

École Nationale de la Statistique et de l'Administration Economique

Projet de séries temporelles



Modélisation ARIMA d'une série temporelle : Indice CVS-CJO de la production industrielle (base 100 en 2015) - Fabrication de bière

Louis Becquart, Ludovic Guyot
31 août 2023

Table des matières

1	Partie I : Les données	3
1.1	Question 1	3
1.2	Question 2	3
1.3	Question 3	4
2	Partie II : Modèles ARMA	4
2.1	Question 4	4
2.2	Question 5	5
3	Partie III : Prévision	6
3.1	Question 6	6
3.2	Question 7	7
3.3	Question 8	8
3.4	Question 9	8
4	Annexe	8

1 Partie I : Les données

1.1 Question 1

Nous travaillons sur un indice de production industriel issue des données de l'INSEE¹, la fabrication de bière. Notre série est indicé en base 100 avec pour référence l'année 2015, et est déjà corrigée des variations saisonnières et des jours ouvrés. Cet indice est disponible de janvier 1990 à février 2023 avec une récurrence mensuelle, pour un total de 398 observations. Nous avons supprimé les 2 dernières observations, c'est à dire de novembre 2022 à février 2023, afin de réaliser des prédictions dites "in sample". Aucune transformation n'a été effectué sur la série à ce stade. Travailler sur cette série nous permettra d'estimer l'évolution de la production de bière en France dans les années à venir. Notre étude pourrait alors s'avérer utile dans le cadre de débats de santé public, par exemple si l'on venait à observer une hausse importante de la production de bière. Cette dernière pourrait en effet être révélatrice d'une hausse parallèle de la consommation de ce produit qui, nous le rappelons, doit rester modérée...

1.2 Question 2

Avant de modéliser la série, il est nécessaire de vérifier sa stationnarité. Dans le cas où la série ne serait pas stationnaire, il faut alors la différencier afin d'obtenir une série modélisable. Nous avons décidé d'utiliser le test augmenté de Dickey-Fuller. Cette méthode permet de tester l'hypothèse H_0 qui suppose qu'une racine unitaire est présente dans le modèle auto-régressif, c'est à dire $\rho = 1$ dans l'équation² :

$$(1.1) \quad \Delta X_t = \alpha + (\rho - 1)\Delta X_{t-1} + \varepsilon_t$$

Cette hypothèse revient à supposer que la série n'est pas stationnaire. Ainsi, dans le cas où on rejetterait H_0 ³, on pourrait conclure à la stationnarité. Cependant, il faut choisir le bon modèle parmi les 3 suivants pour que le test ne soit pas biaisé :

$$(1.2) \quad \Delta X_t = (\rho - 1)X_{t-1} + \varepsilon_t$$

$$(1.3) \quad \Delta X_t = \alpha + (\rho - 1)X_{t-1} + \varepsilon_t$$

$$(1.4) \quad \Delta X_t = \alpha + \beta t + (\rho - 1)X_{t-1} + \varepsilon_t$$

En régressant notre série sur le temps, on observe que seul la constante est significative. On choisit donc de retenir le 2^{ème} modèle, i.e le modèle avec constante sans tendance temporelle. Il faut encore vérifier que les résidus issus de la régression de la spécification de ce test ne sont pas autocorrélés. Pour ce faire, on utilise la fonction `Qtests` pour effectuer des tests de Ljung-Box sur la série chronologique des résidus de la spécification du test ADF jusqu'à un certain nombre de décalages k choisi a priori. Cette méthode teste l'hypothèse nulle que ces résidus sont des bruits blancs (i.e série aléatoire sans corrélation). Si on rejette l'hypothèse H_0 pour au moins un des résidus alors on considère que la spécification est fausse et on recommence le processus avec un lag supplémentaire dans la spécification du test de Dickey-Fuller. Le test devient donc un "Augmented Dickey-Fuller test" :

$$(1.5) \quad \Delta X_t = \alpha + \gamma X_{t-1} + \sum_{i=1}^p \delta_i \Delta X_{t-i} + \varepsilon_t$$

Pour trouver le réel p optimal de cette spécification, nous utilisons la fonction `adfTest_valid`. La fonction utilise une approche itérative. Elle commence avec p égal à 0, puis teste si les résidus obtenus après le test ADF sont des bruits blancs en utilisant la fonction `Qtests` décrite précédemment. Si les résidus passent tous le test de bruits blancs, la boucle s'arrête et la fonction renvoie le résultat du test ADF. Sinon, elle incrémente p de 1 et répète le processus jusqu'à ce que les résidus passent le test de bruits blancs.

