

## Projet séries temporelles

# L'indice des prix à la consommation des produits d'entretien, hygiène beauté, dans la grande distribution en France métropolitaine

Réalisé par :

- Kamal *Samnoui*

Enseignant :

- *Prof.* Laurence *Broze*

ANNÉE UNIVERSITAIRE :

2020/2021

28 DÉCEMBRE 2020

---

# Table des matières

<b>Introduction</b>	<b>3</b>
<b>Méthode de décomposition</b>	<b>5</b>
1 Tableaux de Buys-Ballot . . . . .	6
2 Détermination de la tendance . . . . .	7
2.1 Moyenne mobile . . . . .	7
3 Estimation des variations saisonnières . . . . .	9
<b>Méthode de lissage</b>	<b>11</b>
1 Méthode de Holt . . . . .	11
2 Méthode de Holt-Winters additive . . . . .	13
3 Méthode de Holt-Winters multiplicative . . . . .	14
4 Comparaison des trois méthodes . . . . .	15
<b>Méthode de Box-Jenkins</b>	<b>17</b>
1 Analyse de la série . . . . .	17
2 Estimation du modèle . . . . .	21
3 Validation du modèle . . . . .	26
4 Prévision . . . . .	27
<b>Conclusion</b>	<b>29</b>
<b>Sommaire des figures</b>	<b>29</b>
<b>Sommaire des tableaux</b>	<b>30</b>

---

---

# *Introduction*

Une série temporelle, ou une série chronologique, est une suite de valeurs numériques appelées des observations, représentant l'évolution d'une quantité spécifique au cours du temps. De telles suites d'observations peuvent être exprimées mathématiquement afin d'en analyser le comportement, généralement pour comprendre son évolution passée et pour en prévoir le comportement futur. Les séries temporelles sont présentes dans de nombreux domaines d'application : en démographie, en économie, en médecine...

Dans ce rapport, nous nous intéressons à l'indice des prix à la consommation des produits d'entretien, d'hygiène et de beauté dans la grande distribution en France métropolitaine. Cet indice est l'instrument de mesure de l'inflation. Il permet d'estimer, entre deux périodes données, la variation moyenne des prix des produits consommés. Il s'agit d'une inflation quand l'indice est supérieur à 100, et de déflation quand l'indice est inférieur à 100. Dans ce rapport, (IPC) fait référence à l'indice des prix à la consommation.

Les données de la série que nous allons étudier dans la suite ont été mesurées ou relevées entre Janvier 2005 et Octobre 2020, nous disposons donc de 190 mesures, et elles ont été obtenues à partir du site de l'INSEE<sup>1</sup>. Nous représentons ci dessous l'évolution de l'indice des prix à la consommation de l'année 2005 à l'année 2020 :

---

1. INSEE : Institut National de la Statistique et des Études Économiques  
<https://www.insee.fr/fr/statistiques/serie/001768736>

---



FIGURE 1 – Évolution de l'IPC des produits d'entretien

Les données stockées ne se répartissent pas de manière égale entre les différents mois, et elles varient entre 96.1 en Novembre 2005 à 103.92 en Août 2012. Par ailleurs, nous constatons qu'il n'y a pas de saisonnalité dans l'évolution de l'IPC, la série croît globalement jusqu'en Mars 2010, suivie d'une faible décroissance jusqu'en Mars 2011, puis elle croît jusqu'en Août 2012, suivie d'une légère décroissance jusqu'en Octobre 2020. Cette augmentation de l'IPC qui dépasse 102 entre Juin 2011 et Avril 2014 est due à la crise économique mondiale de l'été 2011, donc la perte du pouvoir d'achat de la monnaie, ce qui explique l'augmentation des prix des produits d'entretien dans cette période. À partir d'Août 2015 on remarque un gain du pouvoir d'achat de la monnaie qui se traduit par une baisse de l'IPC (moins de 100). nous pouvons également remarquer que la série n'est pas stationnaire.

---

## *Méthode de décomposition*

Dans cette partie, nous étudions deux étapes. En première étape, on va s'intéresser aux différentes méthodes de décomposition, dans le but de déterminer une tendance globale qui donne une idée du comportement global de la série et de son évolution durable (par exemple globalement croissante ou décroissante), ensuite désaisonnaliser la série pour permettre d'évaluer la direction de la tendance, grâce à des comparaisons de mois en mois. Dans la deuxième étape, la série sera découpée en un échantillon d'estimation et un échantillon de prévision (comportant les 12 dernières observations), pour faire des prévisions.

Notons  $(X_t)_{t \in \mathcal{T}}$  la série temporelle où  $\mathcal{T} = \mathbb{N}$  ou  $\mathbb{Z}$ , correspondant aux valeurs de l'indice des prix à la consommation.

Si on trace deux courbes joignant les maxima (minima) distants de  $s$  périodes. Les deux courbes ne seront pas parallèles, on a le graphique suivant.



FIGURE 2 – Coubres joignant les maxima et minima de la série temporelle

---

Donc un modèle multiplicatif est plus adapté. Notre série chronologique se décompose de la façon suivante :

$$X_t = T_t \times I_t \times \epsilon_t$$

où

$T_t$  : La tendance déterministe (composante fondamentale).

$I_t$  : La composante saisonnière déterministe.

$\epsilon_t$  : La composante résiduelle aléatoire.

Un passage au logarithme du modèle multiplicatif le ramène à un modèle additif :

$$\log X_t = \log T_t + \log S_t + \log \epsilon_t$$

## 1 Tableaux de Buys-Ballot

il s'agit d'un tableau donnant deux dimensions au temps : une dimension annuelle et une dimension mensuelle. Pour les séries présentant à la fois tendance et mouvement saisonnier. Il va permettre d'analyser plus précisément notre série. voici ci-dessous le tableau de Buys-Ballot de notre série.

X(t)	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	moyenne
Janvier	96,6	96,52	97,42	98,73	100,95	101,81	101,31	103,85	103,79	103	100,4	99,28	98,77	98,59	98,63	97,47	99,82
Février	96,6	96,84	97,49	99,02	101,09	101,93	101,26	103,76	103,76	102,89	100,39	99,28	98,79	98,69	98,58	97,74	99,881875
Mars	96,7	97,36	97,59	99,31	101,22	102	101,45	103,7	103,65	102,66	100,26	99,13	98,77	98,7	98,45	97,49	99,9025
Avril	96,5	97,51	97,59	99,63	100,96	101,91	101,45	103,69	103,69	102,2	100,29	99,03	98,59	98,65	98,47	97,66	99,86375
Mai	96,4	97,38	97,64	99,56	100,9	101,81	101,9	103,42	103,4	101,58	100,09	98,83	98,44	98,63	98,27	97,32	99,723125
Juin	96,4	97,37	97,76	99,6	100,58	101,85	102,33	103,62	103,4	101,47	99,81	98,89	98,72	98,32	98,37	97,02	99,719375
Juillet	96,4	97,37	97,68	100,04	100,86	101,69	102,75	103,63	103,2	101,41	99,85	98,77	98,77	98,19	98,39	97,47	99,779375
Août	96,5	97,54	98,05	100,31	101,34	101,73	103,12	103,92	103,51	101,57	100	99	98,94	98,41	98,49	97,84	100,016875
Septembre	96,4	97,51	97,93	100,24	101,4	101,54	102,99	103,71	103,22	101,27	99,85	98,84	98,59	98,11	98,31	96,78	99,793125
Octobre	96,2	97,5	98,01	100,34	101,55	101,48	103,31	103,73	103,08	101,01	99,77	98,94	98,78	98,18	98,24	96,96	99,8175
Novembre	96,1	97,69	98,16	100,71	101,75	101,37	103,77	103,86	103,17	100,84	99,64	98,84	98,63	98,26	98,03		100,0546667
Décembre	96,3	97,71	98,26	100,7	101,57	101,44	103,86	103,84	103,19	100,45	99,66	99,02	98,78	98,66	98,35		100,1193333
moyenne	96,425	97,3583	97,798	99,8492	101,18	101,713	102,458	103,728	103,4217	101,696	100,001	98,9875	98,714	98,4492	98,38167	97,375	

TABLE 1 – Tableau de Buys-Ballot contient les valeurs  $X_t$

Le tableau Buys-Ballot, nous indique que la moyenne des valeurs correspondant à l'année 2012 est la plus élevée sur la période de 2005 à 2020 par rapport aux autres années. C'est l'année qui suit la crise économique de 2011. En revanche, la moyenne des valeurs la plus basse a été enregistrée à l'année 2005.

