

LEFRANC Romain

DENIZAN Maxime



ENSAE 2^{ème} année

Année scolaire 2020 - 2021

Projet de Séries Temporelles

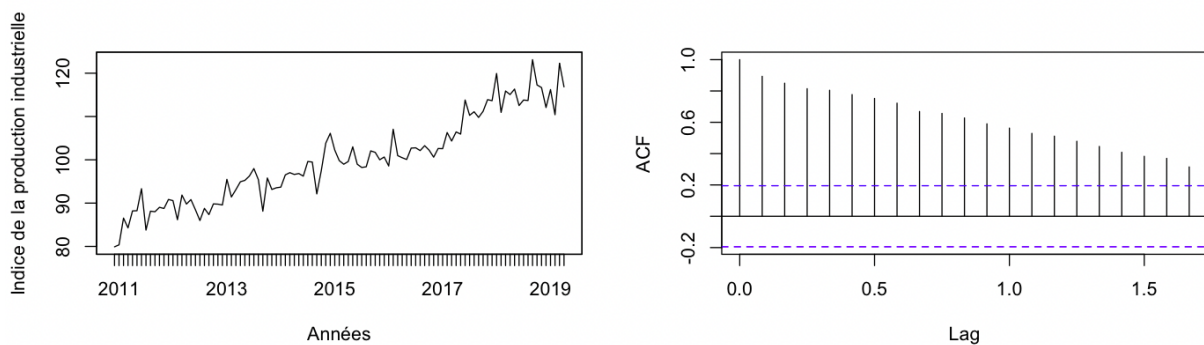
*Modélisation ARIMA de l'indice
de production industrielle*

14 mai 2021

Partie I - Les Données

Notre étude porte sur un indice de production industrielle corrigé des variations saisonnières et jours ouvrables (CVS-CJO). La série retenue permet de suivre l'évolution mensuelle de l'Indice CVS-CJO de la production industrielle (base 100) de parfums et de produits pour la toilette, poste 20.42 selon la classification de la NAF, entre novembre 2010 et mars 2019 et comporte donc 101 observations. Avant de commencer à étudier la série, il était nécessaire de la traiter afin de faciliter sa manipulation en la formalisant de manière pratique (Cf. Annexe II.1). Après ce traitement, nous avons représenté graphiquement la série avec et sa fonction d'auto-corrélation pour essayer d'identifier une saisonnalité ou une tendance. Nous la noterons dans la suite comme suit : $(X_t)_{t \in T}$, avec $T = 1, \dots, T$ l'ensemble contenant toutes les dates où la série est observée.

FIGURE 1 – Représentation graphique de la série étudiée - Indice CVS-CJO de la production industrielle de parfums et de produits de toilette entre le 11/2010 et le 03/2019 et son autocorrélogramme



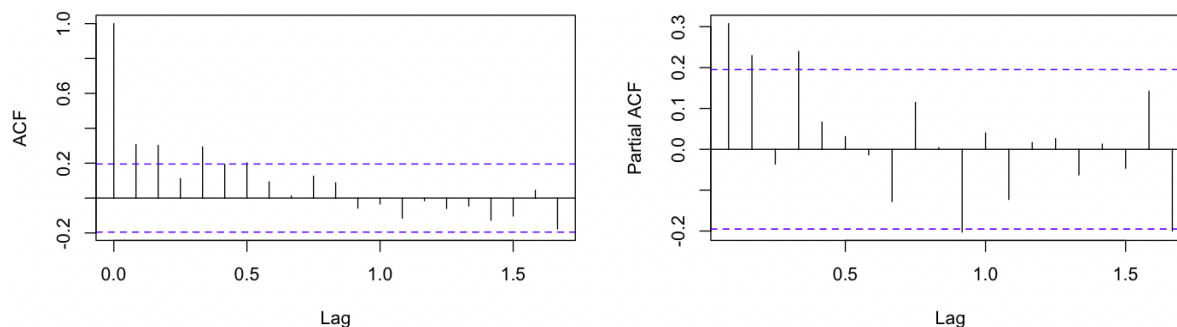
La série est représentée graphiquement sur la Figure 1 ci-dessus. Graphiquement, elle semble présenter une tendance linéaire, mais pas une saisonnalité. On a donc l'intuition que la série $(X_t)_{t \in T}$ n'est pas stationnaire. En outre, la décroissance lente du graphe de la fonction d'auto-corrélation obtenu semble mettre en évidence la non-stationnarité de la série $(X_t)_{t \in T}$. Pour vérifier la supposition de la tendance linéaire, nous avons régressé les indices sur ses dates.