1. <https://www.insee.fr/fr/statistiques/serie/010537309#Telechargement>

2. Dans la suite de ce paper, $(X_t)_{t=1}^T$ correspond à notre série temporelle non-différenciée, et $(\Delta X_t)_{t=1}^T$ à sa différence première.

3. Nous raisonnerons dans ce travail avec un niveau de confiance de 95%

Nous concluons avec cette fonction qu'il faut inclure 9 lags dans la spécification du test ADF. Donc p est égal à 9 dans la spécification précédente. Avec cette spécification, on accepte l'hypothèse de racine unitaire car γ n'est pas significativement différent de 0 dans l'équation 1.5, ce qui revient à dire que ρ est proche de 1. Notre série est donc non-stationnaire. Nous créons une série différenciée : $(\Delta X_t)_{t=1}^T$. Nous réitérons tous le processus d'étude de stationnarité sur cette nouvelle série. $(\Delta X_t)_{t=1}^T$ est bien stationnaire. $(\Delta X_t)_{t=1}^T$ sera donc la série utilisée pour modéliser notre modèle $ARMA(p, q)$. X_t sera donc défini par un $ARIMA(p, 1, q)$ avec p et q à définir.

1.3 Question 3

On représente tout d'abord la série non-différenciée $(X_t)_{t=1}^T$, avec $T=396$. On obtient alors la [Figure 1](#).

On observe que la production de bière en France ne semble pas avoir beaucoup évolué depuis les années 1990. Elle était environ de 100 dans les années 1990 et semble revenir à ce niveau depuis 2015. Entre ces deux périodes, la production de bière a connu une forte baisse, avec un minimum à la fin des années 2000, avant de reprendre une pente croissante. L'Association des Brasseurs de France explique cette reprise par une hausse de l'offre ainsi que sa diversification en France à partir des années 2010. La série ne semble pas stationnaire graphiquement, du fait de cette baisse de l'activité observée pendant une vingtaine d'années. On représente la série différenciée $(\Delta X_t)_{t=1}^T$ sur la [Figure 2](#).

Cette représentation graphique confirme les résultats du test augmenté de Dickey-Fuller : la série semble désormais stationnaire et peut être traitée avec un modèle ARMA. Cette dernière est centrée sur 0 et l'ampleur des variations de la série semble plutôt constant dans le temps.

2 Partie II : Modèles ARMA

2.1 Question 4

On applique la méthode Box-Jenkins afin de sélectionner le bon modèle $ARMA(p, q)$ sur notre série différenciée $(\Delta X_t)_{t=1}^T$. On représente d'abord la fonction d'autocorrélation (ACF) et la fonction d'autocorrélation partielle (PACF) afin de déterminer les ordres maximums vraisemblables p_{max} et q_{max} pour notre modèle. q_{max} se détermine à partir de l'ACF. En effet, un modèle MA d'ordre q implique que la fonction d'autocorrélation de la série modélisée est nulle pour tout h strictement supérieure à q , soit $MA(q) \Leftrightarrow \rho(h) = 0, \forall h > q$ où $\rho(h)$ représente l'ACF d'ordre h d'une série temporelle centrée et stationnaire. La [Figure 3](#) nous amène à sélectionner $q_{max} = 2$. La PACF d'ordre h , $\rho(h)$, représente l'autocorrélation entre ΔX_t et ΔX_{t-h} , corrigée de l'influence des variables intermédiaires. La PACF permet donc en d'observer la relation "pure" entre un ΔX_t avec ses lags, et nous permet ainsi de déterminer p_{max} , le nombre de lags maximum vraisemblable à considérer pour représenter notre série différenciée. Après représentation de la PACF sur la [Figure 3](#), on choisit $p_{max} = 9$. On estimera donc tous les modèles ARMA d'ordre inférieurs ou égaux à (p_{max}, q_{max}) :

$$(2.1) \quad \Delta X_t = \sum_{i=1}^p \phi_i \Delta X_{t-i} + \sum_{j=1}^q \psi_j \epsilon_{t-j} + \epsilon_t$$

Avec ϵ_t un bruit blanc.