Dans notre tableau, nous ne disposons pas des données de Septembre et Décembre 2020. En calculons la moyenne d'indice des prix à la consommation de ces deux mois entre la période 2005 et 2019 et les autres mois entre 2005 et 2020. La moyenne atteint son maximum au mois de Décembre. À l'inverse c'est au mois de Juin que la moyenne est la plus faible.

Nous représentons maintenant le graphique associé au tableau de Buys-Ballot, où chaque année donne lieu à une courbe

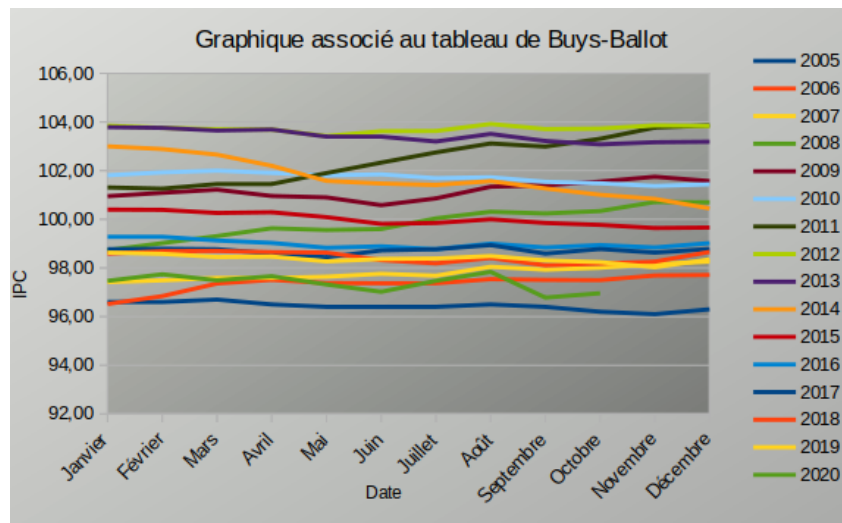


FIGURE 3 – Graphique associé au tableau de Buys-Ballot

D'après ce graphique on peut confirmer que l'année 2012 a enregistré les valeurs les plus élevées au cours la période d'étude entre Juin et Décembre. D'autre part l'année 2005 a enregistré les valeurs les plus faibles. On remarque également des valeurs importantes chaque mois d'Août et Décembre de chaque année.

## 2 Détermination de la tendance

### 2.1 Moyenne mobile

Dans ce rapport, nous utiliserons une moyenne mobile pour déterminer la tendance globale. Une moyenne mobile permet de lisser une série chronologique. Elle permet d'éliminer les fluctuations les moins significatives. On calcule des moyennes mobiles d'ordre 1, d'ordre 2, d'ordre 3, etc. L'ordre est le nombre de périodes (années, trimestres, mois...) sur lesquelles la moyenne mobile est calculée. Dans notre cas, les données sont mensuelles, donc nous considérons la moyenne mobile d'ordre 12 :

$$T_t = \frac{1}{24}T_{t-6} + \frac{1}{12}(X_{t-5} + \dots + X_{t+5}) + \frac{1}{24}T_{t+6}$$

Notons que nous n'aurions pas les valeurs de la tendance pour les premières et les dernières dates.

Voici le tableau de Buys-Ballot qui contient les valeurs des tendances moyenne mobile d'ordre 12.

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	moyenne
Janvier		96,780	97,580	98,760	100,704	101,683	101,623	103,523	103,680	102,690	100,584	99,389	98,791	98,648	98,390	97,838	100,04431
Février		96,864	97,615	98,953	100,781	101,734	101,725	103,593	103,645	102,534	100,454	99,303	98,788	98,602	98,402	97,772	100,051
Mars		96,954	97,653	99,143	100,873	101,756	101,844	103,657	103,608	102,372	100,329	99,219	98,775	98,560	98,413	97,681	100,05578
Avril		97,054	97,692	99,336	100,971	101,759	101,980	103,704	103,560	102,205	100,218	99,142	98,758	98,515	98,424	97,564	100,05894
Mai		97,175	97,733	99,540	101,065	101,740	102,157	103,725	103,505	102,021	100,117	99,074	98,743	98,475	98,417		100,24896
Juin		97,300	97,775	99,748	101,145	101,719	102,358	103,728	103,449	101,810	100,034	99,014	98,724	98,454	98,395		100,2608
Juillet	96,422	97,396	97,853	99,942	101,217	101,693	102,564	103,725	103,389	101,588	99,954	98,966	98,707	98,451	98,333		100,01319
Août	96,428	97,460	97,971	100,120	101,288	101,644	102,774	103,723	103,320	101,375	99,861	98,925	98,695	98,448	98,250		100,01878
Septembre	96,466	97,497	98,107	100,286	101,355	101,593	102,972	103,720	103,242	101,171	99,768	98,889	98,688	98,433	98,175		100,02414
Octobre	96,535	97,510	98,263	100,421	101,427	101,551	103,159	103,718	103,139	100,991	99,668	98,856	98,688	98,415	98,101		100,02956
Novembre	96,618	97,524	98,428	100,533	101,505	101,535	103,316	103,718	103,001	100,850	99,563	98,821	98,698	98,393	98,028		100,03533
Décembre	96,700	97,551	98,585	100,629	101,595	101,559	103,433	103,708	102,845	100,718	99,473	98,798	98,689	98,380	97,932		<b>100,03961</b>
moyenne	96,528	97,255	97,938	99,784	101,160	101,664	102,492	<b>103,687</b>	103,365	101,694	100,002	99,033	98,729	98,481	98,272	97,7138	

TABLE 2 – Tableau de Buys-Ballot contient les valeurs  $T_t$

Ce tableau nous indique que la moyenne des valeurs correspondant à l'année 2012 est la plus élevée par rapport aux autres années. Nous avons également le mois Décembre le mois où les valeurs les plus élevées ont été enregistrées.

Nous représentons ci-dessous, la courbe de la tendance globale et la courbe de la série temporelle.

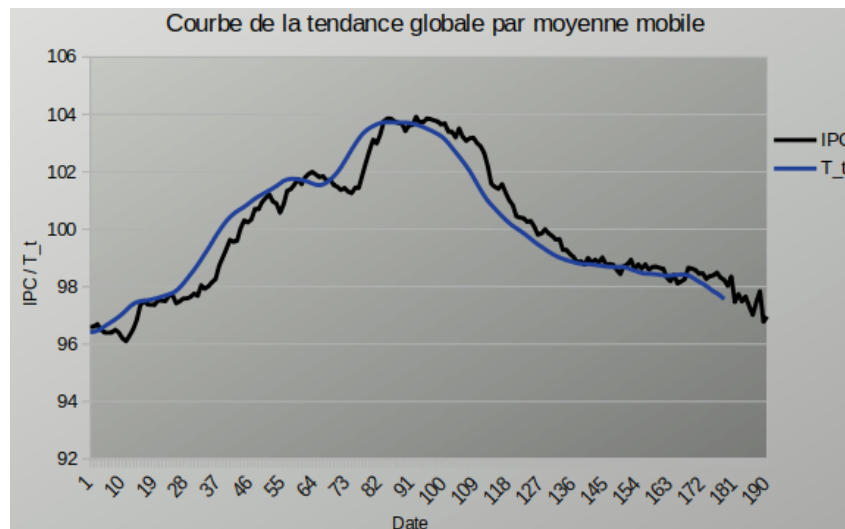


FIGURE 4 – Courbe de la tendance globale par moyenne mobile

Nous remarquons que cette méthode de lissage agit comme un filtre passe bas car elle élimine le bruit. À l'évidence, la tendance n'est pas linéaire, de Janvier 2005 au Juin 2012, l'indice des prix à la consommation a tendance à augmenter progressivement, tandis-



que vers la fin de Juin de l'année 2012, l'indice des prix à la consommation à diminuer lentement.

