



ENSAE PARIS

Projet de Series Temporelles

Modelisation ARIMA d'une serie temporelle

Auteurs :

Thomas SALOMON

Ilies EL JAOUHARI

3 juillet 2022

Table des matières

I	Données	1
I.1	Description	1
I.2	Remarques préliminaires et transformation	1
I.3	Représentation	2
II	Modèle ARMA	3
II.1	Etude des autocorrélogrammes	3
II.2	Sélection du modèle	3
II.3	Tests de normalité	4
III	Prévision	4
III.1	Hypothèses utilisées	4
III.2	Région de confiance	4
IV	Annexe	5
IV.1	Autocorrélogrammes de la série transformée	5
IV.2	Résidus du modèle AR(3)	6
IV.3	Prévision	7

I Données

I.1 Description

La série que nous avons choisie est l'indice brut de la production industrielle (en base 100) d'extraction d'hydrocarbures avec une granularité mensuelle entre Janvier 1990 à Janvier 2022. Elle compte donc 385 observations et sera notée $(Y_t)_{t \in T, T = \{1, \dots, 385\}}$. Celle-ci est disponible via le lien suivant : <https://www.insee.fr/fr/statistiques/serie/010537206>

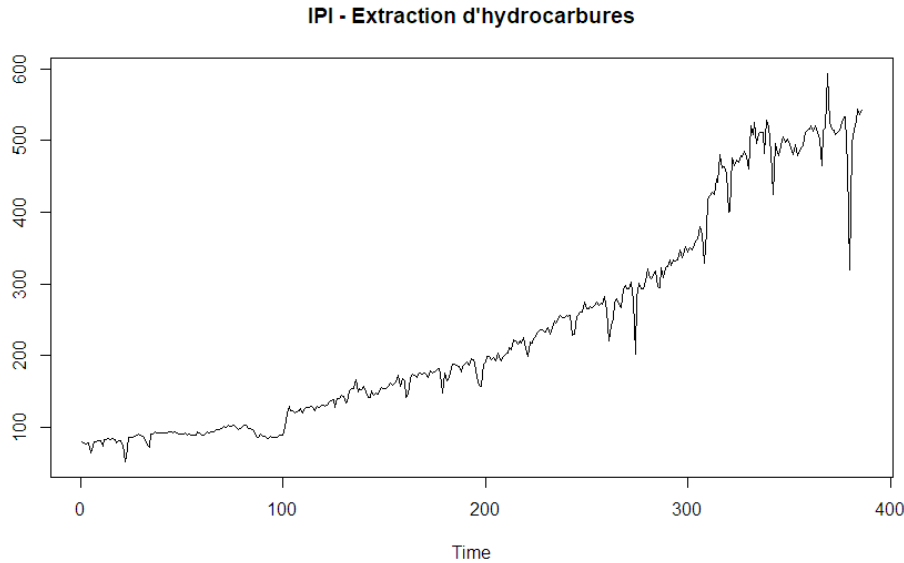


FIGURE 1 – Représentation graphique : Indice de la production industrielle de l'extraction d'hydrocarbures (base 100)

Cet indicateur appartient à la large gamme d'indices de production industrielle réalisés par l'INSEE et qui relèvent du règlement européen des statistiques de court terme.

I.2 Remarques préliminaires et transformation

On peut immédiatement observer une tendance croissante dans la série et une certaine saisonnalité à plus court terme. Dans un second temps, on voit qu'elle expose une certaine hétéroscédasticité. Une première ACF ainsi qu'une décomposition additive de notre série confirme bien cette tendance ainsi qu'une saisonnalité annuelle de faible amplitude. On corrige l'hétéroscédasticité avec une transformation logarithmique. Pour éliminer la tendance déterministe nous effectuons une première différenciation à l'ordre 1 qui nous donne l'autocorrélogramme complet suivant :

Nous pouvons maintenant effectuer nos 3 tests de stationnarité afin de confirmer (ou rejeter) notre hypothèse : Augmented Dickey-Fuller (ADF) , Phillips-Perron (PP) ainsi que le test de stationnarité Kwiatkowski-Phillips-Schmidt-Shin (KPSS). Il est important de noter que l'hypothèse nulle des deux premiers tests est la présence d'une racine unitaire alors que dans le cas du test KPSS l'hypothèse nulle est la stationnarité de notre série. Les résultats sont les suivants :