Après estimation des 30 spécifications⁴, on ne conserve que celles qui sont valides et bien ajustées. On s'intéresse tout d'abord à la significativité statistique des coefficients associés aux ordres maximums. Ainsi, pour un $ARMA(p, q)$, on réalise un test de Student classique afin de déterminer si l'hypothèse de nullité des coefficients ϕ_p et ψ_q peut être rejetée. Si ce n'est pas le cas, par exemple pour le test sur le coefficient ϕ_p , on conclut qu'un modèle $ARMA(p-1, q)$ pourrait être plus adapté pour notre série, et on ne retient donc pas le modèle $ARMA(p, q)$. Une fois que l'on a vérifié l'ajustement des modèles, on ne conserve que ceux que l'on considère comme valides, à partir d'un test du Portmanteau. On souhaite conserver uniquement les modèles ne présentant pas d'autocorrélation dans les résidus. Pour cela, on compare la statistique $Q = n \sum_{h=1}^H \hat{\rho}_\epsilon^2(h)$, avec $\hat{\rho}_\epsilon^2$ la fonction d'autocorrélation des résidus, au quantile $\chi_{1-\alpha}^2(H - p - q)$. On ne conserve que les modèles pour lesquels $Q > \chi_{1-\alpha}^2(H - p - q)$.

4. Il y a en effet $p * q$ spécifications

En appliquant ces étapes pour notre série, avec $p_{max} = 9$ et $q_{max} = 2$, on retient trois modèles : ARMA(9,0), ARMA(2,1) et ARMA(2,2). Afin de sélectionner notre modèle finale, on s'intéresse aux critères d'informations AIC et BIC, permettant la pénalisation de notre modèle selon le nombre de paramètres utilisés afin d'éviter une sur-paramétrisation. On obtient le tableau suivant :

	ARMA(9,0)	ARMA(2,1)	ARMA(2,2)
AIC	2,631.025	2,626.257	2,624.127
BIC	2,674.792	2,646.152	2,648.000
Adjusted R^2	0.340	0.342	0.347

TABLE 1 – Critères d'information et R^2 ajusté des modèles valides et bien ajustés

L'ARMA(2,2) minimise l'AIC et l'ARMA(2,1) minimise le BIC. On exclu donc le modèle ARMA(9,0) de nos modèles finaux. Aucun de ces deux critères n'étant vraiment plus fiable que l'autre, on ne peut donc pas départager les deux modèles restants à partir de ces critères d'informations. BIC pénalise davantage le nombre de paramètre, tandis que AIC a tendance à favoriser les modèles plus complexes. Afin de choisir notre modèle final, on s'intéresse au R^2 ajusté de nos deux modèles, représentant la qualité de la prévision dans l'échantillon de ces derniers. ARMA(2,2) ayant un meilleur R^2 ajusté, on préférera ce modèle à l'ARMA(2,1). Le modèle retenu s'écrit donc, $\forall t \in [4, T]$:

$$(2.2) \quad \Delta X_t = \phi_1 \Delta X_{t-1} + \phi_2 \Delta X_{t-2} + \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2}$$

Ce modèle nous renvoie les estimations suivantes :

<i>Dependent variable :</i>	
	X_t
$\hat{\phi}_1$	1.002*** (0.163)
$\hat{\phi}_2$	-0.242*** (0.058)
$\hat{\psi}_1$	-1.713*** (0.165)
$\hat{\psi}_2$	0.760*** (0.138)
Observations	395
<i>Note :</i> *p<0.1 ; **p<0.05 ; ***p<0.01	

TABLE 2 – Estimations de l'ARMA(2,2)

2.2 Question 5

Le modèle estimé s'écrit, $\forall t \in [4, T]$:

$$(2.3) \quad \Delta X_t = \hat{\phi}_1 \Delta X_{t-1} + \hat{\phi}_2 \Delta X_{t-2} + \epsilon_t + \hat{\psi}_1 \epsilon_{t-1} + \hat{\psi}_2 \epsilon_{t-2}$$

En remplaçant les estimations du modèle par leur valeurs, on a alors :

$$(2.4) \quad \Delta X_t = 1.002 \cdot \Delta X_{t-1} - 0.242 \cdot \Delta X_{t-2} + \epsilon_t - 1.713 \cdot \epsilon_{t-1} + 0.760 \cdot \epsilon_{t-2}$$