TABLE 1 – Résultats de la régression linéaire des indices sur les dates

Coefficients	Estimate	Std. Error	t value	Pr(> t)
Intercept	-7738.8585	269.2530	-28.74	<2e-16
dates	3.8901	0.1336	29.11	<2e-16

Le coefficient associé à la tendance linéaire (dates) est effectivement bien significatif et positif, comme on s'y attendait. En notant les résidus de la régression comme $(Z_t)_{t \in T}$ (dont la représentation graphique est en Annexe I, Figure 8), ses fonctions d'autocorrélation et d'auto-corrélation partielle (Figure 2) nous montrent que l'autocorrélation d'ordre 1 - totale ou partielle, c'est la même chose - est d'un peu plus de 0.3, soit petite et loin d'être égale à 1 (plus de détails sur l'Annexe II.2). La série $(Z_t)_{t \in T}$ semble donc stationnaire.

FIGURE 2 – Fonctions d'autocorrélation et d'autocorrélation partielle de $(Z_t)_{t \in T}$, la série corrigée de sa tendance



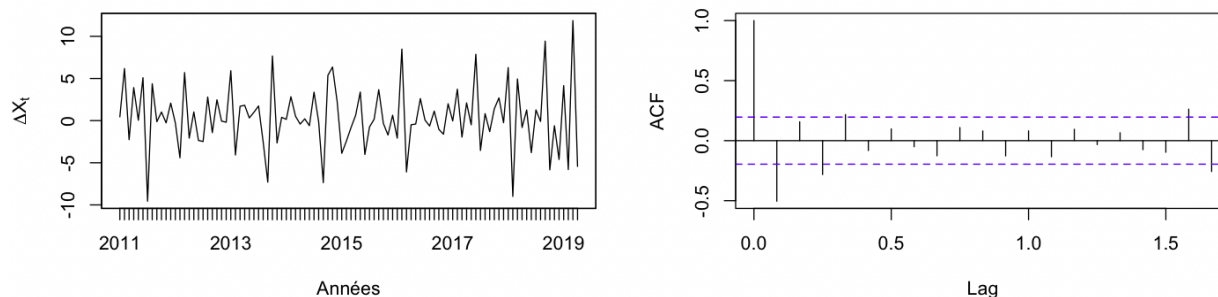
Par conséquent, la prochaine étape est d'appliquer un test de racine unitaire pour vérifier cette supposition. Nous avons donc appliqué le test de Dickey-Fuller Augmenté (Cf. Annexe III.1) avec constante et tendance non nulles étant donnée la significativité du coefficient de la régression effectuée. Cependant, pour que ce test soit valide, il faut s'assurer au préalable que les résidus sont bien non autocorrélés, ce qui a été fait et le résultat détaillé est montré dans l'Annexe II.3, où l'absence d'autocorrélation des résidus n'est pas rejetée à un seuil de 95% en considérant le test ADF avec 8 retards. Maintenant, on interprète le test de racine unitaire pour plus de sécurité.

TABLE 2 – Résultat du test de Dickey-Fuller augmenté sur la série avec 8 retards

PARAMETER : Lag Order	STATISTIC : Dickey-Fuller	P-value
8	-2.135	0.5207

La racine unitaire n'est pas rejetée au seuil de 95% pour (X_t) avec 8 retards ($p\text{-value} > 0.05$). Nous dirons que la série n'est pas stationnaire. La série est donc au moins intégrée d'ordre 1. Nous nous intéressons à présent à la série différenciée, que nous noterons $(\Delta X_t)_{t \in T}$ par la suite.

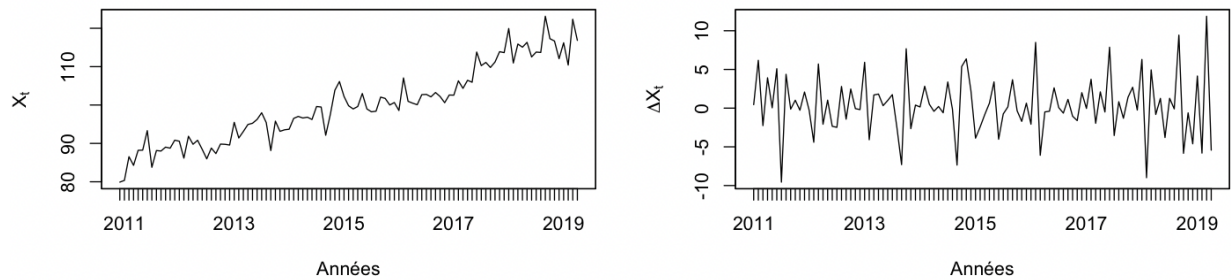
FIGURE 3 – Représentation graphique de la série différenciée $(\Delta X_t)_{t \in T}$ et son autocorrélogramme



De la même manière, nous nous intéressons à la régression de la série différenciée sur les dates (cf. Annexe II.4). Cette régression ne rejette pas l'hypothèse de nullité des coefficients, nous pouvons donc appliquer le test de Dickey-Fuller Augmenté (cf. Annexe II.4) dans le cas sans constante ni tendance, en vérifiant l'absence d'autocorrélation des résidus.