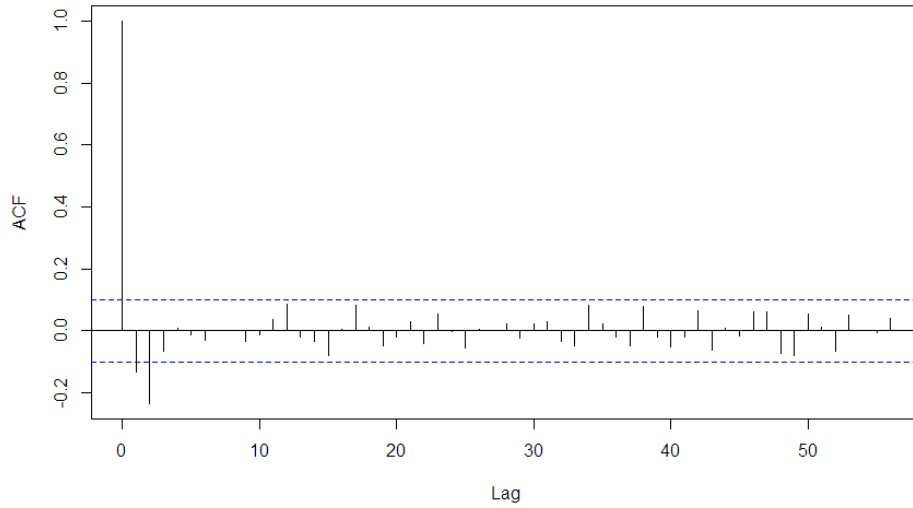


FIGURE 2 – ACF de la série différenciée à l'ordre 1

Test	t-value	$\mathbb{P}(\geq t)$
ADF	-9.5541	<0.01
PP	-336.7	<0.01
KPSS	0.024532	>0.1

Les tests ADF et Phillips-Perron donne des p-values inférieures à 0.01 ce qui permet de rejeter l'hypothèse H_0 à 1%. Quant au test KPSS la p-value supérieures à 0.1 nous empêche de rejeter l'hypothèse H_0 à 10%. L'hypothèse de stationnarité est donc vérifiée à tous les niveaux usuels et nous pouvons poursuivre l'étude avec la nouvelle série transformée :

$$X_t = (1 - B) \log(Y_t), t \in T$$

I.3 Représentation

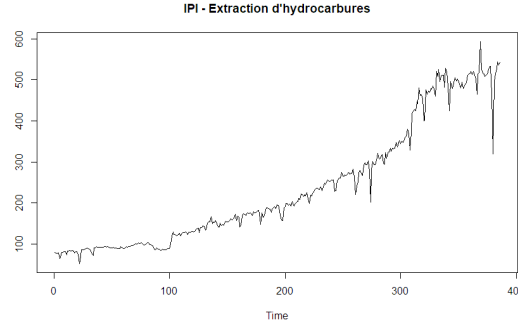


FIGURE 3 – Y_t

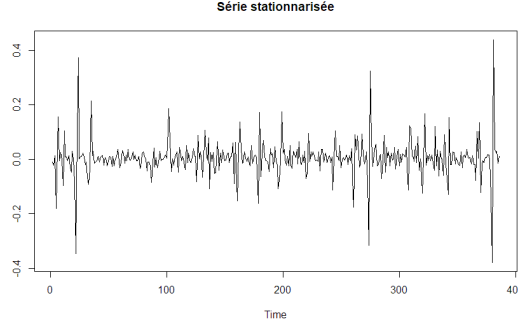


FIGURE 4 – X_t

FIGURE 5

II Modèle ARMA

II.1 Etude des autocorrélogrammes

Nous allons maintenant tenter d'identifier les paramètres de notre futur modèle ARMA grâce aux ACF et PACF (en Annexe). Ici nous utilisons deux propriétés parallèles des modèles $MA(q)$ et $AR(p)$: $\rho(h) = 0, \forall h > q$ et $r(h) = 0, \forall h > p$. Les ACF et PACF nous permettent de mettre en lumière la significativité de chaque ordre. L'autocorrélogramme complet montre que l'ordre $q \leq 1$ est significativement non-nul à 5% malgré les pics à 12, 13 et 18. Pour l'autocorrélogramme partiel, on peut observer que les ordres $p \in \{0, 1, 2, 3, 4\}$ sont tous significatifs à 5% ce qui n'est pas le cas pour les ordres supérieurs mis à part pour les ordres 11 et 12.

II.2 Sélection du modèle

Afin d'estimer les modèles à notre série, nous effectuons un premier test de significativité des coefficients. Pour cela nous utilisons la library *lmtest* et la fonction *coefstest* qui nous donne les modèles suivants : $MA(1)$, $MA(2)$, $AR(1)$, $ARMA(1,1)$, $AR(2)$, $ARMA(2,1)$, $AR(3)$.