En remplaçant les ΔX_t par $X_t - X_{t-1}$ notre modèle final ARIMA(2,1,2) s'écrit, $\forall t \in [4, T]$:

$$(2.5) \quad X_t = X_{t-1} + 1.002 \cdot (X_{t-1} - X_{t-2}) - 0.242 \cdot (X_{t-2} - X_{t-3}) + \epsilon_t - 1.713 \cdot \epsilon_{t-1} + 0.760 \cdot \epsilon_{t-2}$$

De plus, il est possible d'identifier les racines du polynôme associé à notre modèle ARMA(2,2) sur la série différenciée en résolvant le polynôme du second degré issue de la partie auto-régressive du modèle :

$$(2.6) \quad 0.242 \cdot x^2 + 1.002 \cdot x + 1 = 0$$

Un modèle est causal si toutes les racines de la partie AR sont en dehors du cercle unité. Avec notre modèle, les 2 racines obtenues (Voir [Figure 4](#)) sont effectivement en dehors du cercle unité ce qui permet de conclure que le modèle est bien causal.

3 Partie III : Prévision

3.1 Question 6

Notons $(Z_t)_{t=1}^T$ avec $Z_t = \Delta X_t$, notre série différenciée pour simplifier les expressions. Notre série différenciée suit un modèle ARMA(2,2) que l'on a estimé et donc vérifie l'équation $\forall t \in [4, T]$:

$$(3.1) \quad Z_t = \hat{\phi}_1 Z_{t-1} + \hat{\phi}_2 Z_{t-2} + \epsilon_t + \hat{\psi}_1 \epsilon_{t-1} + \hat{\psi}_2 \epsilon_{t-2}$$

Les meilleure prédiction de Z_{T+1} et Z_{T+2} sachant toutes les valeurs précédentes à T , T incluse, sont respectivement :

$$(3.2) \quad \hat{Z}_{T+1|T} = \hat{\phi}_1 Z_T + \hat{\phi}_2 Z_{T-1} + \hat{\psi}_1 \epsilon_T + \hat{\psi}_2 \epsilon_{T-1}$$

$$(3.3) \quad \hat{Z}_{T+2|T} = \hat{\phi}_1 \hat{Z}_{T+1} + \hat{\phi}_2 Z_T + \hat{\psi}_2 \epsilon_T$$

Avec $\epsilon_{t-i} = Z_{t-i} - \hat{Z}_{t-i}$ ⁵

On souhaite calculer pour chaque prédiction la variance de l'erreur de prédiction, *i.e* $V(Z_{T+1} - \hat{Z}_{T+1|T})$ et $V(Z_{T+2} - \hat{Z}_{T+2|T})$ pour construire l'intervalle de confiance.

Il faut donc préalablement calculer l'erreur de prédiction qu'on notera sous forme matricielle :

$$(3.4) \quad \begin{pmatrix} Z_{T+1} - \hat{Z}_{T+1|T} \\ Z_{T+2} - \hat{Z}_{T+2|T} \end{pmatrix} = \begin{pmatrix} \epsilon_{T+1} \\ \hat{\phi}_1(Z_{T+1} - \hat{Z}_{T+1|T}) + \epsilon_{T+2} + \hat{\psi}_1 \epsilon_{T+1} \end{pmatrix} \quad \text{6} = \begin{pmatrix} \epsilon_{T+1} \\ (\hat{\phi}_1 + \hat{\psi}_1) \epsilon_{T+1} + \epsilon_{T+2} \end{pmatrix} \quad \text{7}$$

La première ligne de la matrice vient logiquement de la définition des ϵ_t alors que la deuxième ligne est plus complexe car elle dépend de l'erreur de prédiction faite en $T+1$.

On peut maintenant calculer la variance grâce au fait que les résidus ϵ_t sont supposés ne pas être corrélés et avoir une variance constante : $V(\epsilon_t) = \sigma^2$. En effet, notre étude des résidus a montré que ceux-ci correspondaient à des bruits blancs ⁸. Donc :

$$V(Z_{T+1} - \hat{Z}_{T+1|T}) = V(\epsilon_{T+1}) = \sigma^2$$

et

$$V(Z_{T+2} - \hat{Z}_{T+2|T}) = V((\hat{\phi}_1 + \hat{\psi}_1) \epsilon_{T+1} + \epsilon_{T+2}) = \sigma^2(1 + (\hat{\phi}_1 + \hat{\psi}_1)^2)$$

5. il n'y a pas de « moving average » d'ordre 1 dans $\hat{Z}_{T+2|T}$ car on ne connaît pas Z_{T+1} .

