Projet de Séries Temporelles

Kim Antunez et Alain Quartier-la-Tente

31/03/2020

Contents

1 Partie 1 : Les données **1**
1.1 Question 1 : description de la série choisie 1
1.2 Questions 2 et 3 2

2 Partie 2 : Modèles ARIMA **3**

3 Partie 3 : Prévisions **4**
3.1 Question 6, 7 et 8 : construction d'un intervalle de confiance 4
3.2 Question 9 : question ouverte 5

A Annexe 1 : tests supplémentaires sur la qualité des modèles **5**

1 Partie 1 : Les données

1.1 Question 1 : description de la série choisie

Pour ce projet, nous avons choisi de travailler sur la série d'indice de production industrielle (IPI) dans l'industrie automobile (identifiant : 010537940). Il s'agit d'une série au niveau A64 de la nomenclature d'activités française révision 2 (NAF rév. 2), poste CL1. Cela concerne aussi bien la production des constructeurs de voitures particulières, de véhicules de loisir, de véhicules utilitaires que les équipementiers spécialisés, les carrossiers, les assembleurs ou les prestataires de services d'aménagement de véhicules automobiles. Cette production intègre donc la filière complète, y compris moteurs et organes mécaniques en amont, dès lors qu'ils sont principalement destinés à des véhicules automobiles (à l'exception des parties de moteur).

C'est un indice de Laspeyres chaîné avec des pondérations annuelles en valeur ajouté. Il est de base 2015. L'IPI dans l'industrie automobile est calculé à partir de l'enquête mensuelle de branche. Il est calculé par agrégation de séries "élémentaires" calculées à un niveau plus fin. Ces séries élémentaires sont estimées en volume : la série d'IPI dans l'industrie automobile ne tient donc pas compte des variations de prix.

Les séries de l'IPI sont corrigées des variations saisonnières et des jours ouvrables (CVS-CJO) à partir de la méthode X13-ARIMA. La désaisonnalisation est réalisée de manière indirecte : elle est effectuée à un niveau fin et les agrégats CVS-CJO sont ensuite calculés directement à partir de ces séries en agrégeant les séries CVS-CJO. Cette désaisonnalisation est réalisée par sous-période pour prendre en compte le fait que la structure économique des séries a beaucoup évolué en 30 ans, et donc qu'il serait peut pertinent d'appliquer un seul modèle de désaisonnalisation sur l'ensemble de la période. Ainsi, les modèles utilisés pour la désaisonnalisation commencent en 2005 et ces modèles sont utilisées pour estimer les séries CVS-CJO à partir de 2012.

Les séries CVS-CJO avant et après 2012 n'étant pas évalués sur les mêmes modèles, et pour éviter des ruptures liées à ce changement de modèle, l'idéal serait d'étudier notre série après janvier 2012. En revanche, cela laisserait une faible profondeur temporelle risquant de fragiliser l'estimation de nos modèles ARIMA. C'est pourquoi nous allons étudier la série d'IPI dans l'industrie automobile entre **janvier 2010 et décembre 2019**[1], c'est-à-dire sur **120 observations**.

Nous n'effectuerons pas de correction de point atypique ou de transformation logarithmique.

