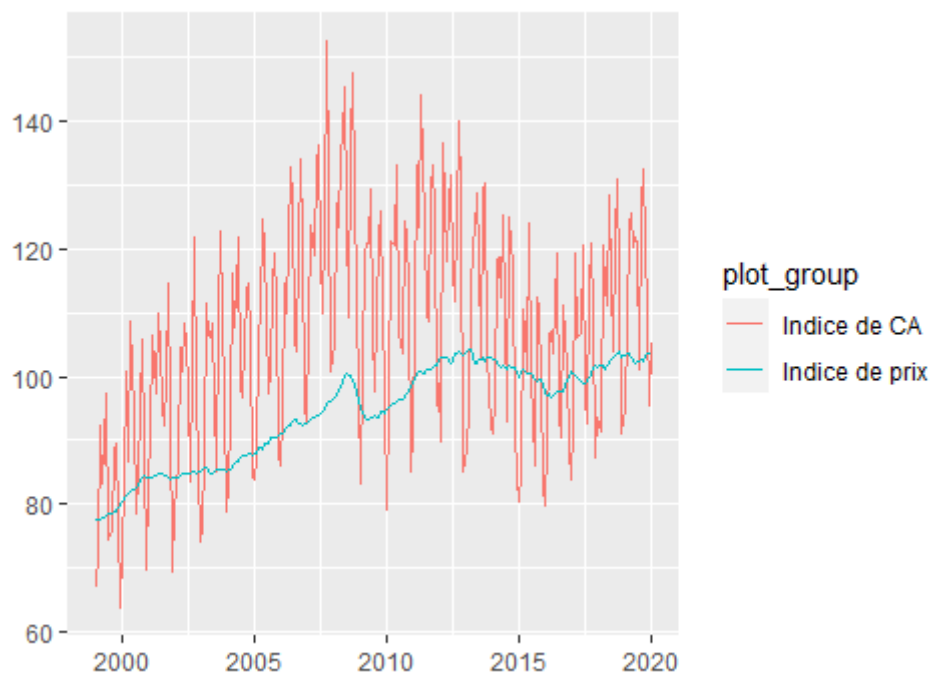
Partie 3: Analyse plus poussée

a) Analyse de la série avec un modèle avec variable explicative (Fonction d'intervention, ARMAX ou Cointegration/MCE).

Les modèles de la famille ARMA permettaient de modéliser une relation de court terme (CT) sur notre série, à savoir que la valeur actuelle était déterminée par un ensemble limité de valeurs passées. La cointégration va nous permettre d'aller plus loin dans notre étude et de définir une relation de long terme (LT) en introduisant une seconde série. Ici nous avons décidé d'intégrer le série sur l'indice de production industrielle française, *Indice de prix* dans la figure 14, pour voir si cette évolution extérieure a influencé les fluctuations de notre série principale *extractive*, *Indice de CA* dans la figure 14.

Pour qu'une relation de long terme (LT) existe entre nos deux séries, il faut qu'elles soient toutes les deux non-stationnaires intégrées de même ordre et que leurs tendances stochastiques soient liées.

Figure 14: Evolution des deux séries concernées par la cointégration



On le sait déjà, la série *Indice de CA*, qui n'est autre que la série *extractive*, n'est pas stationnaire, et on constate que la seconde série ne l'est pas non plus. Les deux séries ont une tendance à la hausse et leurs variations semblent corrélées.

Comment trouver cette corrélation ? Pour cela on va tester la cointégration.

Après différenciation, les deux séries sont stationnaires (voir *Table 8*, Annexe, p27).