En considérant la série (ΔX_t) avec 6 retards, les résidus sont non autocorrélés et la racine unitaire est rejetée à un seuil de 95 %. Nous pouvons conclure que la série différenciée est stationnaire : la transformation sur la série originale est terminée. Ci-dessous, on peut voir notre série avant $(X_t)_{t \in T}$ et après $(\Delta X_t)_{t \in T}$ ces transformations.

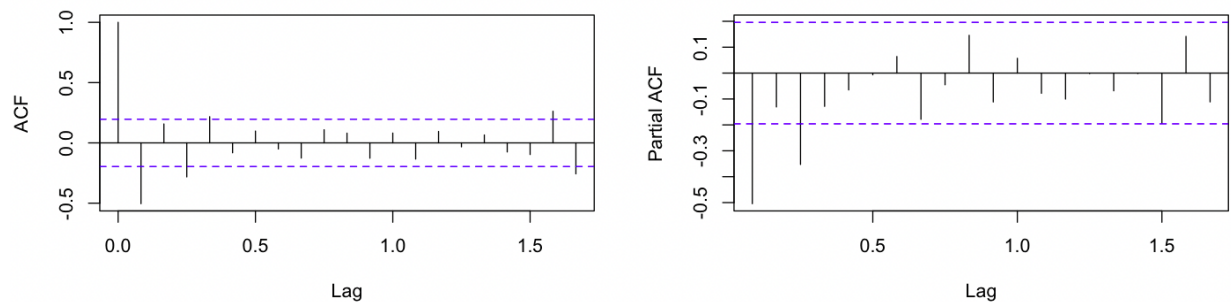
FIGURE 4 – Représentation graphique de la série $(X_t)_{t \in T}$ (série initiale, premier graphique), et de la série $(\Delta X_t)_{t \in T}$ (série différenciée, second graphique)



Partie II - Modèle ARMA

Du fait que l'hypothèse de stationnarité a été bien confirmée dans la Partie I et sans la nécessité d'élever une racine unitaire supplémentaire, la série corrigée $(\Delta X_t)_{t \in T}$ est intégrée d'ordre $d = 0$. La détermination des paramètres pertinents pour une modélisation se fait grâce aux auto-corrélogrammes (complets et partiels) disponibles ci-dessous :

FIGURE 5 – Fonctions d'autocorrélation et d'autocorrélation partielle de $(\Delta X_t)_{t \in T}$, notre série différenciée stationnaire



Par l'ACF et le PACF de la série corrigée $(\Delta X_t)_{t \in T}$ représentés sur la Figure 5, on voit que les fonctions d'auto-corrélation totales sont significatives (i.e. plus grandes que les bornes $\pm \frac{1.96}{\sqrt{n}}$ de l'intervalle de confiance d'un test de nullité de l'auto-corrélation à un seuil de 95%) jusqu'à $q^* = 20$ et les autocorrélations partielles le sont jusqu'à $p^* = 3$.

Alors, on peut conclure que si la série corrigée suit un modèle $\text{ARIMA}(p,d,q)$, elle suit au plus un $\text{ARIMA}(3,0,20)$. C'est pourquoi, les sous-modèles possibles sont les ARIMA tels que $p \leq p^*$, $d \leq d^*$ (ici $d^* = 0$), et $q \leq q^*$.