1.2 Questions 2 et 3

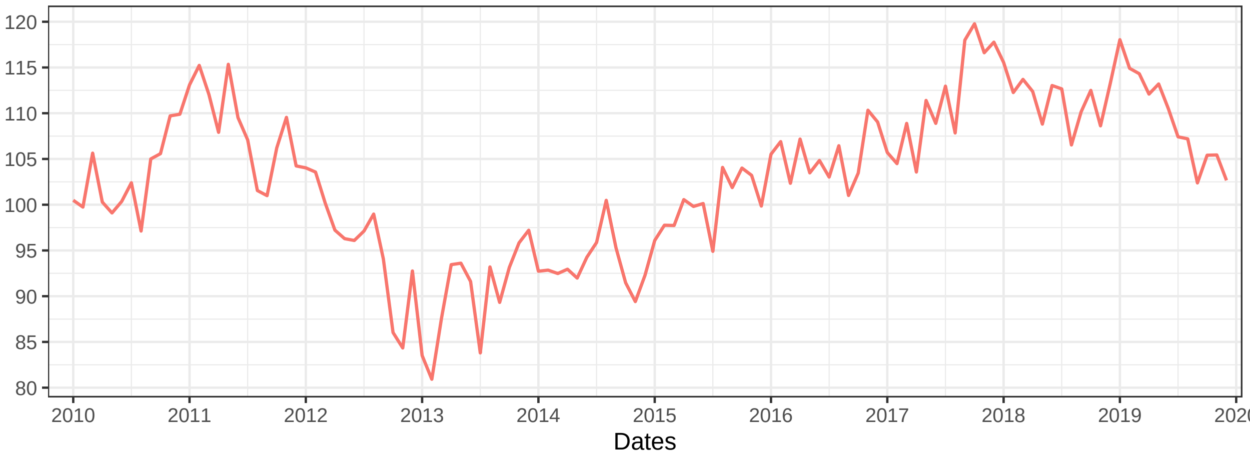

Figure 1: IPI dans l'automobile (CVS-CJO sans traitement)

Le graphique 1 ne montre pas de tendance linéaire déterministe nette sur la période 2010-2020 : on observe plutôt une alternance entre des périodes de tendance croissante (2010-2011, 2013-2018) et de tendance décroissance (2011-2013 et 2018-2020). La série de l'IPI dans l'automobile semble plutôt montrer une tendance stochastique : elle n'est sûrement pas stationnaire. Ceci est vérifié en faisant le test Dickey-Fuller augmenté (ADF) dans le cas avec une constante non nulle et sans tendance, puisqu'on ne rejette pas l'hypothèse de présence de racine unitaire au seuil de 5 % (tableau 1). C'est également confirmé par le test de racine unitaire de Phillips-Perron, non rejeté au seuil de 5 %, et par le test de racine unitaire de Kwiatkowski-Phillips-Schmidt-Shin (KPSS), rejeté au seuil de 5 % (ici l'hypothèse alternative est la non-stationnarité de la série).

Table 1: Tests de racine unitaire et de stationnarité sur la série d'IPI dans l'automobile

Test	Statistique	p-valeur	
Dickey-Fuller augmenté [a]	-1.678	0.434	
Phillips-Perron	-2.578	0.336	
KPSS	0.892	0.010	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

[a] Le test ADF a été fait en rajoutant 2 retards. De cette façon les résidus utilisés dans ce test sont bien indépendants et le test ADF est bien interprétable

D'après le graphique 1, la série différenciée semble stationnaire. C'est confirmé par le test de Dickey-Fuller augmenté, effectué avec une constante nulle et sans tendance, le test de Phillips-Perron et le test KPSS (tableau 2).

[1] Les derniers points étant souvent sujets à révisions, nous avons préféré ne pas prendre en compte les points de janvier et février 2020.

Table 2: Tests de racine unitaire et de stationnarité sur la série différenciée d'IPI dans l'automobile

Test	Statistique	p-valeur	
Dickey-Fuller augmenté [a]	-10.261	0.010	**
Phillips-Perron	-15.132	0.010	**
KPSS	0.074	0.100	.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

[a] Le test ADF a été fait en rajoutant 1 retard. De cette façon les résidus utilisés dans ce test sont bien indépendants et le test ADF est bien interprétable

