



Projet de séries temporelles

Elaboré par:
KHENISSI Nour
TELMINI Mohamed Fyras

Année universitaire :2021/2022
Classe: 2A

1 Partie I: les données

1.1 Présentation des données

La série étudiée¹ dans ce projet correspond à l'indice de la production industrielle de l'industrie chimique en France selon la référence de la base 100 en 2015 et de la méthodologie correspondante². La série est mensuelle et contient les valeurs de l'indice entre 1990 et 2021.



Figure 1: évolution de l'indice de la production industrielle de l'industrie chimique en France entre 1990 et 2021

1.2 Stationnarisation de la série

Nous remarquons des fluctuations assez importantes dans la série, chose qui nous indique la présence d'un caractère hétéroscédastique. nous corrigerons cela par une transformation logarithmique ayant pour effet de stabiliser la variance. Nous étudions donc la série $(Y_t)_{t \in T}$, $Y_t = \text{Log}(X_t)$.

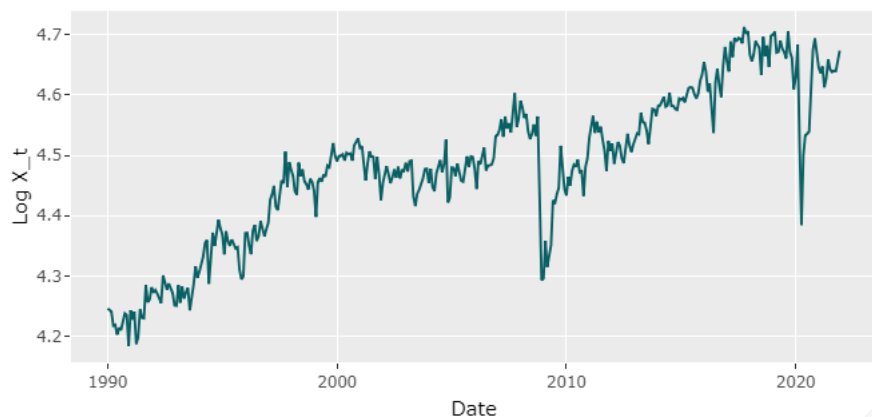


Figure 2: Représentation de la série transformée

La décomposition de la série et la représentation de l'autocorrélogramme nous mènent à constater la présence d'une tendance croissante ainsi qu'une saisonnalité mensuelle.