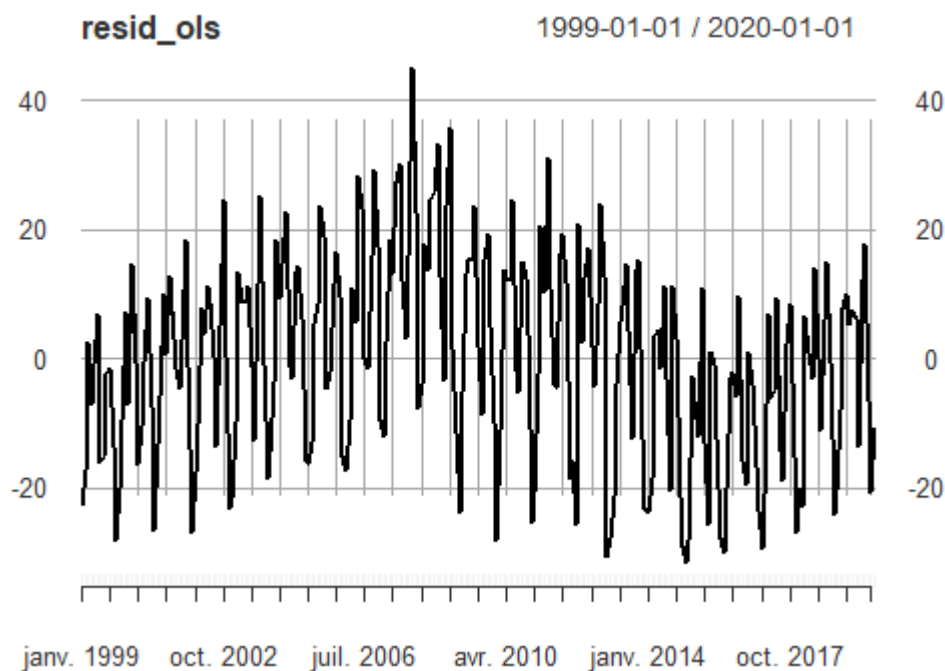
On dit qu'elles sont cointégrées d'ordre 1 et on note:

$$\star X_t \sim I(1) \Leftrightarrow (1 - L)X_t \sim I(0) \text{ , } X_t \text{ représentant } \textit{Indice de CA}$$

$$\star Y_t \sim I(1) \Leftrightarrow (1 - L)Y_t \sim I(0) \text{ , } Y_t \text{ représentant } \textit{Indice de prix}$$

On doit être capable d'exprimer X_t en fonction de Y_t , et le résidu z_t de cette relation doit être intégré d'ordre 0.

Figure 15: Série extraite du résidu de la cointégration



La série semble stationnaire en variance et en moyenne avant et après les ruptures de niveau et de tendance nettement visibles autour de 2008-2009 et de 2014. Un tel phénomène risque de fausser les résultats du test de cointégration. On décide donc

de procéder au test de RU via le test de Zivott & Andrews qui est un test ADF prenant compte des éventuelles ruptures (voir *Table 2*, Annexe, p26).

On compare le résultat du test obtenu avec le t-tabulé au niveau 5% de Engle et Yoo et on conclut à la stationnarité de la série des résidus. En conclusion, les deux séries sont cointégrées si l'on retire les effets des ruptures.

On a la relation de cointégration suivante: $z_t = \beta_1 X_t + \beta_2 Y_t$

b) Réalisation d'une prévision avec le MCE.

D'après le théorème de représentation de Granger (1981), les séries cointégrées peuvent toujours être représentées sous la forme d'un modèle à correction d'erreur (MCE). L'avantage d'un MCE dans un cas de cointégration est d'obtenir des prévisions plus fiables que dans le cas d'une estimation standard de relation de long terme. En effet, les résultats de l'estimation de cette relation de long terme sont faussés par la non-stationnarité des séries. Ainsi, le MCE lit la relation d'équilibre de long terme impliquée par la cointégration avec le mécanisme d'ajustement de la dynamique de court terme. Cette dernière décrit comment les variables réagissent lorsqu'elles sortent de l'équilibre de long terme.

On va donc tenter de réaliser dans cette dernière partie un MCE.

Soient $W_t = (X_t, Y_t)' \sim CI(1, 1)$, alors le MCE s'écrit:

$$\begin{aligned}\Delta X_t &= \alpha_1 z_{t-1} + \sum_{i=1}^{p_x} a_i \Delta X_{t-i} + \sum_{i=1}^{p_y} c_i \Delta Y_{t-i} + \varepsilon_{X_t} \\ \Delta Y_t &= \alpha_2 z_{t-1} + \sum_{i=1}^{p'_x} a'_i \Delta X_{t-i} + \sum_{i=1}^{p'_y} c'_i \Delta Y_{t-i} + \varepsilon_{Y_t}\end{aligned}$$

où α_1 et α_2 représentent la force de rappel vers l'équilibre de LT.

Les résidus de la relation de cointégration entre X_t et Y_t sont définis par:

$$z_{t-1} = X_{t-1} - bY_{t-1}$$

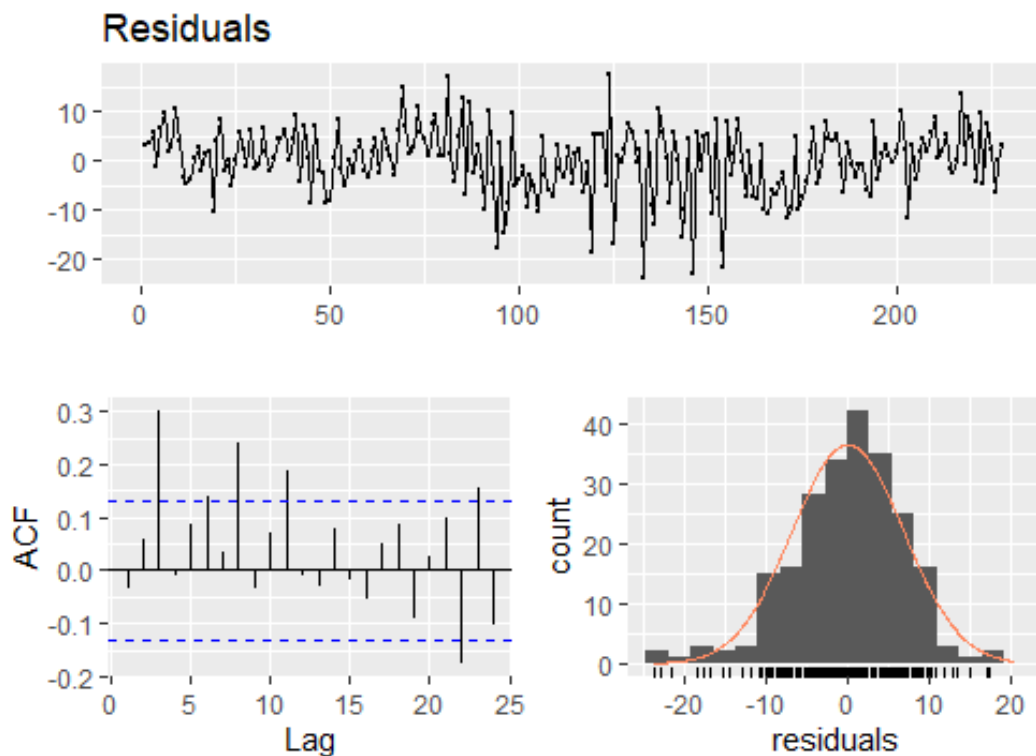
Les variables sont cointégrées, on estime alors la relation de court terme:

$$\Delta X_t = \alpha_1 \hat{z}_{t-1} + \sum_{i=1}^{p_x} a_i \Delta X_{t-i} + \sum_{i=1}^{p_y} c_i \Delta Y_{t-i} + \varepsilon_{X_t},$$

la relation de long terme étant celle représentée par la relation de cointégration.

L'estimation du modèle a été faite de deux manières différentes. A la "main", en utilisant une forme de processus "backward" qui consiste à supprimer petit à petit les coefficients associés aux décalages non-significatifs du modèle. Et avec la fonction *ecm* définie dans R, une fois les coefficients sélectionnés. On finit par conserver les deux premiers décalages de *dprix*, les résidus z_{t-1} de la relation ainsi que les 12 coefficients autorégressifs du modèle. La représentation des résidus de ce nouveau modèle est visible en figure 16 ci-dessous. On remarque encore une fois qu'ils sont fortement autocorrélés (pics pour les 3, 8 et 11 premiers décalages). La loi qu'ils suivent se rapproche tout de même d'une loi normale.

Figure 16: Résidus du MCE (extraits de la sortie de la fonction "ecm")



Conclusion

La série étudiée nous a permis d'observer l'évolution de l'indice de CA de l'industrie extractive française. Nous avons pu voir une attitude saisonale plutôt régulière, malgré une perturbation en 2008, qui est restée légère contrairement à ce qu'on aurait pu penser. En effet, 2008 est l'année de la crise financière de l'automne 2008, qui fait suite à la crise des subprimes de l'été 2007. C'est pourquoi l'intégration de la série de l'indice de production industrielle dans notre étude était pertinente dans notre recherche d'un modèle de prévision optimal. Mais avant cela, il était nécessaire d'analyser la série seule pour déterminer son type (DS) afin de la stationnariser correctement et éviter des problèmes d'autocorrélation.

Dans cette approche, nous avons décidé d'effectuer une prévision grâce au modèle naïf, de lissage exponentiel et des processus ARMA, qui prenaient en compte directement la saisonnalité de la série. Une autre approche aurait été de désaisonnaliser la série avant de la stationnariser et de mettre en place nos

modèles, avec par exemple le filtre moyenne mobile dans notre cas de saisonnalité rigide (marquée et répétitive).