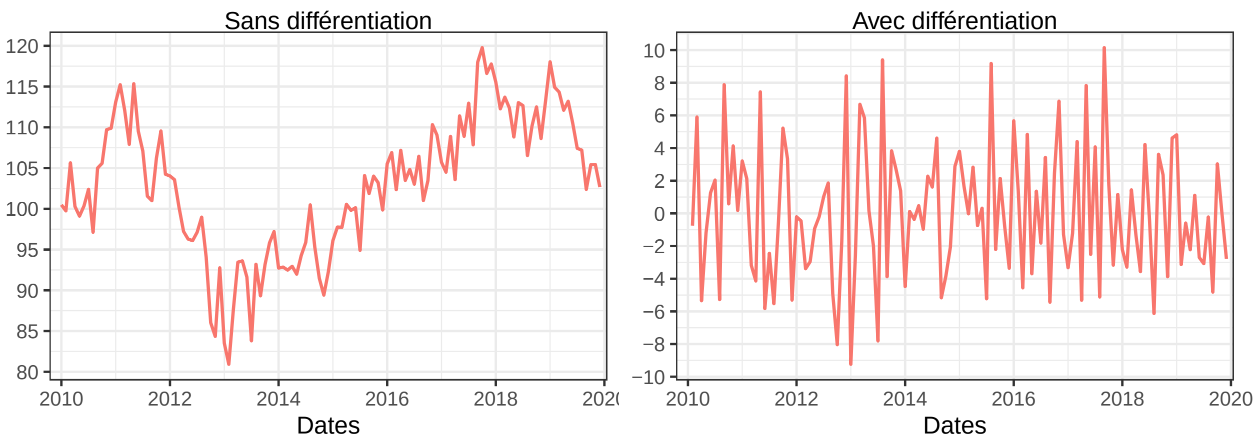

Figure 2: IPI dans l'automobile (CVS-CJO) sans et avec différentiation

2 Partie 2 : Modèles ARIMA

Pour déterminer les ordres maximaux, p_{max} et q_{max}, du modèle $ARMA(p, q)$ suivi par la série différenciée de l'IPI dans l'automobile, analysons les autocorrélogrammes et les autocorrélogrammes partiels (graphique 3). À partir de retard 2 (inclus), tous les autocorrélogrammes ne sont pas significatifs à 5 % : on en déduit que $p_{max} = 1$. À partir de retard 2 (inclus), tous les autocorrélogrammes partiels ne sont pas significatifs à 5 % : on en déduit que $q_{max} = 1$. Pour savoir quel modèle choisir nous allons tester tous les modèles $ARMA(p, q)$ tels que $p \leq 1$ et $q \leq 1$.

Quatre modèles ARMA ont donc été testés, sans ajout de constante :

- *ARMA(0,0)* : les résidus de ce modèle ne sont pas indépendants (tableau 4) -> **modèle non retenu**
- *ARMA(1,0)* : les résidus de ce modèle ne sont pas indépendants (tableau 4) -> **modèle non retenu**
- *ARMA(0,1)* : les résidus de ce modèle sont indépendants (tableau 4) et le coefficient associé au MA(1) est significatif (tableau 7) -> **modèle retenu**
- *ARMA(1,1)* : les résidus de ce modèle sont indépendants (tableau 4) mais le coefficient associé au AR(1) n'est significatif (tableau 7) -> **modèle non retenu**

En somme il n'y a qu'un seul modèle ARMA valide sur l'IPI différencié : c'est le modèle ARMA(0,1). Pour l'IPI automobile, X_t on retient donc le modèle ARIMA(0,1,1) suivant :

$$\Delta X_t = \varepsilon_t - \underset{(0{,}09)}{0{,}38}\ \varepsilon_{t-1}$$

ε_t est bien un bruit blanc : les $(\varepsilon_t)_t$ sont indépendants (tableau 4), homoscédastiques (tableau 5) et suivent aussi une loi normale (tableau 6).