Cependant, pour qu'un modèle soit valide, il faut que ses résidus aient des propriétés similaires à celles des bruits blancs. C'est pourquoi, pour chaque modèle considéré, nous étudions sa validité en regardant la blancheur de ses résidus par le Test de Validité et la significativité des coefficients estimés par le Test de Nullité des Coefficients. (Cf. Annexe III.2 et III.3)

Finalement, nous avons effectué ces tests de façon itérative par un algorithme qui nous a permis d'obtenir 11 candidats potentiels, puis nous avons calculé leur AIC et leur BIC :

TABLE 3 – Les modèles ARMA(p,q) potentiels de la série corrigée, c'est-à-dire valides (résidus non autocorrélés) et bien ajustés (coefficients du modèle significatifs), leur AIC et leur BIC

Modèles candidats	AIC	BIC
ARMA(3,0)	515.0363	528.0621
ARMA(0,1)	514.9598	522.7753
ARMA(1,3)	514.4186	530.0496
ARMA(0,5)	514.6496	532.8858
ARMA(2,5)	515.0707	538.5172
ARMA(0,7)	514.8111	538.2576
ARMA(2,7)	513.8034	542.4603
ARMA(3,7)	512.5055	543.7676
ARMA(0,10)	517.6748	548.9368
ARMA(0,18)	510.7317	562.8351
ARMA(0,20)	514.7314	572.0451

Parmi tous ces modèles restants, on peut utiliser les critères d'information AIC ou BIC pour déterminer quel modèle est le plus adapté (i.e., celui qui minimise les critères), puisque ces critères mesurent la qualité d'un modèle statistique. À partir de la Table 3, nous pouvons conclure que le modèle ARIMA(0,1,18) minimise l'AIC pour la série $(X_t)_{t \in T}$ et le modèle ARIMA(0,1,1) minimise le BIC pour la série $(X_t)_{t \in T}$, ce qui n'est pas une surprise, car normalement BIC choisit le modèle le plus "économique".

On choisira de poursuivre avec le modèle ARIMA(0,1,1). L'expression du modèle proposé est alors donnée par :

$$(1 - B)X_t = (1 - 0,5931)\varepsilon_t$$

Partie III - Prévisions

Dans cette partie, on suppose que les résidus de la série sont Gaussiens et on note T la dernière date de notre série. On souhaite trouver l'équation vérifiée par la région de confiance de niveau α sur les futures valeurs (X_{T+1}, X_{T+2}) . On note $\Phi_1 = 0,5931$ le coefficient de notre modèle proposé. On a donc à partir de notre modélisation :

$$\Delta X_t = X_t - X_{t-1} = (1 - \Phi_1 B)\varepsilon_t = \varepsilon_t - \Phi_1 \varepsilon_{t-1} \quad (1)$$

On cherche la prévision optimale des valeurs futures (X_{T+1}, X_{T+2}) , notées (X_{T+1}^T, X_{T+2}^T) .

Avec l'équation (1) on a donc :

$$\begin{aligned} X_{T+1} &= X_T + \varepsilon_{T+1} - \Phi_1 \varepsilon_T \\ X_{T+2} &= X_{T+1} + \varepsilon_{T+2} - \Phi_1 \varepsilon_{T+1} \end{aligned} \quad (2)$$

Par ailleurs, la prévision linéaire optimale de X_{T+h} est donnée par :

$$X_{T+h}^T = \mathbb{E}L(X_{T+h}|X_T, \dots, X_1), \quad \text{ici pour } h \in \{1, 2\} \quad (3)$$

Comme ε_{T+1} et ε_{T+2} sont des innovations (car $\Phi_1 < 1$), ils sont des bruits blancs. Ils sont alors orthogonaux à $\mathcal{H}_X(T)$ et donc $\mathbb{E}L(\varepsilon_{T+1}|X_T \dots X_1) = \mathbb{E}L(\varepsilon_{T+2}|X_T \dots X_1) = 0$.

On peut ainsi écrire les prévisions linéaires optimales de (X_{T+1}, X_{T+2}) avec les formules (2) et (3) précédentes et en utilisant les propriétés des espérances conditionnelles :

$$\begin{aligned} X_{T+1}^T &= X_T - \Phi_1 \varepsilon_T \\ X_{T+2}^T &= X_{T+1}^T + 0 = X_T - \Phi_1 \varepsilon_T \end{aligned} \quad (4)$$

Remarque : on estime les deux valeurs par la même, mais les erreurs de prédictions différeront. Les erreurs de prédictions sont alors :

$$\begin{aligned} e_{T+1} &= X_{T+1} - X_{T+1}^T = \varepsilon_{T+1} \\ e_{T+2} &= X_{T+2} - X_{T+2}^T = \varepsilon_{T+2} + (1 - \Phi_1) \varepsilon_{T+1} \end{aligned} \quad (5)$$

Pour déterminer la région de confiance, nous avons besoin de calculer les variances et covariances de ces erreurs de prédiction dans le but d'obtenir la matrice variance-covariance Σ .