### 3 Estimation des variations saisonnières

Après avoir déterminé la tendance par la méthode des moyennes mobiles, nous allons nous intéresser à la composante saisonnière ou saisonnalité  $I_t$  pour  $t \in \mathcal{T}$ , correspondant à un phénomène qui se répète à un intervalle de temps périodique, d'où le terme de variations saisonnières.

Pour estimer les variations saisonnières, sachant que le modèle est supposé multiplicatif, nous commençons par calculer la moyenne des données sans tendance pour chaque mois, pour cela on divise la série  $X_t$  par la tendance globale déterminée par la méthode des moyennes mobiles précédemment  $T_t$  de la série  $X_t$ . Ensuite nous reportons ces valeurs sur le tableau de Buys-Ballot, qui servira à calculer les coefficients saisonniers. Ainsi, on calcule les moyennes des  $X_t/T_t$  notées par  $M_t$  avec  $t = 1, \dots, 12$  pour chaque mois et on fait la moyenne finale des  $M_t$  que nous noterons  $M$ . Enfin on corrige les moyennes  $M_t$  obtenus pour qu'ils aient une moyenne égale à 1 (Principe de conservation des aires) on calculons  $M_t/M$ . Le résultat ainsi obtenu correspond aux coefficients des variations saisonnières  $I_t$ . Les coefficients  $I_t$  s'appliquent de manière égale tous les ans. On obtient :

$$\left\{ \begin{array}{l} I_{Janvier} = I_1 = I_{13} = \dots = 0,9999 \\ I_{Fevrier} = I_2 = I_{14} = \dots = 1,0005 \\ I_{Mars} = I_3 = I_{15} = \dots = 1,0006 \\ I_{Avril} = I_4 = I_{16} = \dots = 1,0003 \\ I_{Mai} = I_5 = I_{17} = \dots = 0,9988 \\ I_{Juin} = I_6 = I_{18} = \dots = 0,9989 \\ I_{Juillet} = I_7 = I_{19} = \dots = 0,9992 \\ I_{Aot} = I_8 = I_{20} = \dots = 1,0014 \\ I_{Septembre} = I_9 = I_{21} = \dots = 0,9997 \\ I_{Octobre} = I_{10} = I_{22} = \dots = 0,9998 \\ I_{Novembre} = I_{11} = I_{23} = \dots = 1,0002 \\ I_{Decembre} = I_{12} = I_{24} = \dots = 1,0008 \end{array} \right.$$

Pour obtenir la série corrigée des variations saisonnières (CVS), on divise la série temporelle  $X_t$  par la série des variations saisonnières  $I_t$ .

Tous ces calculs sont faits en utilisant le tableau de Buys-Ballot. Voici ce Tableau de la série CVS ci-dessous.

X_t / I_t	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Moyenne
Janvier	96,61	96,53	97,43	98,74	100,96	101,82	101,32	103,86	103,80	103,01	100,41	99,29	98,78	98,60	98,64	97,48	99,832691
Février	96,55	96,79	97,44	98,97	101,04	101,88	101,21	103,71	103,71	102,84	100,34	99,23	98,74	98,64	98,53	97,69	99,833774
Mars	96,64	97,30	97,53	99,25	101,16	101,94	101,39	103,64	103,59	102,60	100,20	99,07	98,71	98,64	98,39	97,43	99,842758
Avril	96,47	97,48	97,56	99,60	100,93	101,88	101,42	103,66	103,66	102,17	100,26	99,00	98,56	98,62	98,44	97,63	99,833879
Mai	96,51	97,49	97,75	99,67	101,02	101,93	102,02	103,54	103,52	101,70	100,21	98,94	98,55	98,74	98,38	97,43	99,838239
Juin	96,51	97,48	97,87	99,71	100,69	101,96	102,44	103,74	103,51	101,58	99,92	99,00	98,83	98,43	98,48	97,13	99,830131
Juillet	96,48	97,45	97,76	100,12	100,94	101,77	102,83	103,71	103,28	101,49	99,93	98,85	98,85	98,27	98,47	97,55	99,859381
Août	96,36	97,40	97,91	100,17	101,20	101,59	102,97	103,77	103,36	101,43	99,86	98,86	98,80	98,27	98,35	97,70	99,875128
Septembre	96,43	97,54	97,96	100,27	101,43	101,57	103,02	103,74	103,25	101,30	99,88	98,87	98,62	98,14	98,34	96,81	99,824295
Octobre	96,22	97,52	98,03	100,36	101,57	101,50	103,33	103,75	103,10	101,03	99,79	98,96	98,80	98,20	98,26	96,98	99,840708
Novembre	96,08	97,67	98,14	100,69	101,73	101,35	103,75	103,84	103,15	100,82	99,62	98,82	98,61	98,24	98,01		100,0389
Décembre	96,23	97,63	98,18	100,62	101,49	101,36	103,78	103,76	103,11	100,37	99,58	98,94	98,70	98,58	98,27		<b>100,04138</b>
Moyenne	96,425	97,3584	97,79833	99,8491	101,18	101,713	102,458	<b>103,7275</b>	103,422	101,696	100,001	98,9875	98,714	98,4492	98,38169	97,38402	

TABLE 3 – Tableau de Buys-Ballot pour la série CVS

Nous remarquons toujours que la moyenne des valeurs correspondant à l'année 2012 est la plus élevée par rapport aux autres années, et le mois Décembre le mois où les valeurs les plus élevées ont été enregistrées.

Nous représentons ci-dessous, la courbe de la série corrigée des variations saisonnières.

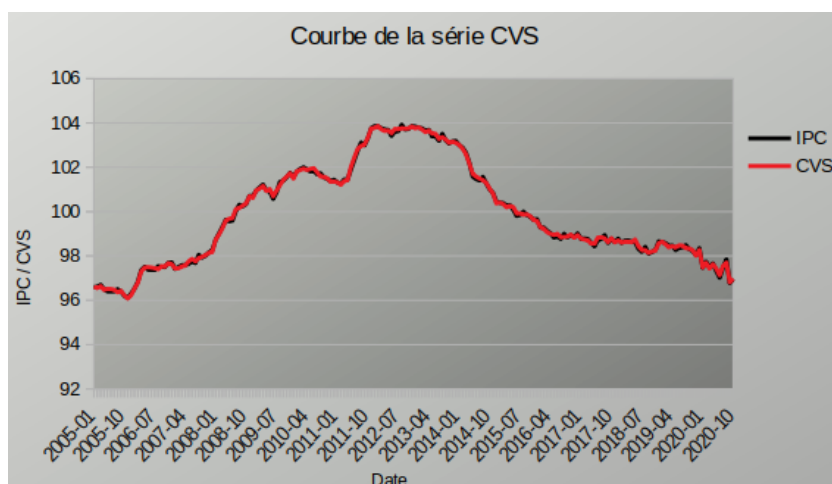


FIGURE 5 – Série corrigée des variations saisonnières

À l'évidence les deux courbes sont presque superposées. Les variations saisonnières n'ont pas un grand effet sur la série.