Liste des tableaux et figures

Figure 1: Evolution de l'industrie extractive	3
Figure 2: Fonction d'autocorrélation	4
Figure 3: Fonction d'autocorrélation partielle	4
Figure 4: Schéma de décomposition	5
Figure 5: Série temporelle de l'industrie extractive après première différenciation	8
Figure 6: Fonction d'autocorrélation de dext	9
Figure 7: Graphique de saisonnalité annuelle	9
Figure 8: Répartition mensuelle de l'indice de CA de l'industrie extractive	10
Figure 9: Résidus provenant du modèle naïf saisonnier	11
Figure 10: Résidus provenant du modèle Holt-Winters multiplicatif	12
Figure 11: Prévisions sur un an de l'indice de chiffre d'affaire de l'industrie extractive avec le modèle Holt-Winters multiplicatif	13
Figure 12: Corrélogramme de la série dext	15
Figure 13: Prévisions de l'indice de chiffre d'affaire de l'industrie extractive sur an avec le modèle SARIMA(0,1,1)(2,1,3)	18
Figure 14: Evolution des deux séries concernées par la cointégration	19
Figure 16: Résidus du MCE (extraits de la sortie de la fonction "ecm")	22

Annexe

Table 0: Test de rupture de niveau et de tendance pour la date 01/01/2009

	p-valeur	significativité
Intercept	< 2e-16	oui
mytrend	7.09e-05	oui
coredata(d2009m1)	0.0931	oui au seuil de 10% ¹

¹Conclusion: rupture de niveau et de tendance après 2008-10-01.

Table 1: Stratégie séquentielle

Test (a1): Test ADF sur extractive

	t-Statistic	t-tabulé	Conclusion
Modèle 3 - tendance	-0.549	2.79	tendance non-significative
Modèle 2 - constante	2.685	2.53	constante significative
Modèle 2 - test de RU	-2.5285	-2.87	RU

Test (a2): Test ADF sur dext

	t-Statistic	t-tabulé	Conclusion
Modèle 3 - tendance	-1.012	2.79	tendance non significative
Modèle 2 - constante	1.280	2.53	constante non significative
Modèle 1 - aucune - test de RU	-5.3718	-1.95	série stationnaire

Test (b): KPSS

	KPSS	t-kpss	Conclusion
extractive	0.305	0.146	RU
dext	0.0337	0.146	pas de RU, série de type DS

Table 2: Test de RU de Zikey-Andrews pour les séries cointégrées

Variables	t-Statistic	Valeur critique au niveau 5%	Conclusion
dca	-15.8621	-5.08	I(1)
dprix	-10.7629	-5.08	I(1)

Table 3: Significativités des coefficients du modèle $SARIMA_{(0,1,1)}(2, 1, 3)$

coefficients	ma1	sar1	sar2	sma1	sma2	sma2
t-Statistic	-15.768	13.395	-77.628	-13.241	7.153	-5.266

Table 4: Test d'autocorrélation des résidus du modèle $SARIMA_{(0,1,1)}(2, 1, 3)$

	P-valeur du test de Ljung-Box	Conclusion
résidus	0.01646	Non-autocorrélation des résidus au niveau 10%

Table 5: Test de normalité des résidus du modèle $SARIMA_{(0,1,1)}(2, 1, 3)$

	P-valeur du test de Jarque-Bera	Conclusion
résidus	3.621e-12	Normalité des résidus au niveau 5%

Table 6: RMSE des prévisions sur les données tests

modèles	$SARIMA_{(0,1,1)}(2, 1, 3)$	$SARIMA_{(0,1,1)}(2, 1, 2)$	Naïf	ARIMA (2, 1, 4)
RMSE	6.902	6.525	23.536	11.450

Table 7: Tests d'autocorrélation sur les résidus des prévisions des modèles (S)ARIMA

modèles	$SARIMA_{(0,1,1)}(2, 1, 3)$	$SARIMA_{(0,1,1)}(2, 1, 2)$	Naïf	ARIMA (2, 1, 4)
p-valeurs	0.003704	0.0015	2.2e-16	0.02041

Table 8: Test de cointégration

	t-Statistic	Valeur critique Engle et Yoo	Conclusion
résidus	-10.224	-3.37	I(0)

Bibliographie

- *Tests de racine unité et stationnarisation des séries non stationnaires, présentation générale et application au cas des séries agricoles*, C. Tavéra, 1991: www.persee.fr
- *L'humanité épuise les ressources naturelles*, Le Figaro, 2011 : www.lefigaro.fr
- *Unit Root Test with Structural Breaks*, Aptech, 2019 : www.aptech.com
- *Les séries temporelles, analyse et modélisation d'agorécosystèmes*, S. Parent, 2020: essicolo.github.io/ecologie-mathematique-R
- *Séries temporelles avancées*, L. Ferrara: lo.ferrara.free.fr

Série pour la cointégration: [Base 2015 – Données mensuelles brutes](#)