$$\text{Var}(e_{T+1}) = \text{Var}(\varepsilon_{T+1}) = \sigma_\varepsilon^2$$

$$\text{Var}(e_{T+2}) = \text{Var}(\varepsilon_{T+2} + (1 - \Phi_1) \varepsilon_{T+1}) = \sigma_\varepsilon^2 (1 + (1 - \Phi_1)^2)$$

$$\text{Cov}(e_{T+1}, e_{T+2}) = (1 - \Phi_1) \sigma_\varepsilon^2$$

Ici on a fait l'hypothèse que les processus d'innovation suivent une loi Gaussienne et qu'ils sont i.i.d. Par linéarité et stabilité de la loi Normale on obtient ainsi que :

$$W = \begin{pmatrix} e_{T+1} \\ e_{T+2} \end{pmatrix} \sim \mathcal{N}(0, \Sigma)$$

où

$$\Sigma = \begin{pmatrix} \sigma_\varepsilon^2 & (1 - \Phi_1) \sigma_\varepsilon^2 \\ (1 - \Phi_1) \sigma_\varepsilon^2 & \sigma_\varepsilon^2 \end{pmatrix}$$

Si l'on fait l'hypothèse que $\sigma_\varepsilon^2 > 0$, on a que Σ est inversible. On peut donc faire le test :

$$H_0 : W = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \text{contre} \quad H_1 : W \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Soit $q_{\chi_2^2}^{1-\eta}$ le quantile d'ordre $\alpha = 1 - \eta$ d'une loi de χ_2^2 . Le test d'hypothèse sera donc :

$$\Phi_n = \begin{cases} 0, & \text{si } \widehat{W} \Sigma^{-1} \widehat{W} < q_{\chi_2^2}^{1-\eta} \\ 1, & \text{sinon} \end{cases}$$

La statistique du test vérifie $W \Sigma^{-1} W \sim \chi_2^2$, \widehat{W} est la valeur empirique de W . Alors, la région de confiance de niveau α sur les valeurs futures (X_{T+1}, X_{T+2}) est donnée par

$$R_\alpha(X_1, \dots, X_T) = \left\{ \widehat{W} \Sigma^{-1} \widehat{W} < q_{\chi_2^2}^\alpha \right\}$$

Lorsqu'on a un test de niveau η , en prenant un point de vue univarié, on peut bien construire un intervalle de confiance au niveau $1 - \eta$ pour X_{T+1}^T et pour X_{T+2}^T .

Comme $e_{T+1} = \varepsilon_{T+1}$, l'erreur de prévision suit une loi $\mathcal{N}(\mu_\varepsilon, \sigma_\varepsilon^2)$. Soit $q_{\mathcal{N}(0,1)}^{1-\eta}$ le quantile d'ordre $1 - \eta$ d'une loi $\mathcal{N}(0, 1)$. Sous l'hypothèse $\mu_\varepsilon = 0$, nous pouvons donner un intervalle

de niveau de confiance $1 - \eta$ pour X_{T+1}^T par la formule :

$$IC_{1-\eta}(X_{T+1}^T) = \left[\hat{X}_{T+1}^T - q_{\mathcal{N}(0,1)}^{1-\eta} \sigma_\varepsilon ; \hat{X}_{T+1}^T + q_{\mathcal{N}(0,1)}^{1-\eta} \sigma_\varepsilon \right] \quad (6)$$

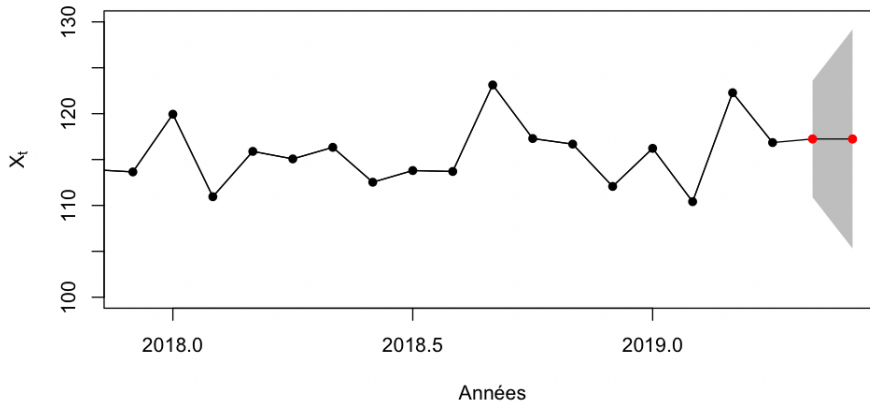
Par un raisonnement similaire, on a un intervalle de niveau de confiance $1 - \eta$ pour X_{T+2}^T par la formule :

$$IC_{1-\eta}(X_{T+2}^T) = \left[\hat{X}_{T+2}^T - q_{\mathcal{N}(0,1)}^{1-\eta} \sigma_\varepsilon \sqrt{1 + (1 - \Phi_1)^2} ; \hat{X}_{T+2}^T + q_{\mathcal{N}(0,1)}^{1-\eta} \sigma_\varepsilon \sqrt{1 + (1 - \Phi_1)^2} \right] \quad (7)$$