Pour poursuivre la sélection des modèles pertinents, nous allons maintenant vérifier la blancheur des résidus pour ceux sélectionnés précédemment. Les p-value du test de Ljung-Box sont calculées à l'aide de la library *FitAR*. La vérification de la blancheur des résidus pour chaque modèle nous conduit à ne conserver que les modèles $MA(2)$, $ARMA(2,1)$, $AR(3)$.

Les tests de blancheur des résidus nous ayant amené à conserver ces 3 modèles, nous allons utiliser les critères d'information d'Akaike et bayésien afin de sélectionner le plus optimal. Notre choix se portera sur le modèle minimisant les deux (s'il en existe un).

Modèle	AIC	BIC
$MA(2)$	-1018.105	-1006.245
$ARMA(2,1)$	-1017.674	-1001.861
$AR(3)$	-1010.348	-994.5355

Le modèle ARIMA(3,1,0) minimise les deux critères d'information pour la série $\log(Y_t)$. Nous utiliserons donc celui-ci dans la suite de l'étude.

II.3 Tests de normalité

Conformément à l'étude, réalisée par Nornadiah Mohd Razali1 et Yap Bee Wah1 sur la comparaison des puissances entre les tests de normalité de Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors et Anderson-Darling. Nous avons choisi de tester la normalité de nos résidus à l'aide des tests de Shapiro-Wilk et Anderson-Darling, car même si le test de Shapiro-Wilk est le plus puissant conformément aux résultats de l'étude, celui-ci semble plus adapté aux échantillons de petite taille. Les résultats donnent des p-value inférieures à $2.2e-16$, ce qui invalide H_0 de normalité des résidus. Ceux-ci sont compréhensibles aux vues de la figure de densité et du QQ-plot qui montrent des écarts anormaux aux niveaux des queues.(Annexe).

III Prévision

III.1 Hypothèses utilisées

Nous noterons à partir d'ici T la longueur de la série X_t . Les résidus sont supposés gaussiens suivant une loi $\mathcal{N}(0, \sigma^2)$. Grâce à l'étude précédente, nous avons trouvé le modèle AR(3) pour notre série X_t . Celui-ci s'écrit :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \epsilon_t$$

Les prédictions en $T+1$ et $T+2$ sachant les précédentes se notent donc :

$$\begin{cases} \hat{X}_{T+1|T} = \phi_1 X_T + \phi_2 X_{T-1} + \phi_3 X_{T-2} \\ \hat{X}_{T+2|T} = \phi_1 \hat{X}_{T+1|T} + \phi_2 X_T + \phi_3 X_{T-1} \end{cases}$$

III.2 Région de confiance

Par définition des résidus, nous savons qu'ils sont orthogonaux à toutes les informations passées (X_{T-1}, X_{T-2}, \dots) on obtient donc que : $\mathbb{E}[\epsilon_T + h | X_T, X_{T-1}, \dots] = 0 \forall h > 0$. Maintenant calculons l'erreur de prédiction : $X_{T+1} - \hat{X}_{T+1|T}$ et $X_{T+2} - \hat{X}_{T+2|T}$

On pose : $X = \begin{pmatrix} X_{T+1} \\ X_{T+2} \end{pmatrix}$ et $\hat{X} = \begin{pmatrix} \hat{X}_{T+1|T} \\ \hat{X}_{T+2|T} \end{pmatrix}$

Dans la partie 2 nous avons vu que notre série transformée suivait un AR(3), on déduit que :

$$\hat{X} - X = \begin{pmatrix} X_{T+1} - \hat{X}_{T+1|T} \\ X_{T+2} - \hat{X}_{T+2|T} \end{pmatrix} = \begin{pmatrix} \epsilon_{T+1} \\ \epsilon_{T+2} \end{pmatrix}$$

Calculons maintenant la variance des résidus :

$$\mathbb{V}[\hat{X} - X] = \begin{pmatrix} \mathbb{V}[\epsilon_{T+1}] \\ \mathbb{V}[\epsilon_{T+2}] \end{pmatrix} = \begin{pmatrix} \sigma_{T+1}^2 \\ \sigma_{T+2}^2 \end{pmatrix}$$

Notons Σ la matrice d'autocorrélation de $\hat{X} - X$

$$\Sigma = \begin{pmatrix} \sigma_t^2 & 0 \\ 0 & \sigma_t^2 \end{pmatrix}$$

Et comme $\sigma_t^2 > 0$, la matrice d'autocorrélation est inversible. D'après le cours, nous savons que $(X - \hat{X})\Sigma^{-1}(X - \hat{X})$ suit une loi $\chi(2)$, ce qui nous permet d'obtenir la région de confiance :

$$\{X \in \mathbb{R}^2 | (X - \hat{X})\Sigma^{-1}(X - \hat{X}) < q_{\chi^2(2)}^{1-\alpha}\}$$

où $q_{\chi^2(2)}^{1-\alpha}$ est le quantile d'ordre $1 - \alpha$ d'une loi $\chi^2(2)$.