6. car : $\begin{pmatrix} Z_{T+1} - \hat{Z}_{T+1|T} \\ Z_{T+2} - \hat{Z}_{T+2|T} \end{pmatrix} = \begin{pmatrix} \hat{\phi}_1 Z_T + \hat{\phi}_2 Z_{T-1} + \epsilon_{T+1} + \hat{\psi}_1 \epsilon_T + \hat{\psi}_2 \epsilon_{T-1} - (\hat{\phi}_1 Z_T + \hat{\phi}_2 Z_{T-1} + \hat{\psi}_1 \epsilon_T + \hat{\psi}_2 \epsilon_{T-1}) \\ \hat{\phi}_1 Z_{T+1|T} + \hat{\phi}_2 Z_T + \epsilon_{T+2} + \hat{\psi}_1 \epsilon_{T+1} + \hat{\psi}_2 \epsilon_T - (\hat{\phi}_1 \hat{Z}_{T+1} + \hat{\phi}_2 Z_T + \hat{\psi}_2 \epsilon_T) \end{pmatrix}$

7. car $Z_{T+1} - \hat{Z}_{T+1|T} = \epsilon_{T+1}$

8. Voir Question 7

Le vecteur $\begin{pmatrix} Z_{T+1} - \hat{Z}_{T+1|T} \\ Z_{T+2} - \hat{Z}_{T+2|T} \end{pmatrix}$ suit une $\mathcal{N}(0, \Sigma)$ car les résidus sont supposés gaussiens et sont donc indépendants par hypothèse. On peut ainsi écrire Σ comme :

$$(3.5) \quad \Sigma = \begin{pmatrix} \sigma^2 & (\hat{\phi}_1 + \hat{\psi}_1)\sigma^2 \\ (\hat{\phi}_1 + \hat{\psi}_1)\sigma^2 & \sigma^2(1 + (\hat{\phi}_1 + \hat{\psi}_1)^2) \end{pmatrix}^9$$

La matrice Σ est inversible car $\det(\Sigma) \neq 0$ ¹⁰. On peut donc calculer :

$$(3.6) \quad \begin{pmatrix} Z_{T+1} - \hat{Z}_{T+1|T} \\ Z_{T+2} - \hat{Z}_{T+2|T} \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} Z_{T+1} - \hat{Z}_{T+1|T} \\ Z_{T+2} - \hat{Z}_{T+2|T} \end{pmatrix} = \frac{\epsilon_{T+1}^2}{\sigma^2} + \frac{\epsilon_{T+2}^2}{\sigma^2}$$

Ainsi, en tant que somme du carré de deux variables indépendantes suivant des lois normales centrées et réduites, on a :

$$\begin{pmatrix} Z_{T+1} - \hat{Z}_{T+1|T} \\ Z_{T+2} - \hat{Z}_{T+2|T} \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} Z_{T+1} - \hat{Z}_{T+1|T} \\ Z_{T+2} - \hat{Z}_{T+2|T} \end{pmatrix} \sim \chi^2(2)$$

Et donc :

$$P\left[\begin{pmatrix} Z_{T+1} - \hat{Z}_{T+1|T} \\ Z_{T+2} - \hat{Z}_{T+2|T} \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} Z_{T+1} - \hat{Z}_{T+1|T} \\ Z_{T+2} - \hat{Z}_{T+2|T} \end{pmatrix} \leq q_{\chi^2(2)}(\alpha)\right] = \alpha$$

Où $q_{\chi^2(2)}(\alpha)$ est le α^{ieme} quantile d'une loi du χ^2 avec 2 degrés de libertés. La région de confiance au niveau α pour les prédictions (Z_{T+1}, Z_{T+2}) , est donc :

$$(3.7) \quad \left\{ \begin{pmatrix} x1 \\ x2 \end{pmatrix} \in R^2 \mid \begin{pmatrix} x1 - \hat{Z}_{T+1|T} \\ x2 - \hat{Z}_{T+2|T} \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} x1 - \hat{Z}_{T+1|T} \\ x2 - \hat{Z}_{T+2|T} \end{pmatrix} \leq q_{\chi^2(2)}(1 - \alpha) \right\}$$