Pour obtenir ces intervalles de confiance et la région R_α , nous avons eu besoin de supposer plusieurs hypothèses. D'abord, nous avons supposé que la variance des résidus était connue et strictement positive. Si l'on ne connaissait pas la variance (σ_ε inconnu), il faudrait faire une estimation sur σ_ε en utilisant son estimateur de vraisemblance ou sa variance empirique. Dans les deux cas, la statistique utilisée pour obtenir les intervalles de confiance ne suit pas une loi Normale, mais suit une loi t-student qui peut être obtenue à partir du Théorème de Cochran. Nous avons également supposé que les résidus suivaient une loi normale et qu'ils étaient i.i.d., ce qui n'est pas normalement vrai en pratique. C'est pour cela qu'il existe des tests de Normalité comme le Test de χ_2 et le Test de Kolmogorov pour avoir hypothèse plausible avant de les considérer en loi normale. Enfin, nous avons supposé que notre modèle était parfaitement identifiable.

Le graphique ci-dessous donne la visualisation de la région de confiance (en grise) et les prédictions (en rouge) pour chaque horizon considéré pour = 95% de la production de parfums pour les mois d'avril et mai 2019.

FIGURE 6 – Prévision de (X_{T+1}, X_{T+2}) en utilisant un ARIMA(0,1,1) et intervalle de confiance univarié à 95 %



Maintenant on considère une série stationnaire $(Y_t)_{t \in T}$, où $(Y_{t+1})_{t \in T}$ est disponible plus rapidement que $(X_{t+1})_{t \in T}$. Dans ce cas là, $(Y_{t+1})_{t \in T}$ nous aidera à améliorer la prévision de $(X_{t+1})_{t \in T}$ juste si la série $(Y_t)_{t \in T}$ cause instantanément $(X_t)_{t \in T}$ au sens Granger, où la causalité se définit par l'utilité positive d'une variable pour en prévoir une autre. Mathématiquement, pour que ce cas arrive, il faut que :

$$\mathbb{E}L[X_{T+1} | (Y_{T+1}, \dots, Y_1), (X_T, \dots, X_1)] \neq \mathbb{E}L[X_{T+1} | (X_T, \dots, X_1)] \quad (8)$$

Cette condition se caractérise par la corrélation entre les deux résidus d'un modèle VAR de (X, Y) , et on peut tester cette hypothèse grâce à un test de Wald.