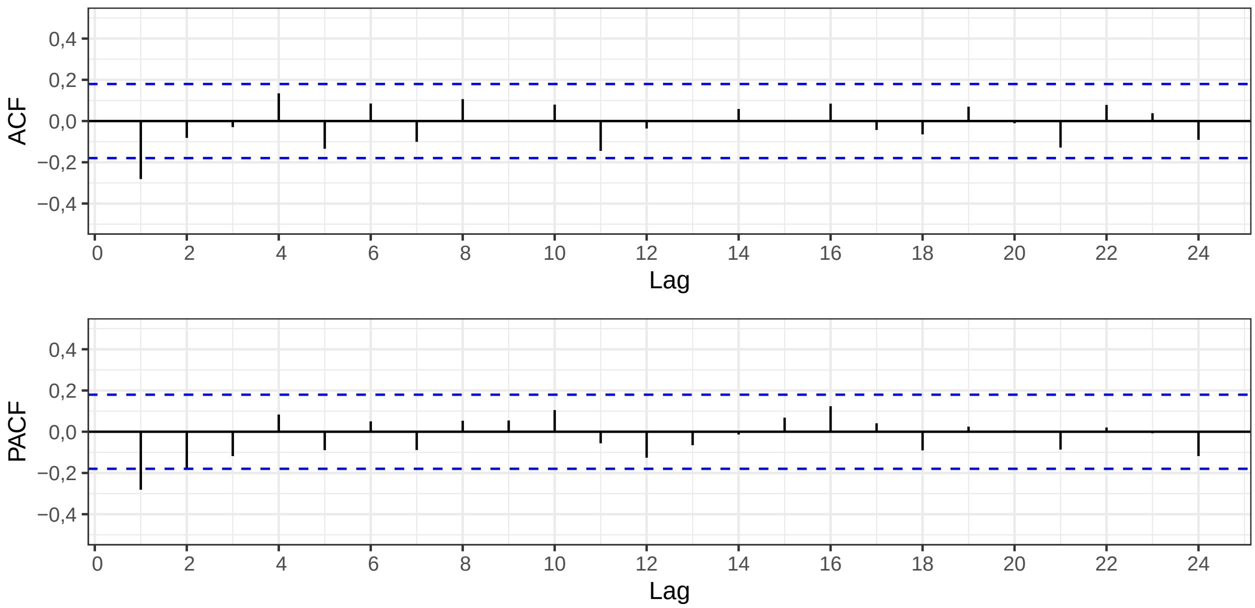

Figure 3: Autocorrélogrammes (ACF) et autocorrélogrammes partiels (PACF) pour la série différenciée de l'IPI dans l'automobile

Parmi l'ensemble des modèles testés, l'ARIMA(0,1,1) est aussi le modèle qui minimise les critères d'information (tableau &ref(tab:aicbic)).

Table 3: Critères d'information des modèles ARIMA sur l'IPI de l'automobile

	ARIMA(0,1,0)	ARIMA(1,1,0)	ARIMA(0,1,1)	ARIMA(1,1,1)
AIC	672.439	664.677	660.932	662.345
BIC	675.219	670.235	666.490	670.683

3 Partie 3 : Prévisions

3.1 Question 6, 7 et 8 : construction d'un intervalle de confiance

On cherche désormais à faire une prévision de X_t à l'horizon $T+2$. Notons $\hat{\theta}_1$ le coefficient associé à la partie MA de notre modèle ARMA(0,1,1), estimé entre janvier 2010 et décembre 2019 ($\hat{\theta}_1 \simeq -0,38$). On a :

$$\Delta X_T = \varepsilon_T + \hat{\theta}_1 \varepsilon_{T-1} \iff X_T = X_{T-1} + \varepsilon_T + \hat{\theta}_1 \varepsilon_{T-1}$$

Les prévisions de X_{T+1} et X_{T+2} réalisées à l'instant T, notées $\hat{X}_{T+1|T}$ et $\hat{X}_{T+2|T}$, vérifient donc l'équation :

$$\begin{cases} \hat{X}_{T+1|T} = X_T - \hat{\theta}_1 \varepsilon_T \\ \hat{X}_{T+2|T} = \hat{X}_{T+1|T} \end{cases}$$

Les erreurs de prévision sont égales à :

$$\begin{cases} \hat{\varepsilon}_{T+1|T} = X_{T+1} - \hat{X}_{T+1|T} = \varepsilon_{T+1} + (\theta_1 - \hat{\theta}_1)\varepsilon_T \\ \hat{\varepsilon}_{T+2|T} = X_{T+2} - \hat{X}_{T+2|T} = \varepsilon_{T+2} + (1+\theta_1)\varepsilon_{T+1} + (\theta_1 - \hat{\theta}_1)\varepsilon_T \end{cases}$$