De façon pratique, pour un niveau de confiance de 95%, on peut réécrire les 2 intervalles de confiance comme :

$$\begin{aligned} & [\hat{Z}_{T+1|T} - 1.96\hat{\sigma}^2; \hat{Z}_{T+1|T} + 1.96\hat{\sigma}^2] \\ & [\hat{Z}_{T+2|T} - 1.96\hat{\sigma}^2\sqrt{1 + (\hat{\phi}_1 + \hat{\psi}_1)^2}; \hat{Z}_{T+2|T} + 1.96\hat{\sigma}^2\sqrt{1 + (\hat{\phi}_1 + \hat{\psi}_1)^2}] \end{aligned}$$

Ce qui nous donne en remplaçant $\hat{\phi}_1$ et $\hat{\psi}_1$ par leur valeur numérique et $\hat{\sigma}^2$ par la variance empirique des résidus, ainsi qu'en ajoutant X_T à $\hat{Z}_{T+1|T}$ pour retrouver $\hat{X}_{T+1|T}$ et en ajoutant $\hat{X}_{T+1|T}$ à $\hat{Z}_{T+2|T}$ pour retrouver $\hat{X}_{T+2|T}$, on a :

$$[83.64124; 109.4807]$$

$$[83.53144; 96.98114]$$

3.2 Question 7

Afin d'obtenir cette région de confiance, on réalise les hypothèses suivantes :

- Les résidus de notre modèle ARIMA(2,1,2), sont supposés être indépendants, identiquement distribués et gaussiens. Cette hypothèse inclut également l'homoscédasticité des résidus du modèle, comme leur variance est supposée constante, égale à σ^2 . D'après la Figure 5, on voit que ces hypothèses semblent vérifiées, et que les résidus se comportent bien comme un bruit blanc. Excepté un point aberrant, on observe en effet une série de résidus homoscédastique, ne présentant pas d'autocorrélation significative, et suivant approximativement une gaussienne centrée en 0.

9. Car $Cov(Z_{T+1} - \hat{Z}_{T+1|T}, Z_{T+2} - \hat{Z}_{T+2|T}) = Cov(\epsilon_{T+1}, (\hat{\phi}_1 + \hat{\psi}_1)\epsilon_{T+1} + \epsilon_{T+2}) = (\hat{\phi}_1 + \hat{\psi}_1)Cov(\epsilon_{T+1}, \epsilon_{T+1}) = (\hat{\phi}_1 + \hat{\psi}_1)\sigma^2$ grâce à l'indépendance des ϵ_t .

10. $\det(\Sigma) = \sigma^4(1 + 2(\hat{\phi}_1 + \hat{\psi}_1)^2)$ et $\sigma > 0$ car la série n'est pas constante

- Les résultats de notre estimation avec l'ARIMA(2,1,2) sont supposés avoir bien identifié les relations autorégressives et de moyenne mobiles sous-jacentes à notre série. En d'autres termes, on obtient nos intervalles de confiances en postulant que les vrais ϕ_i et ψ_i ont été correctement estimés à partir des $\hat{\phi}_i$ et $\hat{\psi}_i$. Concrètement, les $\hat{\phi}_i$ et $\hat{\psi}_i$ doivent être à minima des estimateurs convergents, sans quoi nos intervalles de confiance sont asymptotiquement biaisés.

3.3 Question 8

On obtient la [Figure 6](#). L'intervalle de confiance observé étant assez large, peu de conclusions peuvent être tirées de ce graphique. Il est cependant rassurant d'observer que les vraies valeurs de la série se situent dans l'intervalle de confiance calculé.

De plus, nous avons représenté pour chaque date à partir de 2010 les prédictions mensuelles, sachant les valeurs réelles passées et les erreurs de prédiction passées, à partir de notre modélisation ARIMA(2,1,2) afin de les comparer directement à la série non différenciée. On obtient la [Figure 7](#). Les prédictions du modèle retenu semblent admettre moins de variation que la série en elle-même, et semble plutôt bien ajustée à cette dernière.