---

# *Méthode de lissage*

Afin de prévoir les valeurs futures de la série, c'est à dire prédire l'indice des prix à la consommation aux prochaines dates, on utilise l'estimation de la tendance et celle de la composante saisonnière. Plus précisément, si on souhaite prévoir une valeur de la série à l'instant  $T + h$ , où  $h \geq 0$  est l'horizon, on utilise les estimations de la tendance et de la saisonnalité.

Dans cette partie, on va s'intéresser aux méthodes de lissage exponentiel qui consistent à donner aux observations passées un poids décroissant exponentiellement avec leur ancienneté. Une méthode de lissage exponentiel est une méthode empirique de lissage et de prévision de données chronologiques affectées d'aléas. Nous nous intéressons particulièrement à la méthode de Holt, la méthode de Holt-Winters additive et la méthode de Holt Winters multiplicative

Nous commençons par découper la série en un échantillon d'estimation et en un échantillon de prévision (comportant les 12 dernières observations). L'échantillon d'apprentissage(d'estimation) va nous permettre d'entraîner notre modèle, c'est à dire notre étude sera faite sur cet échantillon afin de choisir une meilleure méthode de lissage. En revanche, l'échantillon de test(de prévision) nous permettra à son tour de tester et valider notre modèle, c'est à dire de comparer les résultats des prévisions avec les observations de cet échantillon.

## 1 Méthode de Holt

Sur l'échantillon d'estimation, on applique la méthode de Holt qui permet de construire des prévisions linéaires. La prévision pour un horizon  $h$  avec la méthode de Holt est donnée par la formule suivante :

$$\hat{X}_t(h) = S_t + hT_t, \quad h = 1, 2, \dots$$

---

Avec :

$\hat{X}_t(h)$  est la valeur de la prévision à l'horizon  $h$ .

$S_t$  est la composante de niveau.

$T_t$  est la composante de tendance.

Cette méthode dépend de deux paramètres compris entre 0 et 1 :  $\alpha$  et  $\gamma$  qui sont appelés Les constantes de lissage. Au temps  $t$ , on estime deux composantes qui sont combinées pour fournir des prévisions, Nous avons :

$$\begin{cases} S_t = \alpha X_t + (1 - \alpha)(S_{t-1} + T_{t-1}) \\ T_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)T_{t-1} \\ \hat{X}_t(h) = S_t + hT_t \quad h = 1, 2, \dots \end{cases}$$

L'initialisation des relations est faite en choisissant  $S_0 = X_1$  ou  $S_0 = \bar{X}$  ainsi que  $T_0 = 0$ . On commence par fixer  $\alpha = 0.85$  et  $\gamma = 0.3$ . Ces constantes seront ensuite optimisées grâce au solveur sur Excel. Pour l'initialisation de la méthode, on considérant que  $S = X_1$  et  $T = 0$ , nous sommes parvenus à faire des prévisions sur l'indice des prix à la consommation pour l'échantillon d'estimation. Les résultats obtenus ci-dessous représentent les valeurs prédites et les valeurs observées :

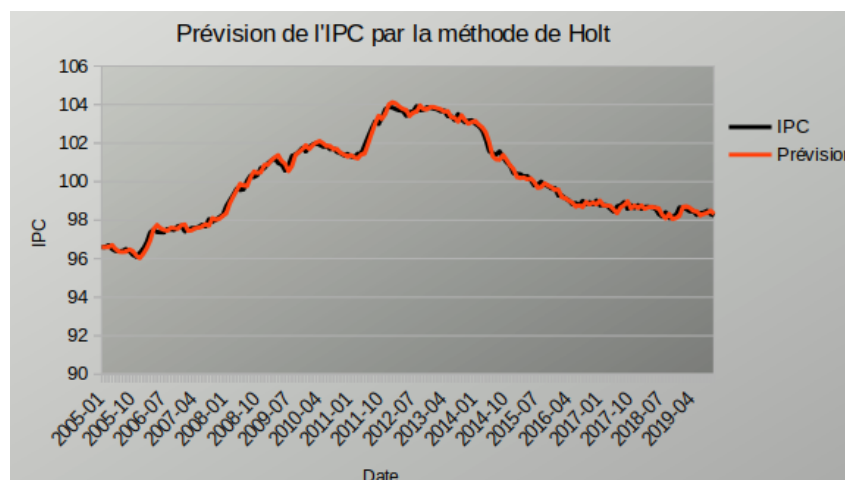


FIGURE 6 – Valeurs prédites et observées par la méthode de Holt

On remarque que les valeurs prédites sont plus proches des valeurs observées au début et à la fin de la période d'étude. L'erreur moyenne absolue en pourcentage (MAPE) obtenue avec ces prévisions est égale à 0,16%

Les valeurs optimales pour les constantes obtenues à l'aide du solveur pour minimiser cette erreur sont :  $\alpha = 0.98$  et  $\gamma = 0.07$ . La nouvelle MAPE correspondante est 0.15%.

## 2 Méthode de Holt-Winters additive

Sur l'échantillon d'estimation, on applique la méthode de Holt-Winters additive, l'idée de cette méthode est d'introduire des coefficients saisonniers. La prévision est donnée par la formule suivante :

$$\hat{X}_t(h) = S_t + hT_t + I_{t+h-s}, \quad h = 1, 2, \dots$$

Avec :

$S_t$  est la composante de niveau

$T_t$  est la composante de tendance

$I_t$  est un coefficient saisonnier et  $s$  est la saison.

La prévision dépend de 3 paramètres compris entre 0 et 1  $\alpha$ ,  $\gamma$  et  $\delta$  tels que :

$$\begin{cases} S_t = \alpha(X_t - I_{t-s}) + (1 - \alpha)(S_{t-1} + T_{t-1}) \\ T_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)T_{t-1} \\ I_t = \delta(X_t - S_t) + (1 - \delta)I_{t-s} \end{cases}$$

Comme précédemment, on commence par fixer  $\alpha = 0.3$ ,  $\gamma = 0.1$  et  $\delta = 0.9$ . Ces constantes seront ensuite optimisées grâce au solveur sur Excel.

Pour l'initialisation de la méthode, pour  $S$  on prend la moyenne des  $s$  ( $s=12$ ) premières observations,  $T = 0$  et pour  $I$  on prend  $X - S$ .

Les résultats obtenus ci-dessous représentent les valeurs prédites et les valeurs observées :

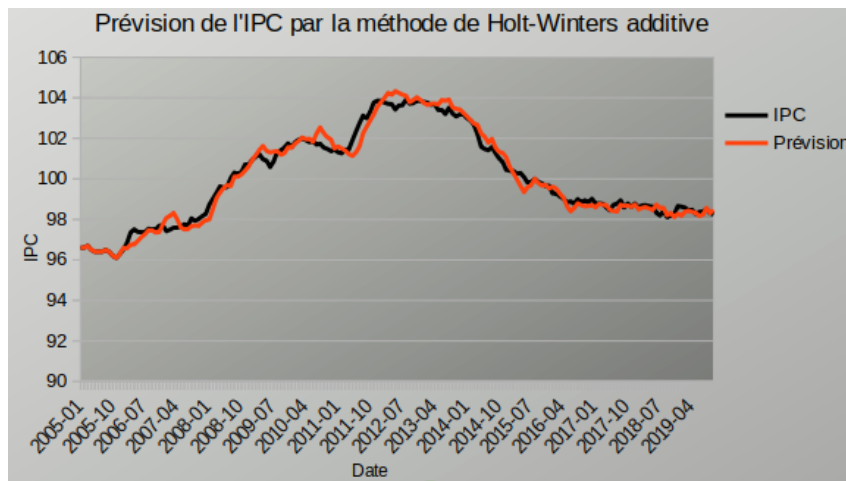


FIGURE 7 – Valeurs prédites et observées par la méthode de Holt-Winters additive

On remarque que les valeurs prédites sont plus proches des valeurs observées au début et à la fin de la période d'étude. L'erreur moyenne absolue en pourcentage (MAPE) obtenue

avec ces prévisions est égale à 0,26%

Les valeurs optimales pour les constantes obtenues à l'aide du solveur pour minimiser cette erreur sont :  $\alpha = 0.88$ ,  $\gamma = 0.05$  et  $\delta = 0.99$ . La nouvelle MAPE correspondante est 0.16%.