¹<https://www.insee.fr/fr/statistiques/serie/010537422>

²<https://www.insee.fr/fr/metadonnees/source/indicateur/p1671/description>

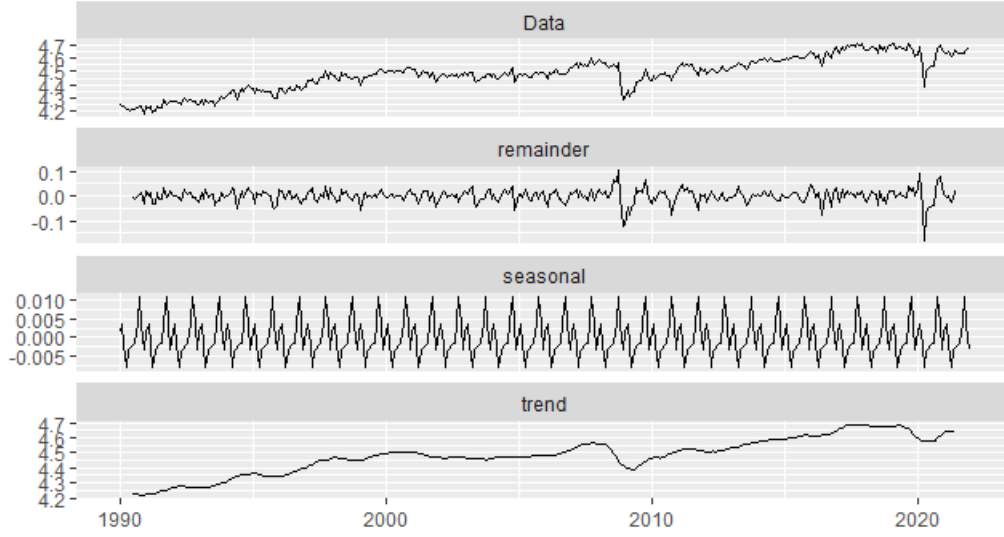


Figure 3: Décomposition de la série

Nous appliquons donc l'opérateur de différenciation saisonnier $\Delta_{12}Y_t = Y_t - Y_{t-12} = \log(X_t) - \log(X_{t-12})$. Nous avons recours par la suite aux tests statistiques qui permettent de vérifier si une série est stationnaire ou non. En effet, il y a deux types de tests différents : les tests de stationnarité, comme le test KPSS, pour lesquels l'hypothèse nulle H_0 est que la série est stationnaire, et les tests de racine unitaire comme le test de Dickey-Fuller et le test augmenté de Dickey-Fuller (ADF) pour lesquels l'hypothèse nulle est que la série a été générée par un processus présentant une racine unitaire, et donc, qu'elle n'est pas stationnaire. Le résultat du test KPSS appliqué à la série $\log(X_t) - \log(X_{t-12})$ conduit à rejeter H_0 au seuil de 10%, nous indiquant que la série n'est pas stationnaire. De ce fait nous procédons à une différenciation de premier ordre de la série $\log(X_t) - \log(X_{t-12})$. On introduit alors $(G_t)_{t \in T}$ définie par :

$$G_t = \Delta(\log(X_t) - \log(X_{t-12})) = (1 - B)^1(1 - B^{12})^1 \log X_t \quad \forall t \in T$$

Afin de vérifier la stationnarité de la série différenciée G_t , nous appliquons de nouveau les tests KPSS et DF. Le test KPSS donne une p-value supérieure à 0,01 nous menant à retenir l'hypothèse nulle à un seuil de 1% et le test DF donne une p-value inférieure à 0.01, d'où le rejet de l'hypothèse de racine unitaire. Il n'est plus donc nécessaire de différencier la série une autre fois et nous supposons dans ce qui suit la stationnarité de la série G_t .

1.3 Comparaison entre la série brute et la série corrigée

Nous représentons la série choisie avant et après la transformation comme l'indique la figure suivante:

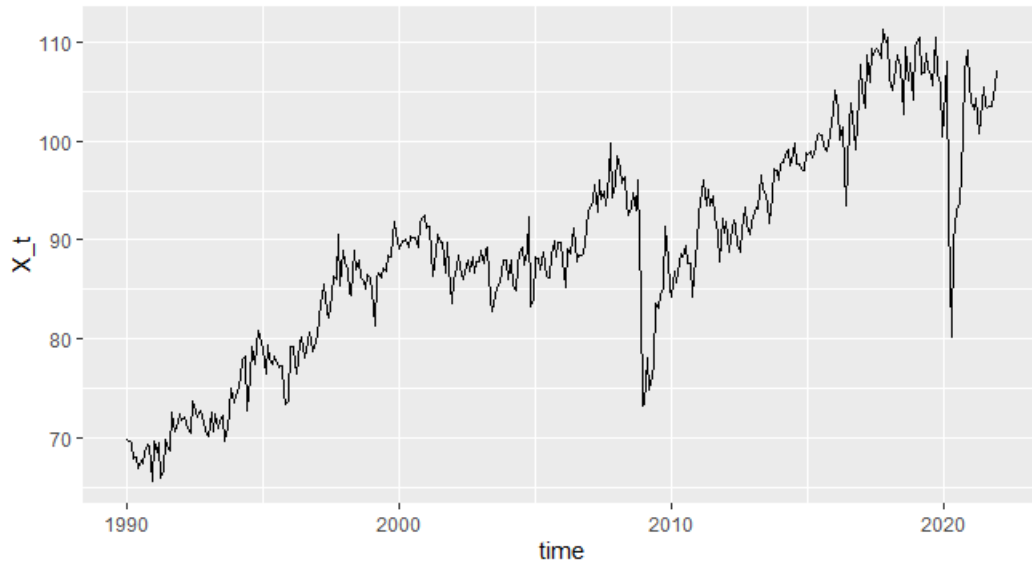


Figure 4: Série avant traitement

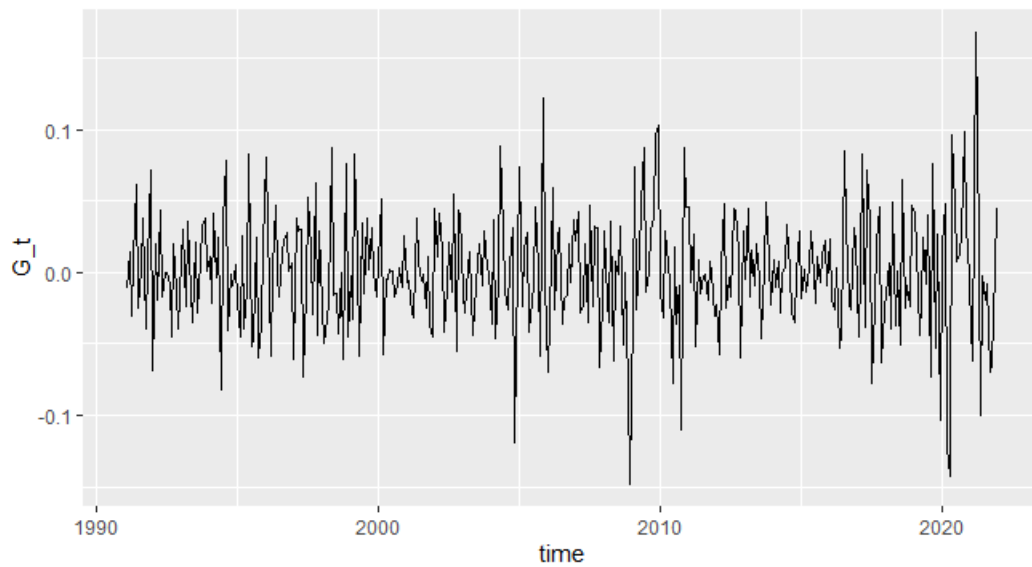


Figure 5: Série après traitement

2 Partie II: Modèles ARMA

2.1 Identification des ordres

Une fois la stationnarisation effectuée, nous cherchons à modéliser la série observée par le modèle SARI-MAs(p,d,q)(P,D,Q) qui semble être le plus adéquat d'après ce qui précède avec $d = D = 1$, $s = 12$ et p, q, P et Q des ordres à déterminer.

Comme première étape nous représentons les autocorrélogrammes complets et partiels:

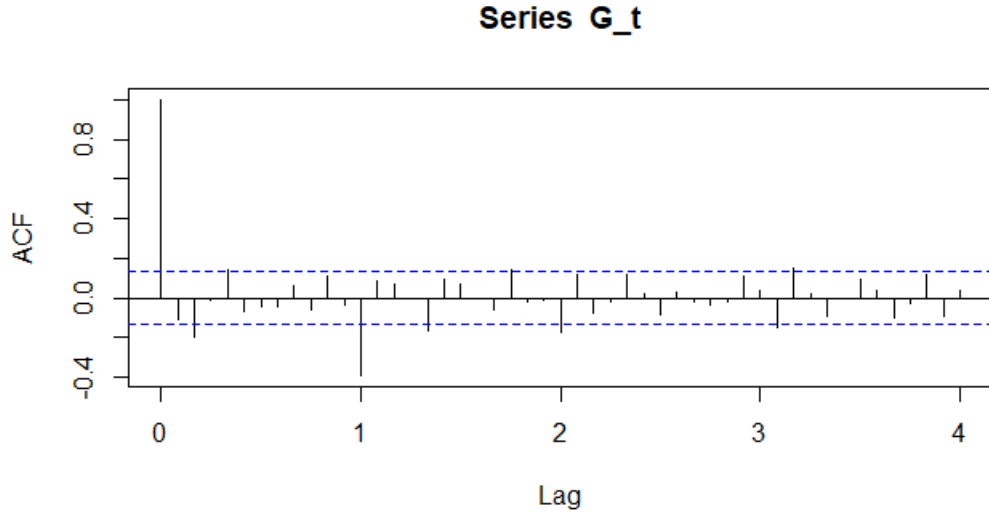


Figure 6: Autocorrelogramme de la série G_t

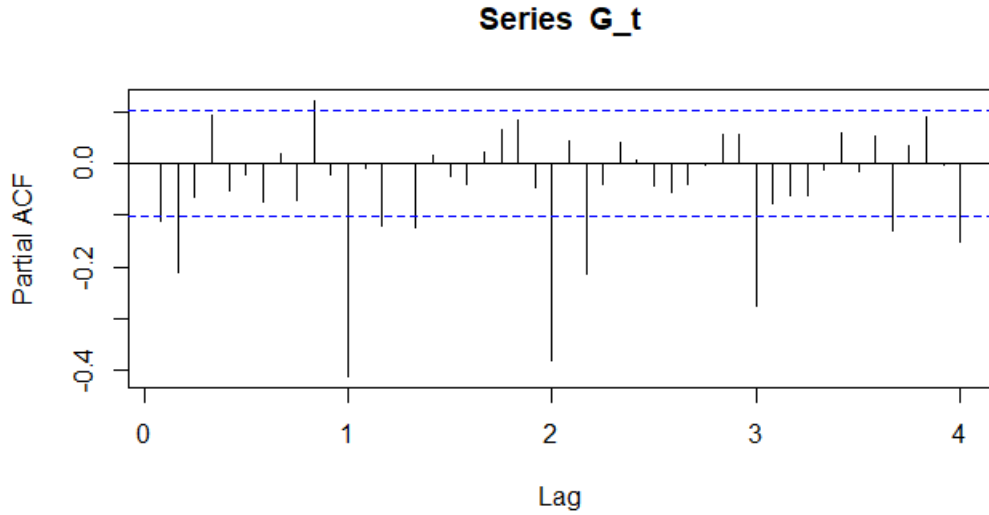


Figure 7: Autocorrelogramme complet de la série G_t

Sachant qu'une propriété caractéristique d'un $MA(q)$ est que $\rho(h) = 0 \forall h > q$, nous retenons $q \in (0, 1, 2, 3)$ puisque les auto-corrélations sont significativement nulles à 5% pour $h > 3$ sauf pour le lag12, ce qui nous mène à choisir $Q \in (0, 1)$. L'étude de l'autocorrélogramme partiel nous montre des autocorrélations significativement nulles à 5% pour $h > 10$ sauf pour les lags 12 et 24. Nous rappelons qu'un $AR(p)$ est caractérisé par $r(h) = 0 \forall h > p$ et donc nous déduisons que $p \in [0..10]$ et que $P \in (0, 1, 2, 3)$.

2.2 Choix du modèle selon les critères AIC/BIC:

Nous avons recours aux critères d'information AIC et BIC pour choisir le meilleur modèle, celui qui minimise ces deux critères, parmi 352 modèles possibles.

	Ordres					
Modèle choisi par	p	d	q	P	D	Q
<i>AIC</i>	7	1	3	3	1	1
<i>BIC</i>	2	1	0	2	1	1

Table 1: Choix des modèles selon les critères AIC-BIC

Nous rappelons que l'idée est de se prémunir contre la surparamétrisation en introduisant un coût à l'introduction de chaque paramètre supplémentaire. Le critère AIC a tendance à privilégier les gros modèles tandis que le critère BIC pénalise davantage la surparamétrisation et ils ne conduisent pas forcément à la même sélection du modèle ce qui est le cas ici. Dans ce sens, nous optons pour la modélisation la plus simple et qui minimise le BIC. Nous retenons donc $\text{SARIMA}_{12}(2,1,0)(2,1,1)$.

Nous effectuons ensuite un test de student sur les coefficients du modèle choisi:

	Estimate	SE	t.value	p.value
ar1	-0.125	0.050	-2.485	0.013
ar2	-0.260	0.051	-5.156	0
sar1	0.002	0.051	0.030	0.976
sar2	-0.257	0.057	-4.548	0
sma1	-1	0.049	-20.350	0

Table 2: Coefficients du modèle $\text{SARIMA}_{12}(2,1,0)(2,1,1)$

Les pvalues obtenues sont inférieures à 0.05, nous rejettons donc l'hypothèse de la nullité des coefficients à un seuil de 5% sauf pour le coefficient sar1 ce qui n'est pas une remise en cause du choix de l'ordre car ne concerne pas les coefficients d'ordre maximal.