Annexes

Annexe I : Plot et ACF des séries étudiées

FIGURE 7 – (X_t) - Série initiale

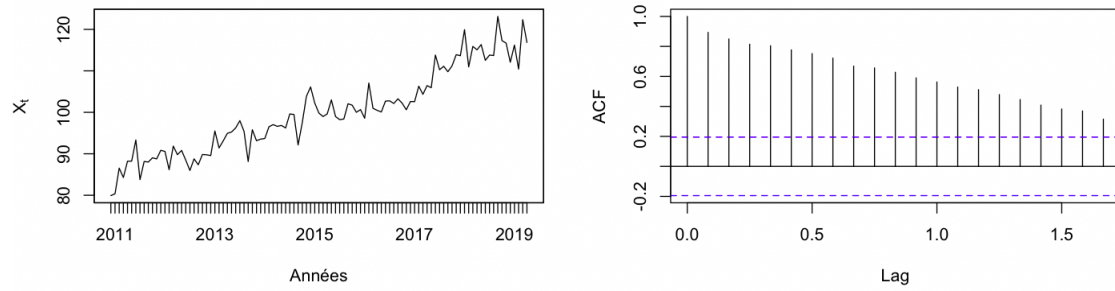


FIGURE 8 – (Z_t) - Série initiale corrigée de sa tendance

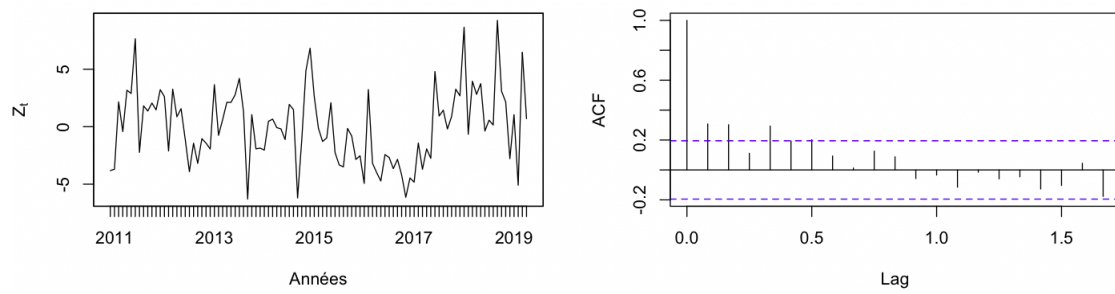
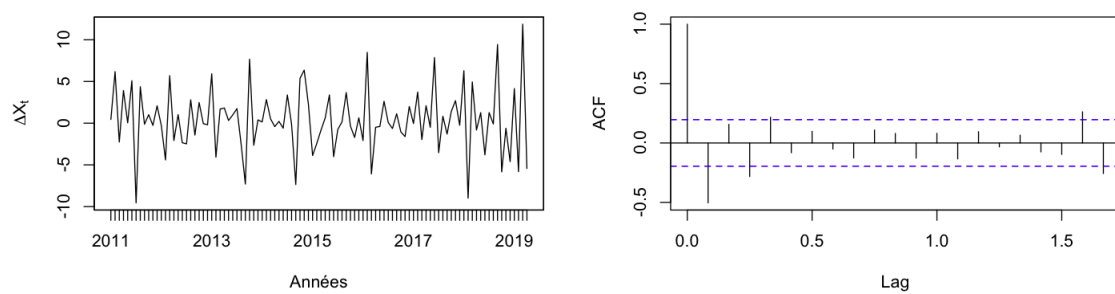


FIGURE 9 – (ΔX_t) - Série initiale différenciée



Annexe II : Résultats pratiques

1. Détails des données sur l'application R

Avant commencer à étudier la série, il était nécessaire de la traiter afin de faciliter sa manipulation en la formalisant de manière pratique. Pour cela, nous avons créé un "double", en utilisant le package "zoo", qui contient les Dates (le csv nous les a donnés en format "integer" et il faudrait changer pour "double") et ses respectives valeurs.

2. Autocorrélation au premier ordre

Si l'autocorrélation du premier ordre est proche de 1, on suspecte une non-stationnarité. Cependant, il est possible que l'autocorrélation soit très forte sans pour autant que la série soit non-stationnaire. Pour vérifier cela, on utilise un test de seconde génération. Le plus connu est le test KPSS. Comme dans notre série l'autocorrélation du premier ordre a été vers 0.3, nous n'avons pas eu besoin de faire ce test là.

3. Non-Autocorrélation des Résidus de la Régression et Test de la racine unitaire

Lorsqu'on fait une régression pour corriger la tendance linéaire de la série temporelle, pour que le test de racine soit valide afin de savoir si la série est stationnaire, il faut vérifier si les résidus de la régression ne sont pas autocorrélés. La tableau ci-dessous montre les p-valeurs du test qui contient l'absence de l'autocorrélation des résidus comme hypothèse nulle jusqu'à l'ordre 24 (deux ans). Les NAs représentent la correction des degrés de liberté du nombre de regressseurs. On voit que l'absence d'autocorrélation est systématiquement rejetée à un seuil de 95% (Q(4) à Q(24)).

	lag	pval
[1,]	1	NA
[2,]	2	NA
[3,]	3	NA
[4,]	4	0.0001063688
[5,]	5	0.0004084230
[6,]	6	0.0002671406
[7,]	7	0.0007508520
[8,]	8	0.0016529324
[9,]	9	0.0017911538
[10,]	10	0.0025365427
[11,]	11	0.0035660254
[12,]	12	0.0062977619
[13,]	13	0.0062871350
[14,]	14	0.0102458771
[15,]	15	0.0157595027
[16,]	16	0.0243784933
[17,]	17	0.0258107737
[18,]	18	0.0259455929
[19,]	19	0.0196185265
[20,]	20	0.0046779564
[21,]	21	0.0065284277
[22,]	22	0.0047596037
[23,]	23	0.0068067180
[24,]	24	0.0012336974

Une méthode itérative nous permet de réaliser le test de Dickey-Fuller jusqu'à ce que les résidus ne soient plus autocorrélés. Il a alors été nécessaire d'inclure 8 retards dans le test ADF. Pour autant, la racine unitaire n'est pas rejetée à un seuil de 95%.