Les intervalles de confiance de niveau α sur $\hat{X}_{T+1|T}$ et $\hat{X}_{T+2|T}$ s'écrivent donc :

$$IC_{1-\alpha}(\hat{X}_{T+h|T}) = \left[\hat{X}_{T+h|T} - q_{1-\frac{\alpha}{2}}\hat{\sigma}_h \ ; \ \hat{X}_{T+h|T} + q_{1-\frac{\alpha}{2}}\hat{\sigma}_h\right]$$

Avec $q_{1-\frac{\alpha}{2}}$ le quantile $1-\frac{\alpha}{2}$ d'une loi $\mathcal{N}(0,1)$ et :

$$\hat{\sigma}_1 = \hat{\sigma}, \quad \hat{\sigma}_2 = \hat{\sigma}\sqrt{1+(1+\hat{\theta}_1)^2} \quad \text{et} \quad \hat{\sigma} = \frac{1}{T-2}\sum_{t=2}^{T}\hat{\varepsilon}_t^2$$

(la somme commence à 2 puisque l'on perd une période en différenciant)

Bilan :

$$IC_{1-\alpha}\left(\begin{pmatrix}\hat{X}_{T+1|T}\\ \hat{X}_{T+2|T}\end{pmatrix}\right) = \left[\begin{pmatrix}X_T-\hat{\theta}_1\varepsilon_T\\ X_T-\hat{\theta}_1\varepsilon_T\end{pmatrix} - \hat{\sigma}q_{1-\frac{\alpha}{2}}\begin{pmatrix}1\\ \sqrt{1+(1+\hat{\theta}_1)^2}\end{pmatrix} ; \begin{pmatrix}X_T-\hat{\theta}_1\varepsilon_T\\ X_T-\hat{\theta}_1\varepsilon_T\end{pmatrix} + \hat{\sigma}q_{1-\frac{\alpha}{2}}\begin{pmatrix}1\\ \sqrt{1+(1+\hat{\theta}_1)^2}\end{pmatrix}\right]$$

Pour obtenir ces intervalles de confiance il faut que les résidus de notre modèle ARIMA soient indépendants, homoscédastiques et suivent une loi normale : ce qui est bien vérifié dans notre cas.

3.2 Question 9 : question ouverte

Granger

A Annexe 1 : tests supplémentaires sur la qualité des modèles

Table 4: Tests de Ljung-Box sur les résidus (tests d'autocorrélation) des modèles ARIMA sur l'IPI de l'automobile

	ARIMA(0,1,0)			ARIMA(1,1,0)			ARIMA(0,1,1)		ARIMA(1,1,1)	
Retards	Statistique	p-valeur		Statistique	p-valeur		Statistique	p-valeur	Statistique	p-valeur
1	9.652	0.002	**							
2	10.473	0.005	**	4.757	0.029	*	0.858	0.354		
3	10.580	0.014	*	4.796	0.091	.	0.925	0.630	0.106	0.744
4	12.843	0.012	*	6.191	0.103		1.985	0.576	1.544	0.462
5	15.122	0.010	**	7.239	0.124		3.024	0.554	2.606	0.456
6	16.041	0.014	*	7.327	0.197		3.165	0.675	2.907	0.573
7	17.341	0.015	*	7.800	0.253		3.566	0.735	3.389	0.640
8	18.808	0.016	*	8.936	0.257		4.884	0.674	4.676	0.586
9	18.809	0.027	*	9.306	0.317		5.129	0.744	4.781	0.687
10	19.645	0.033	*	9.610	0.383		5.338	0.804	5.037	0.754
11	22.433	0.021	*	12.895	0.230		8.684	0.562	8.103	0.524
12	22.609	0.031	*	13.822	0.243		9.649	0.562	8.787	0.552
13	22.610	0.047	*	13.838	0.311		9.649	0.647	8.787	0.642
14	23.078	0.059	.	14.496	0.340		10.455	0.656	9.497	0.660
15	23.082	0.082	.	14.860	0.388		11.003	0.686	9.871	0.704
16	24.077	0.088	.	15.947	0.386		12.232	0.661	11.074	0.680
17	24.339	0.111		16.263	0.435		12.411	0.715	11.227	0.736
18	24.935	0.127		16.911	0.460		12.993	0.737	11.795	0.758
19	25.633	0.141		17.430	0.494		13.173	0.781	11.966	0.802
20	25.649	0.178		17.579	0.551		13.432	0.816	12.208	0.836
21	28.080	0.138		20.078	0.453		15.782	0.730	14.643	0.745
22	28.992	0.145		20.694	0.478		16.064	0.766	14.878	0.783
23	29.209	0.173		20.920	0.526		16.146	0.809	14.914	0.827
24	30.464	0.170		21.907	0.526		17.136	0.803	16.104	0.811