3.4 Question 9

Intuitivement, Y_{T+1} est utile à la prédiction de X_{T+1} si la prédiction avant et après l'introduction de cette nouvelle information dans nos données n'est pas la même. Nous utilisons ici le concept de causalité introduit par Granger en 1969. Selon Granger, la série Y_t est utile à la prédiction de X_t si Y_t cause X_t . Par ailleurs, Y_t cause X_t si et seulement si :

$$(3.8) \quad \hat{X}_{t+h|\{Y_u, X_u, u \leq t\}} \neq \hat{X}_{t+h|\{X_u, u \leq t\}}$$

Pour au moins un $h > 0$ et un $t \in \mathbb{Z}$. Il faut donc que pour un $h > 0$, les valeurs passées de Y_t en t modifient la prédiction de la série X_t pour la période $t+h$. Dans le cadre de cette question, on s'intéresse plus spécifiquement au fait de savoir si Y_{T+1} est utile à la prédiction de X_{T+1} . On parle alors de causalité instantanée au sens de Granger, Y_{T+1} cause X_{T+1} de manière instantanée si et seulement si :

$$(3.9) \quad \hat{X}_{T+1|\{Y_u, X_u, u \leq t\} \cup \{Y_{T+1}\}} \neq \hat{X}_{T+1|\{Y_u, X_u, u \leq t\}}$$

Afin de tester la causalité de Granger, une possibilité est d'effectuer un test de Wald. Ce test permet d'évaluer si les séries d'un Vector Autoregressive Model (VAR) se causent au sens de Granger. Si les séries étudiées sont stationnaires, il est possible de comparer une statistique du sous-espace vectoriel engendré par les colonnes de la matrice spécifiant la relation autorégressive des séries étudiées à un quantile de loi χ^2 . On rejette alors l'hypothèse que la série Y_t ne cause pas la série X_t avec une confiance de $1 - \alpha\%$ lorsque la statistique est supérieure au quantile d'ordre $1 - \alpha$ de la loi du χ^2 retenue.

4 Annexe

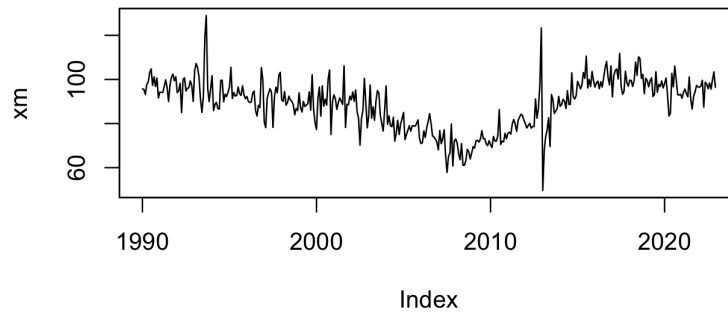


FIGURE 1 – Représentation de la série non différenciée de fabrication de bière en France

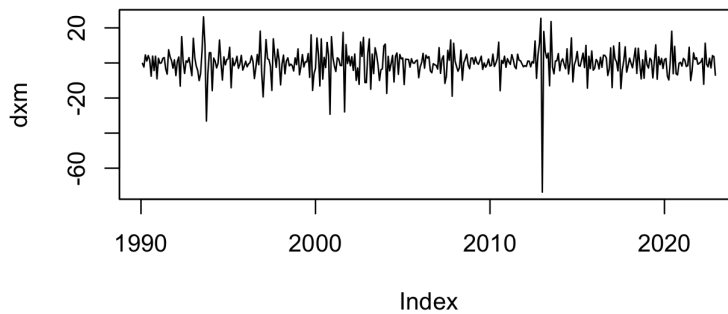


FIGURE 2 – Représentation de la série différenciée de fabrication de bière en France

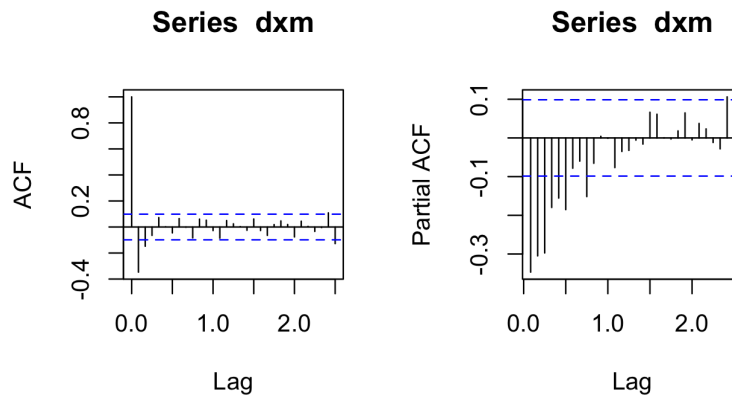


FIGURE 3 – ACF et PACF de la série différenciée de fabrication de bière en France