2.3 Validation du modèle par les tests sur les résidus:

Si l'ajustement de notre modèle est bon, les résidus doivent être proches d'un Bruit blanc gaussien. Pour cela et comme dernière étape de validation, nous conduisons deux tests statistiques:

-Test du Portemanteau/Ljung-box: qui évalue la corrélation entre les résidus et a pour hypothèse nulle H_0 la présence d'un bruit blanc fort.

-Test de Shapiro-Wilk: teste l'hypothèse nulle selon laquelle les résidus sont normalement distribués.

Le test du Portemanteau conduit à ne pas rejeter l'hypothèse de l'indépendance des résidus à un niveau de 5%. Et le test de Shapiro-Wilk donne une pvalue égale à $9.917e^{-10}$ ce qui rejette l'hypothèse de la normalité.

3 Partie III: Pr evision

3.1 Intervalle de confiance

La s erie $(Y_t)_{t \in T}$ est mod elis ee par $\text{SARIMA}_{12}(2,1,0)(2,1,1)$. Donc la s erie $(G_t)_{t \in T}$ suit un processus $\text{ARMA}(p + sP, q + sQ) = \text{ARMA}(26,12)$. On a:

$$(1 - \sum_{i=1}^2 \varphi_i B^i)(1 - \sum_{i=1}^2 \Phi_i B^{12})G_t = (1 - 0 * B^1)(1 - \Psi_1 B^{12})\epsilon_t$$

$$\Leftrightarrow G_t = \sum_{i=1}^2 \varphi_i G_{t-i} + \sum_{i=1}^2 \Phi_i G_{t-12} - \Phi_1 \sum_{i=1}^2 \varphi_i G_{t-12-i} - \Phi_2 \sum_{i=1}^2 \varphi_i G_{t-12-i} + \epsilon_t - \Psi_1 \epsilon_{t-12}$$

La pr evision lin eaire $(\hat{G}_{T+1|T}, \hat{G}_{T+2|T})$ des valeurs (G_{T+1}, G_{T+2}) est donn ee par:

$$\begin{cases} \hat{G}_{T+1|T} = \sum_{i=1}^2 \varphi_i G_{T+1-i} + \sum_{i=1}^2 \Phi_i G_{T+1-12} - \Phi_1 \sum_{i=1}^2 \varphi_i G_{T+1-12-i} - \Phi_2 \sum_{i=1}^2 \varphi_i G_{T+1-12-i} - \Psi_1 \epsilon_{T+1-12} \\ \hat{G}_{T+2|T} = \varphi_1 \hat{G}_{T+1|T} + \varphi_2 G_T + \sum_{i=1}^2 \Phi_i G_{T+2-12} - \Phi_1 \sum_{i=1}^2 \varphi_i G_{T+2-12-i} - \Phi_2 \sum_{i=1}^2 \varphi_i G_{T+2-12-i} - \Psi_1 \epsilon_{T+2-12} \end{cases}$$

d'o  les erreurs de pr ecision sont obtenus par:

$$\begin{cases} G_{T+1} - \hat{G}_{T+1|T} = \epsilon_{T+1} \\ G_{T+2} - \hat{G}_{T+2|T} = \epsilon_{T+2} + \varphi_1 \epsilon_{T+1} \end{cases}$$

Sachant que $G_t = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}$ et notons $(\hat{Y}_{T+1|T}, \hat{Y}_{T+2|T})$ les pr edictions de (Y_{T+1}, Y_{T+2}) , on a:

$$\begin{cases} Y_{T+1} - \hat{Y}_{T+1|T} = G_{T+1} - \hat{G}_{T+1|T} \\ Y_{T+2} - \hat{Y}_{T+2|T} = G_{T+2} - \hat{G}_{T+2|T} \end{cases}$$

$$\Leftrightarrow \begin{cases} Y_{T+1} - \hat{Y}_{T+1|T} = \epsilon_{T+1} \\ Y_{T+2} - \hat{Y}_{T+2|T} = \epsilon_{T+2} + \varphi_1 \epsilon_{T+1} \end{cases}$$