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 5: Tests de Ljung-Box sur le carré des résidus (tests d'homoscédasticité) des modèles ARIMA sur l'IPI de l'automobile

	ARIMA(0,1,0)		ARIMA(1,1,0)		ARIMA(0,1,1)		ARIMA(1,1,1)	
Retards	Statistique	p-valeur	Statistique	p-valeur	Statistique	p-valeur	Statistique	p-valeur
1	2.832	0.092 .						
2	2.843	0.241	2.917	0.088 .	2.032	0.154		
3	2.860	0.414	3.844	0.146	3.569	0.168	3.140	0.076 .
4	3.227	0.521	5.425	0.143	4.164	0.244	3.575	0.167
5	3.233	0.664	5.448	0.244	4.183	0.382	3.576	0.311
6	3.262	0.775	7.479	0.187	6.916	0.227	5.513	0.239
7	3.263	0.860	7.836	0.250	7.515	0.276	6.127	0.294
8	3.270	0.916	8.294	0.307	7.781	0.352	6.290	0.392
9	3.772	0.926	8.613	0.376	7.787	0.455	6.314	0.504
10	3.797	0.956	8.727	0.463	7.838	0.551	6.605	0.580
11	5.286	0.917	9.343	0.500	8.071	0.622	6.897	0.648
12	5.482	0.940	10.431	0.492	9.136	0.609	7.987	0.630
13	5.619	0.959	11.058	0.524	9.537	0.656	8.373	0.680
14	7.532	0.912	12.221	0.510	10.714	0.635	9.807	0.633
15	7.727	0.934	12.252	0.586	10.784	0.703	9.807	0.710
16	7.760	0.956	12.256	0.660	10.802	0.766	9.823	0.775
17	7.866	0.969	12.268	0.725	10.968	0.811	9.890	0.827
18	11.403	0.876	12.791	0.750	11.573	0.825	11.468	0.780
19	11.440	0.908	12.791	0.804	11.788	0.858	11.709	0.817
20	12.066	0.914	13.115	0.833	12.156	0.879	12.086	0.843
21	12.213	0.934	14.050	0.828	13.228	0.867	12.984	0.839
22	13.751	0.910	15.070	0.819	14.231	0.859	14.364	0.812
23	14.907	0.898	15.242	0.852	14.284	0.891	14.389	0.852
24	15.944	0.890	16.450	0.835	16.650	0.826	16.688	0.780

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 6: Tests de Jarque-Bera de normalité des résidus des modèles ARIMA sur l'IPI de l'automobile

	Statistique	p-valeur
ARIMA(0,1,0)	2.381	0.304
ARIMA(1,1,0)	2.414	0.299
ARIMA(0,1,1)	2.363	0.307
ARIMA(1,1,1)	2.241	0.326

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Le test de Jarque-Bera suppose que les résidus soient indépendants et homoscédastiques

Table 7: Estimation des coefficients associés aux modèles ARIMA sur l'IPI de l'automobile

	AR(1)				MA(1)			
	Coefficient	Écart-type	p-valeur		Coefficient	Écart-type	p-valeur	
ARIMA(0,1,0)								
ARIMA(1,1,0)	-0.280	0.088	0.001	**				
ARIMA(0,1,1)					-0.377	0.091	0.000	***
ARIMA(1,1,1)	0.165	0.214	0.439		-0.515	0.183	0.005	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1