4. Non-Autocorrélation des Résidus de la Régression de (ΔX_t) sur les dates et Test de la racine unitaire

Au regard de la représentation graphique de la série différenciée ΔX_t , on a l'intuition qu'elle ne présente ni constante ni tendance. Cela est confirmée par la régression de la série différenciée sur les dates.

TABLE 4 – Résultat de la régression linéaire de (ΔX_t) sur les dates

Coefficients	Estimate	Std. Error	t value	Pr(> t)
Intercept	46.09732	321.52744	0.143	0.886
dates[-1]	-0.02269	0.15956	-0.142	0.887

On peut alors tester la stationnarité de (ΔX_t) avec le test ADF dans le cas sans constante ni tendance, en vérifiant l'absence autocorrélation des résidus. On déduit qu'il faut 6 retards dans le test ADF pour qu'il soit valide. Enfin, le test de la racine unitaire sur la série différenciée avec 6 retards donne :

TABLE 5 – Résultat du test de Dickey-Fuller augmenté sur la série différenciée (ΔX_t) avec 6 retards

PARAMETER : Lag Order	STATISTIC : Dickey-Fuller	P-value
6	-3.685	0.01

Annexe III : Contenu théorique

1. Test de Dickey-Fuller Augmenté

Le test augmenté de Dickey-Fuller ou test ADF est un test statistique qui vise à savoir si une série temporelle est stationnaire, c'est-à-dire si ses propriétés statistiques (espérance, variance, auto-corrélation) varient ou non dans le temps. Ce processus est réalisé en vérifiant s'il y a des racines unitaires dans l'équation caractéristique de la série. En effet, si la série présente une racine unitaire, sa moyenne et variance ne sont pas constantes par rapport au temps.

Une série avec une racine unitaire est non-stationnaire. On a aussi que si les autres racines de l'équation caractéristique sont dans le cercle unitaire (de module plus petit que 1), alors faire une différence première de la série suffit à la rendre stationnaire. Dans notre série, c'est le cas car les résidus de la régression linéaire sont auto corrélés et, donc, nous avons appliqué le Test de Dickey-Fuller Augmenté sans constante ni tendance.

Alors, le test est donné par $H_0 : \Phi = 0$ contre $H_1 : \Phi \neq 0$, où si $\Phi = 0$ on aura au moins une racine unitaire.

Le modèle ADF pour une série $(X_t)_{t \in T}$ est donnée par :

$$\Delta X_t = \Phi X_t + \sum_{l=1}^k \alpha_l \Delta X_{t-l} + \varepsilon_t$$

où k est le nombre de retards nécessaires à considérer pour rendre des résidus non- corrélés (dans notre cas, $k = 6$ avec pour série la série déjà différenciée $(\Delta X_t)_{t \in \{2, \dots, T\}}$.

La statistique du test est associée à l'estimation du coefficient Φ donnée par :

$$\text{t-stat} = \frac{\hat{\Phi}}{\hat{\sigma}(\hat{\Phi})}$$

Si t-stat est supérieure à la valeur critique associée au test 5% alors on ne peut pas rejeter l'hypothèse nulle de racine unitaire (donc de non-stationnarité) de la série à 5%. Le test ADF conclura donc à la non-stationnarité de la série en niveau.

Dans notre cas, nous avons obtenu un résultat concluant pour la stationnarité de $(\Delta X_t)_{t \in \{2, \dots, T\}}$.

2. Test de Validité (Ljung-Box)

Le modèle est valide si ses résidus ne sont pas auto-corrélés. Nous l'avons testé en utilisant le Test de Ljung-Box de l'hypothèse nulle de nullité jointe des auto corrélations jusqu'à un ordre k donné qui utilise une statistique de test qui suit une loi χ^2 à $k - p - q$ degrés de liberté. On ne peut donc l'effectuer que dans la situation où $k > p + q$. Plus précisément, le but est de tester $H_0 : \rho_l = 0 \quad \forall l \in \{p + q, \dots, T\}$ contre $H_1 : \exists l \in \{p + q, \dots, T\} \quad \rho_l \neq 0$

La statistique de ce test s'écrit :

$$Q(k) = T(T+2) \sum_{l=1}^k \frac{\hat{\rho}_l}{T-l}$$

3. Test de nullité des coefficients

On mesure la significativité des coefficients par le rapport entre le coefficient estimé et l'écart-type estimé du coefficient estimé. Pour que le modèle soit bien ajusté, la p-valeur correspondant de ce rapport doit être inférieure à 0.05.