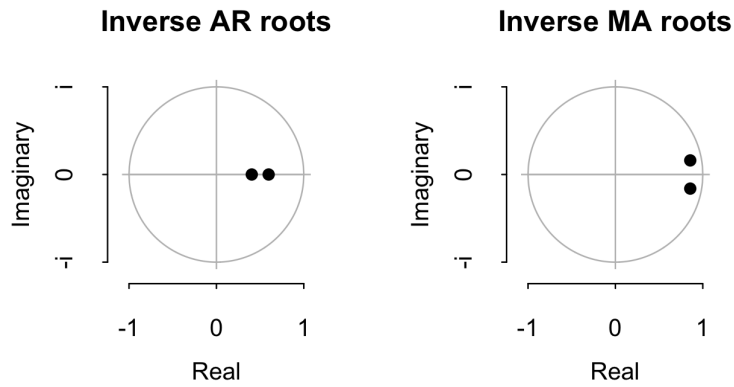


FIGURE 4 – Représentation de l'inverse des racines des polynômes de la partie autorégressive et de la partie moyenne mobile du modèle estimé dans le cercle unité

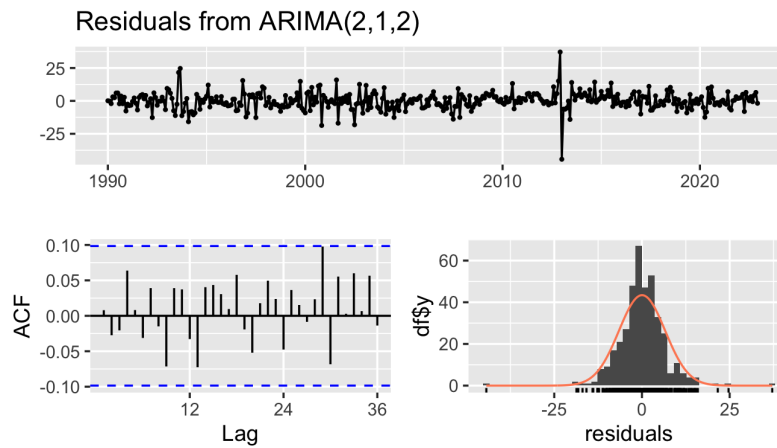


FIGURE 5 – Etude des résidus du modèle retenu ARMA(2,1,2)

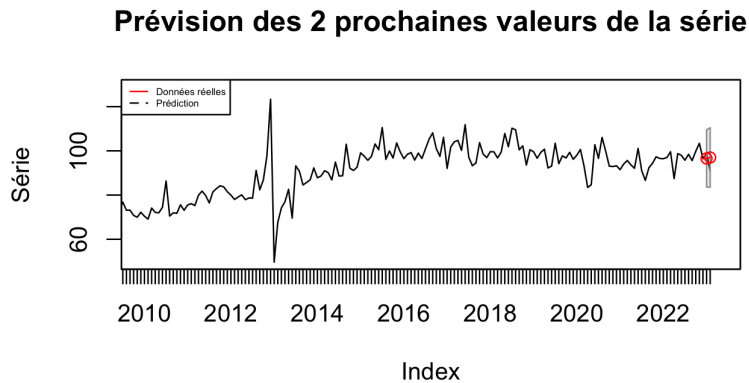


FIGURE 6 – Représentation de la série non différenciée avec les valeurs prédites et les intervalles de confiance à 95% correspondants

Comparaison de la série et des prédictions de l'ARIMA(2,1,2)

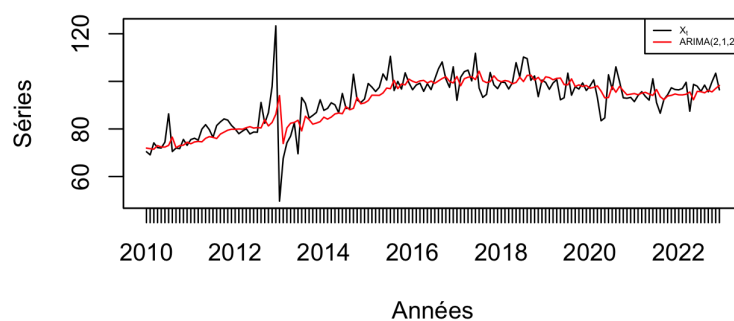


FIGURE 7 – Comparaison des prédictions de l'ARIMA(2,1,2) à partir de 2010 à la série non différenciée