### 3 Méthode de Holt-Winters multiplicative

Sur l'échantillon d'estimation, on applique la méthode de Holt-Winters multiplicative. On va introduire les coefficients saisonniers comme dans le cas du modèle de Holt-Winters additive mais cette fois-ci en multipliant. La prévision à l'horizon  $h$  est donnée par la formule suivante :

$$\hat{X}_t(h) = (S_t + hT_t)I_{t+h-s}, \quad h = 1, 2, \dots$$

Avec :

$S_t$  est la composante de niveau

$T_t$  est la composante de tendance

$I_t$  est un coefficient saisonnier et  $s$  est la saison.

La prévision dépend de 3 paramètres compris entre 0 et 1  $\alpha$ ,  $\gamma$  et  $\delta$  tels que :

$$\begin{cases} S_t = \alpha(X_t/I_{t-s}) + (1 - \alpha)(S_{t-1} + T_{t-1}) \\ T_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)T_{t-1} \\ I_t = \delta(X_t/S_t) + (1 - \delta)I_{t-s} \end{cases}$$

Comme précédemment, on commence par fixer  $\alpha = 0.3$ ,  $\gamma = 0.1$  et  $\delta = 0.9$ . Ces constantes seront ensuite optimisées grâce au solveur sur Excel.

Pour l'initialisation de la méthode, pour  $S$  on prend la moyenne des  $s(s=12)$  premières observations,  $T = 0$  et pour  $I$  on prend  $X/S$ .

Les résultats obtenus ci-dessous représentent les valeurs prédites et les valeurs observées :

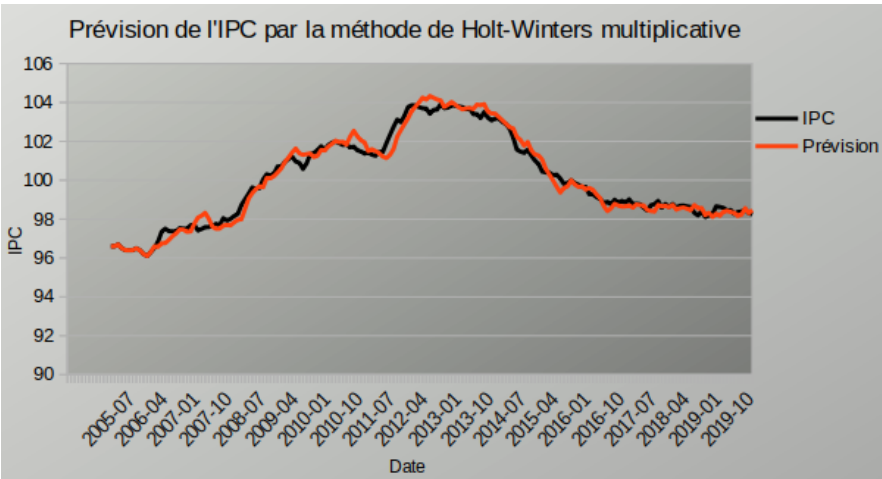


FIGURE 8 – Valeurs prédites et observées par la méthode de Holt-Winters multiplicative

L'erreur moyenne absolue(MAPE) obtenue pour cette méthode avant l'optimisation des paramètres est 0.26%. Afin de rendre notre modèle plus performant, nous avons ensuite cherché les valeurs optimales des constantes  $\alpha$ ,  $\gamma$  et  $\delta$  pour minimiser cette erreur. Les valeurs optimales ainsi que la MAPE trouvées à l'aide du solver sont :  $\alpha = 0.89$ ,  $\gamma = 0.05$  et  $\delta = 0.99$ . La nouvelle MAPE correspondante est 0.15%.

## 4 Comparaison des trois méthodes

Après avoir étudié ces différentes méthodes, nous comparons maintenant les différents résultats afin de choisir la meilleure méthode. On se base sur l'erreur MAPE de chacune de ces approches pour choisir la méthode ayant commis moins d'erreur. Les différents résultats obtenus sont résumés dans le tableau suivant :

Méthode	MAPE
Holt	0,15 %
Holt-Winters multiplicative	0,15 %
Holt-Winters additive	0,16 %

TABLE 4 – Valeur de MAPE pour différentes méthodes de lissage

D'après les résultats du tableau, on propose de choisir la méthode de Holt-Winters multiplicative comme étant la meilleure méthode. Dans la suite, nous utiliserons donc cette méthode pour la prévision des 12 dernières observations.

Après avoir appliqué cette méthode aux 12 dernières données, nous avons obtenu les résultats suivants :

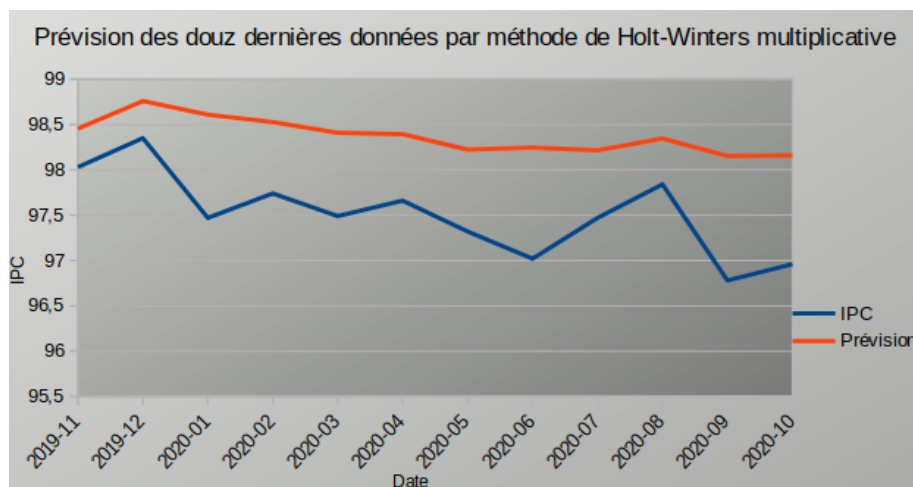


FIGURE 9 – La prévision des douze dernières données

L'erreur moyenne en pourcentage obtenue avec ces 12 dernières observations avec la méthode de Holt-Winters est de : 0,9%. On remarque que la courbe des données prédites et la courbe des données observées, elles ont le même comportement. Nous allons ainsi afficher ci-dessous le graphique sur l'ensemble de l'échantillon comportant les observations et les valeurs prédites.