La prédiction pour les valeurs sont présentes en annexe.

IV Annexe

IV.1 Autocorrélogrammes de la série transformée

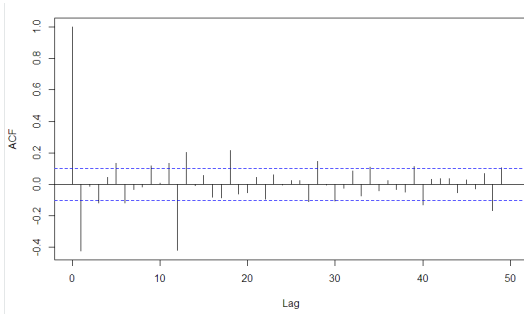


FIGURE 6 – ACF de la série transformée

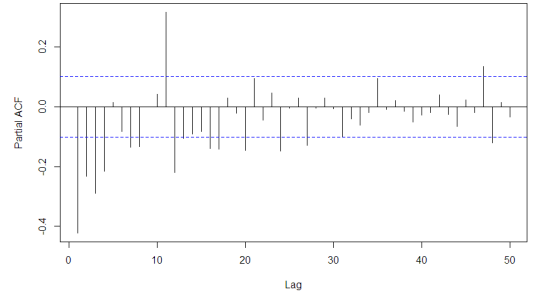


FIGURE 7 – PACF de la série transformée

IV.2 Résidus du modèle AR(3)