On note: $Y = \begin{pmatrix} Y_{T+1} \\ Y_{T+2} \end{pmatrix}$ et $\hat{Y} = \begin{pmatrix} \hat{Y}_{T+1|T} \\ \hat{Y}_{T+2|T} \end{pmatrix}$

Avec l'hypoth ese que les r esidus de la s erie sont gaussiens, on a $Y - \hat{Y} \sim N(0, \epsilon)$ avec Σ la matrice de variance-covariance donn ee par:

$$\Sigma = \begin{pmatrix} \sigma_\epsilon^2 & \sigma_\epsilon^2 \varphi_1 \\ \sigma_\epsilon^2 \varphi_1 & \sigma_\epsilon^2 + \sigma_\epsilon^2 \varphi_1 \end{pmatrix}$$

On suppose que Σ est inversible et on conduit le test ayant pour hypoth ese nulle $H_0: Y=0_{\mathbb{R}^2}$

et de statistique $Y^T \Sigma^{-1} Y \sim \chi^2(2)$ d'o  une r egion de confiance de la forme:

$$\text{IC} = \left\{ y \in \mathbb{R}^2 \mid (y - \hat{Y})^T \Sigma^{-1} (y - \hat{Y}) \leq q_{\chi^2(2)}^{1-\alpha} \right\} \text{ avec } q_{\chi^2(2)}^{1-\alpha} \text{ le quantile d'ordre } 1 - \alpha \text{ de la loi } \chi^2(2)$$

3.2 Hypoth eses

Pour aboutir aux r esultats pr esent es pr ecedemment, nous avons suppos  vraies les hypoth eses suivantes:

- Les residus sont iid et normalement distribu s $\epsilon_t \sim N(0, \sigma_\epsilon^2)$.
- Le mod ele est bien ajust  (les coefficients sont bien identifi s)
- $\sigma_\epsilon^2 > 0$ pour que la matrice Σ soit inversible

Dans notre cas, l'hypoth ese de normalit  des r esidus n'est pas v erifi e, ce qui contrarie notre mod elisation.

3.3 Représentation graphique

La représentation graphique des intervalles de confiances univariés pour un seuil de 5% donne:

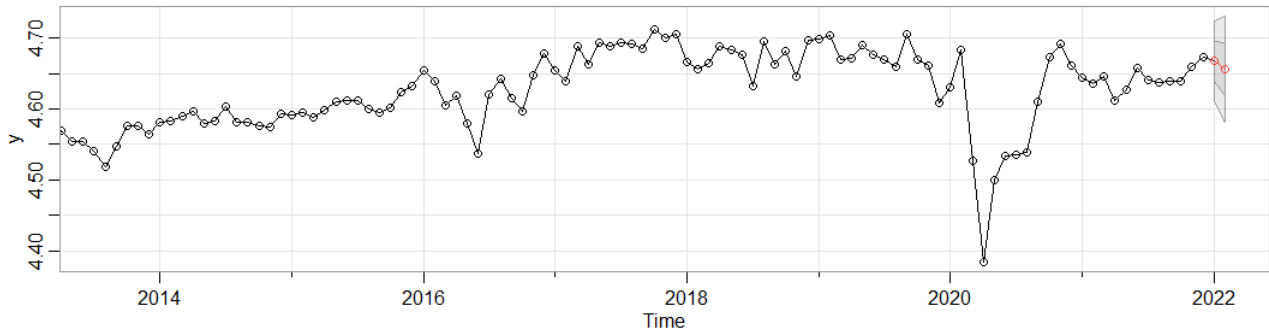


Figure 8: Les intervalles de confiances univariés au seuil de 5%

3.4 Question ouverte:

Nous disons qu'une série $(Y_t)_{t \in T}$ Granger provoque une autre série $(X_t)_{t \in T}$ si les prévisions de la valeur de X_t en fonction de ses propres valeurs passées et sur les valeurs passées de Y_t sont meilleures que les prévisions de X_t basées uniquement sur ses propres valeurs passées. Cette causalité au sens de Granger est définie mathématiquement comme suit: $E[X_t | X_{t-1}, Y_{t-1}] \neq E[X_t | X_{t-1}]$