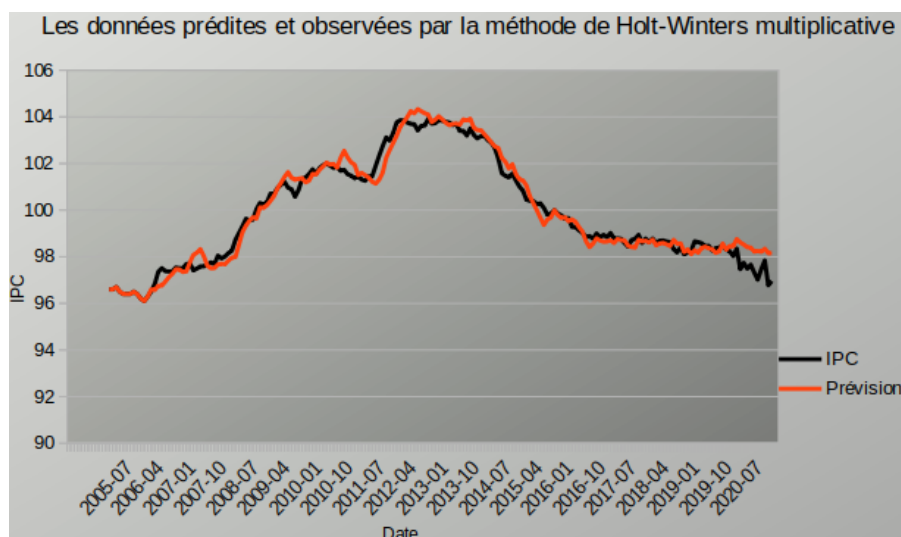


FIGURE 10 – Ensemble de valeurs prédites et observées par Holt-Winters multiplicative



---

# *Méthode de Box-Jenkins*

La mise en oeuvre des résultats théoriques concernant les modèles ARMA a donné lieu à une méthodologie en plusieurs étapes, connues sous le nom de méthode de Box et Jenkins. dans cette partie, on s'intéresse à cette méthode, dont l'approche est différente des méthodes de lissage vues dans la partie précédente. En effet, des modifications sur la série brute seront nécessaires, la notion de stationnarité intervient. Un processus  $(X_t)$  est dit stationnaire si les conditions suivantes sont vérifiées :

- L'espérance est constante au cours du temps :  $E(X_t) = m$  pour tout  $t$ .
- La variance est constante au cours du temps :  $V(X_t) = \sigma^2$  pour tout  $t$ .
- Les covariances entre variables sont stables au cours du temps, c'est à dire ne dépendent que du délai  $k$  :

$$Cov(X_1, X_2) = Cov(X_2, X_3) = \dots = Cov(X_{t+k-1}, X_{t+k})$$

La méthodologie de Box-Jenkins peut se décomposer en trois étapes :

- L'estimation
- La validation
- La prévision

Nous détaillerons chacune de ces étapes dans les sections suivantes.

Dans toute cette partie, nous utiliserons le logiciel Gretl.

## **1 Analyse de la série**

Comme nous l'avons abordé sur la figure 1 dans la première partie, la série n'est pas stationnaire. Pour approfondir cette analyse de non stationnarité, on peut également afficher le corrélogramme de la série.

---

Fonction d'auto-corrélation pour IPC  
 \*\*\*, \*\*, \* indicate significance at the 1%, 5%, 10% levels  
 using standard error  $1/T^{0,5}$

RETARD	ACF		PACF		Q [p. crit.]
1	0,9847	***	0,9847	***	187,1440 [0,000]
2	0,9685	***	-0,0369		369,1375 [0,000]
3	0,9540	***	0,0497		546,6729 [0,000]
4	0,9368	***	-0,1006		718,7840 [0,000]
5	0,9155	***	-0,1365	*	884,0423 [0,000]
6	0,8946	***	0,0001		1042,7064 [0,000]
7	0,8732	***	-0,0376		1194,6867 [0,000]
8	0,8513	***	-0,0007		1339,9451 [0,000]
9	0,8267	***	-0,0934		1477,6753 [0,000]
10	0,8014	***	-0,0309		1607,8384 [0,000]
11	0,7776	***	0,0320		1731,0670 [0,000]
12	0,7541	***	0,0054		1847,5999 [0,000]
13	0,7292	***	-0,0299		1957,1782 [0,000]
14	0,7058	***	0,0345		2060,4248 [0,000]
15	0,6837	***	0,0157		2157,8768 [0,000]
16	0,6622	***	0,0141		2249,8191 [0,000]
17	0,6382	***	-0,0916		2335,7039 [0,000]
18	0,6135	***	-0,0483		2415,5382 [0,000]
19	0,5886	***	-0,0491		2489,4470 [0,000]
20	0,5634	***	-0,0253		2557,5629 [0,000]
21	0,5368	***	-0,0374		2619,7724 [0,000]
22	0,5102	***	-0,0204		2676,2978 [0,000]

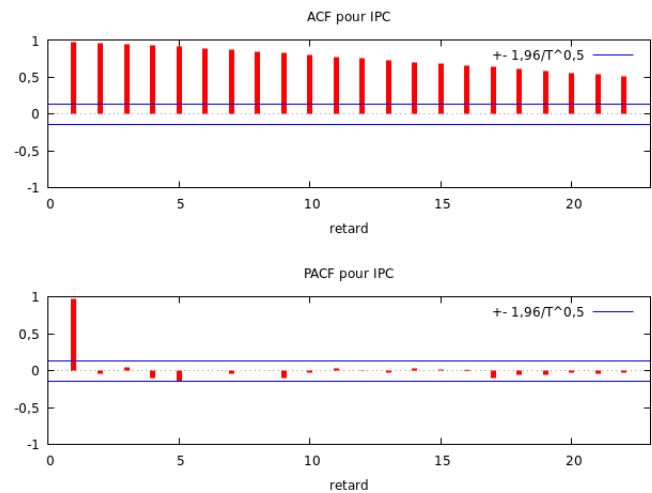


FIGURE 11 – Corrélogramme de la série brute

On remarque que l'autocorrélation est forte pour tous les décalages ( $k$ ) et qu'ils ne sont pas égaux, ce qui est caractéristique d'une série non stationnaire. Afin de rendre notre série stationnaire, nous allons tester d'autres méthodes. Nous pouvons d'abord voir les résultats du test de Dickey-Fuller (test de racine unitaire) qui considère l'hypothèse nulle ( $H_0$ ) comme étant la non stationnarité de la série. Les résultats du test de Dickey-Fuller sont :

```
Test de Dickey-Fuller augmenté pour IPC
testing down from 14 lags, criterion AIC
taille de l'échantillon 177
hypothèse nulle de racine unitaire : a = 1

test avec constante
avec 12 retards de (1-L)IPC
modèle: (1-L)y = b0 + (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -0,0115797
statistique de test: tau_c(1) = -1,52253
p. critique asymptotique 0,5224
Coeff. d'autocorrélation du 1er ordre pour e: 0,006
différences retardées: F(12, 163) = 5,186 [0,0000]

avec constante et tendance temporelle
avec 12 retards de (1-L)IPC
modèle: (1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -0,0131574
statistique de test: tau_ct(1) = -1,74495
p. critique asymptotique 0,7312
Coeff. d'autocorrélation du 1er ordre pour e: 0,014
différences retardées: F(12, 162) = 4,022 [0,0000]
```

TABLE 5 – Test de Dickey-Fuller

Les résultats du test avec constante et tendance temporelle indiquent une p-critique asymptotique égale à 0,7312 qui est supérieure à 0,05. Donc, on ne rejette donc pas l'hypothèse nulle, il y'a donc une racine unitaire. Une première différence est donc nécessaire.

La Figure suivante donne la série obtenue en appliquant une première différence.