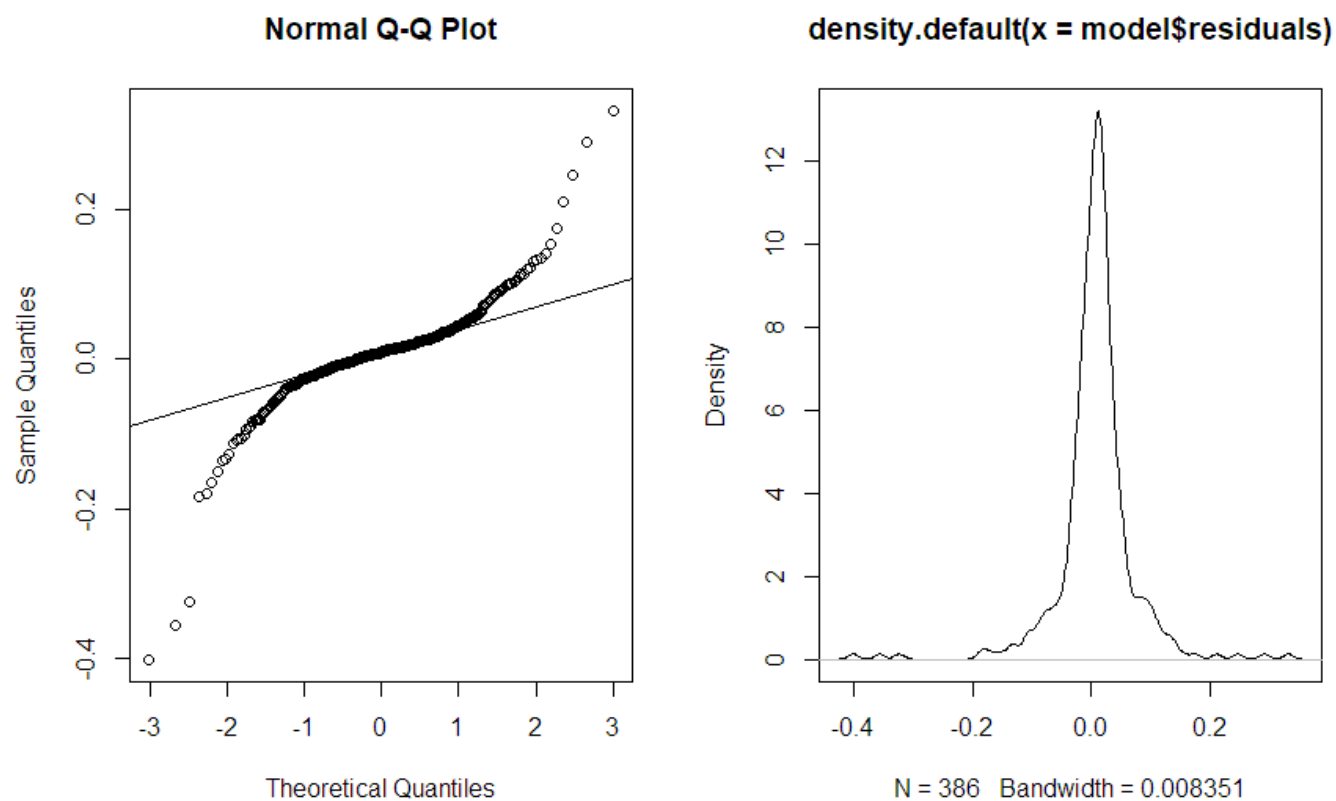


FIGURE 8 – Q-Q plot et densité des résidus

IV.3 Prédiction

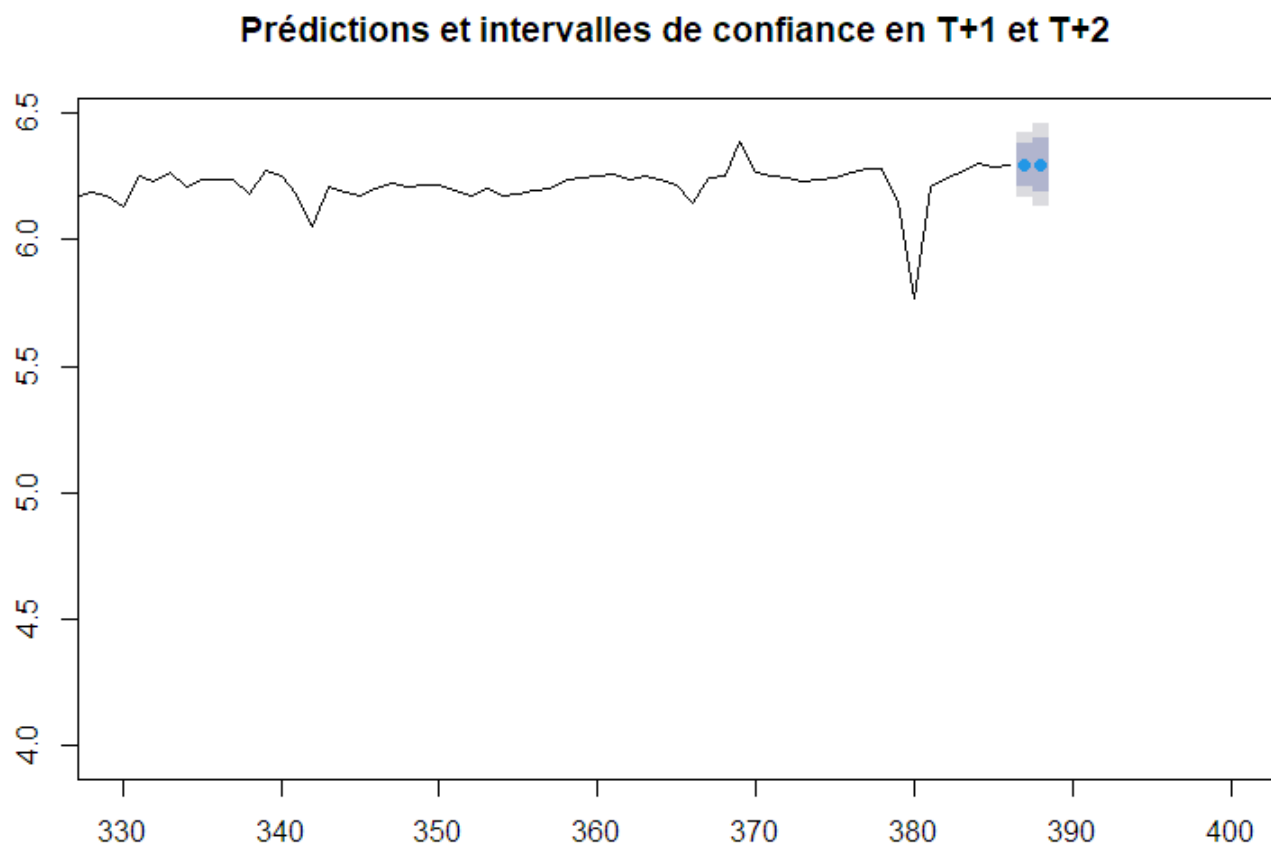


FIGURE 9 – Prédiction X_{T+1} et X_{T+2}