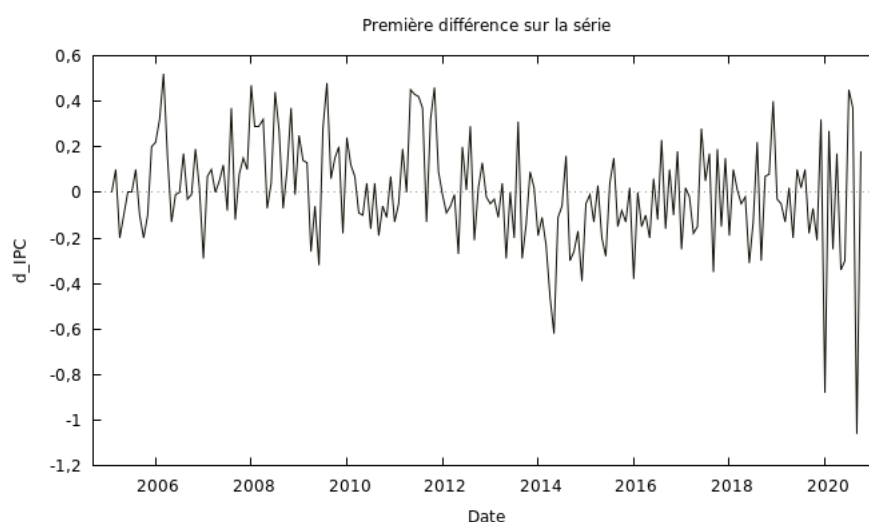


FIGURE 12 – Première différence sur la série brute

Après avoir appliqué une première différence, on peut remarquer que notre série paraît un peu stable au cours du temps. Pour avoir plus de détails sur cette stabilité, nous avons affiché ci-dessous les corrélogrammes de la série de la première différence :

Fonction d'auto-corrélation pour d\_IPC  
 \*\*\*, \*\*, \* indicate significance at the 1%, 5%, 10% levels  
 using standard error  $1/T^{0.5}$

RETARD	ACF	PACF	Q	[p. crit.]
1	0,0321	0,0321	0,1980	[0,656]
2	0,1770 **	0,1762 **	6,2472	[0,044]
3	0,0854	0,0776	7,6642	[0,053]
4	0,2276 ***	0,2008 ***	17,7737	[0,001]
5	0,0463	0,0166	18,1947	[0,003]
6	0,1516 **	0,0847	22,7299	[0,001]
7	0,0245	-0,0185	22,8492	[0,002]
8	0,3044 ***	0,2440 ***	41,3276	[0,000]
9	-0,0857	-0,1327 *	42,8008	[0,000]
10	0,0861	-0,0223	44,2959	[0,000]
11	-0,0543	-0,0902	44,8931	[0,000]
12	0,2640 ***	0,2015 ***	59,1048	[0,000]
13	-0,0200	-0,0037	59,1870	[0,000]
14	0,0953	0,0107	61,0621	[0,000]
15	-0,0310	-0,0455	61,2611	[0,000]
16	0,1917 ***	0,0742	68,9302	[0,000]
17	0,0286	0,0968	69,1019	[0,000]
18	0,0827	-0,0110	70,5442	[0,000]
19	0,0751	0,0999	71,7422	[0,000]
20	0,1592 **	0,0365	77,1551	[0,000]
21	0,0171	0,0519	77,2179	[0,000]
22	0,0611	-0,0497	78,0260	[0,000]

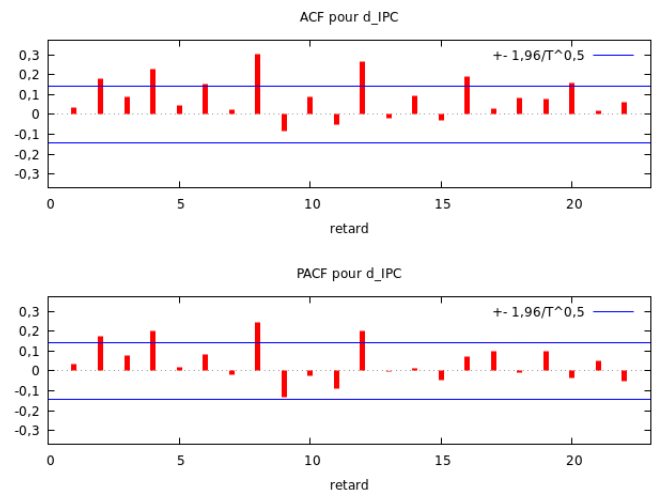


FIGURE 13 – Corrélogramme de la série de la première différence

On peut remarquer qu'il y'a une dépendance au niveau de l'autocorrélation et de l'autocorrélation partielle d'ordre 12, 8 et 4. Cela veut dire que la saisonnalité persiste encore et donc la série n'est pas encore stationnaire. Pour résoudre ce problème de saisonnalité, nous allons appliquer une différence saisonnière de cette première différence. La série issue de cette transformation est donnée par la suivante :

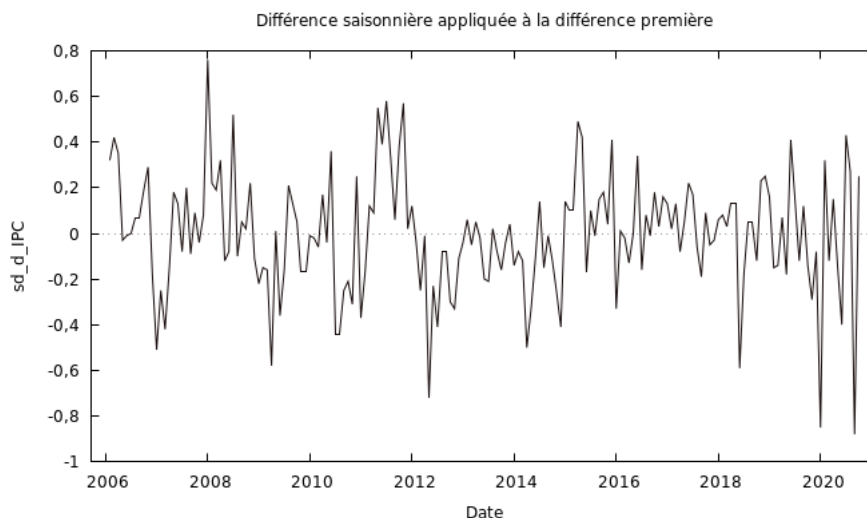


FIGURE 14 – Différence saisonnière appliquée à la première différence

La série obtenue à l'air stationnaire. Si on applique le test de Dickey-Fuller à la série, on trouve une p-critique asymptotique égale à 0.001311 qui est inférieure à 0.05. Donc la série

est stationnaire au sens faible. Pour mieux étudier cette stationnarité, nous allons afficher ci-dessous les corrélogrammes de la série obtenue après avoir appliqué cette différence saisonnière.

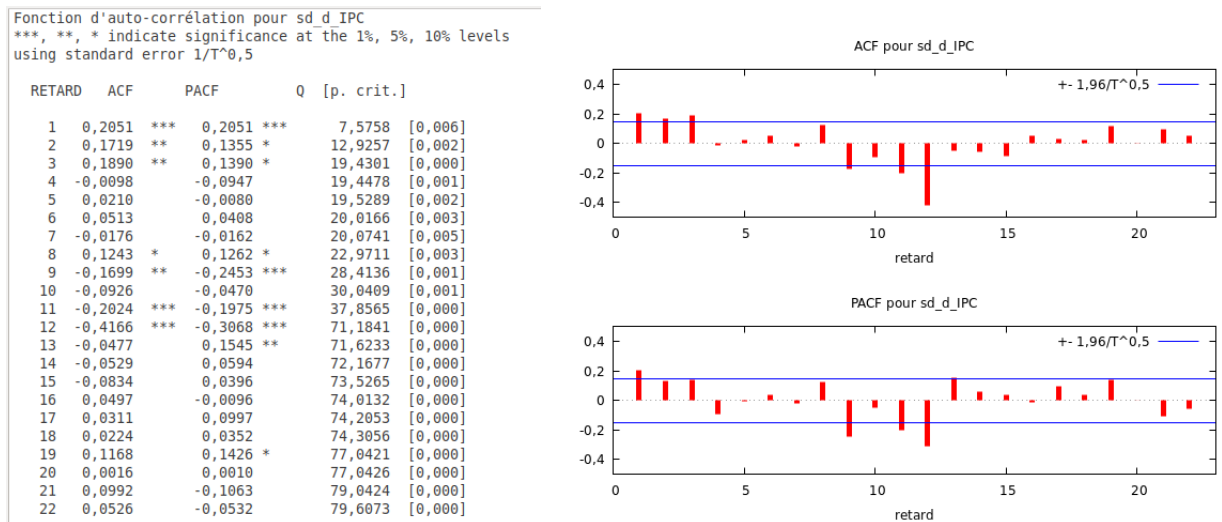


FIGURE 15 – Corrélogramme de la série de la différence saisonnière appliquée à la première différence

Ces résultats semblent satisfaisant, on peut néanmoins conclure que notre série est stationnaire et poursuivre notre étude.

## 2 Estimation du modèle

Après avoir transformé notre série de manière à ce qu'elle devienne stationnaire, on va déterminer les paramètres  $p$  et  $q$  du modèle ARMA à l'aide de la fonction d'autocorrélation et de la fonction d'autocorrélation partielle. Nous allons travailler sur l'échantillon d'estimation, pour déterminer le meilleur modèle ARMA.

On commence par appliquer un AR(1) à notre série. Nous avons obtenu les résultats suivants :

Évaluations de la fonction : 48					
Évaluations du gradient : 15					
Modèle 2: ARMA, utilisant les observations 2005:01-2019:10 (T = 178)					
Estimé à l'aide du filtre de Kalman (MV exacte)					
Variable dépendante: IPC					
Écart type basés sur la matrice hessienne					
	coefficient	erreur std.	z	p. critique	
const	98,1752	1,84820	53,12	0,0000	***
phi_1	0,995450	0,00420937	236,5	0,0000	***
Moy. var. dép.					
100,0312					
Éc. type var. dép.					
2,226065					
Moyenne des innovations					
0,008809					
Éc. type des innovations					
0,200912					
Log de vraisemblance					
30,74844					
Critère d'Akaike					
-55,49688					
Critère de Schwarz					
-45,95153					
Hannan-Quinn					
-51,62599					
	Réel	Imaginaire	Modulo	Fréquence	
AR					
Racine 1	1,0046	0,0000	1,0046	0,0000	

TABLE 6 – Modèle AR(1) appliqué à la série

On remarque que le coefficient estimé  $\varphi_1$  est significatif. On regarde maintenant le corrélogramme des résidus, on a les résultats suivants :

Fonction d'auto-corrélation résiduelle  
 \*\*\*, \*\*, \* indicate significance at the 1%, 5%, 10% levels  
 using standard error  $1/T^{0.5}$

RETARD	ACF	PACF	Q [p. crit.]
1	0,1563 **	0,1563 **	
2	0,1754 **	0,1547 **	10,0187 [0,002]
3	0,1179	0,0741	12,5634 [0,002]
4	0,1734 **	0,1290 *	18,0969 [0,000]
5	0,0713	0,0072	19,0378 [0,001]
6	0,1467 *	0,0930	23,0462 [0,000]
7	0,0743	0,0144	24,0799 [0,001]
8	0,1915 **	0,1358 *	30,9927 [0,000]
9	0,0176	-0,0586	31,0511 [0,000]
10	0,0722	0,0026	32,0453 [0,000]
11	-0,0733	-0,1261 *	33,0751 [0,000]
12	0,1923 **	0,1766 **	40,2150 [0,000]
13	-0,0321	-0,0785	40,4150 [0,000]
14	0,0028	-0,0489	40,4166 [0,000]
15	-0,0145	-0,0114	40,4578 [0,000]
16	0,1538 **	0,1305 *	45,1355 [0,000]
17	0,0839	0,0984	46,5379 [0,000]
18	0,0720	-0,0036	47,5764 [0,000]
19	0,0200	0,0079	47,6573 [0,000]
20	0,1407 *	0,0514	51,6695 [0,000]
21	0,0480	0,0426	52,1398 [0,000]
22	0,0466	-0,0417	52,5853 [0,000]

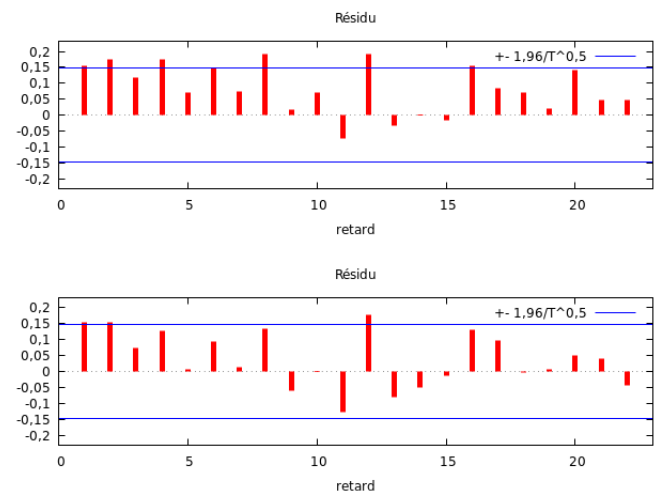


FIGURE 16 – Corrélogramme des résidus du modèle AR(1)

Les résultats du corrélogramme des résidus montrent des dépendances, le modèle estimé n'est donc pas satisfaisant. Dans un première temps, on peut ajouter des coefficients significatifs au modèle. Autrement dit, on va appliquer un AR en rajoutant les coefficients 2 et 12. On obtient les résultats suivants :

Évaluations de la fonction : 54					
Évaluations du gradient : 15					
Modèle 4: ARMA, utilisant les observations 2005:01-2019:10 (T = 178)					
Estimé à l'aide du filtre de Kalman (MV exacte)					
Variable dépendante: IPC					
Écarts type basés sur la matrice hessienne					
	coefficient	erreur std.	z	p. critique	
const	99,1513	1,36657	72,55	0,0000	***
phi_1	1,13774	0,0750240	15,16	6,03e-52	***
phi_2	-0,104042	0,0805373	-1,292	0,1964	
phi_12	-0,0429417	0,0151913	-2,827	0,0047	***
Moy. var. dép.		100,0312			
Ec. type var. dép.		2,226065			
Moyenne des innovations		-0,002361			
Ec. type des innovations		0,191920			
Log de vraisemblance		38,72810			
Critère d'Akaike		-67,45621			
Critère de Schwarz		-51,54729			
Hannan-Quinn		-61,00471			

TABLE 7 – Modèle AR(1) en rajoutant les coefficients 2 et 12

Après avoir ajouter ces deux coefficients dans notre modèle, nous pouvons remarquer que le coefficient  $\varphi_{12}$  est significatif pour le modèle. Pour le coefficient  $\varphi_2$ , on voit qu'il n'est pas significatif pour le modèle. On peut penser à l'appliquer sur la moyenne mobile (MA). Nous avons les résultats suivants :

Évaluations de la fonction : 63					
Évaluations du gradient : 16					
Modèle 5: ARMA, utilisant les observations 2005:01-2019:10 (T = 178)					
Estimé à l'aide du filtre de Kalman (MV exacte)					
Variable dépendante: IPC					
Écarts type basés sur la matrice hessienne					
	coefficient	erreur std.	z	p. critique	
const	99,0832	1,40605	70,47	0,0000	***
phi_1	1,03612	0,0146358	70,79	0,0000	***
phi_12	-0,0460485	0,0154137	-2,987	0,0028	***
theta_2	0,124366	0,0717131	1,734	0,0829	*
Moy. var. dép.		100,0312			
Ec. type var. dép.		2,226065			
Moyenne des innovations		-0,001805			
Ec. type des innovations		0,191216			
Log de vraisemblance		39,37175			
Critère d'Akaike		-68,74350			
Critère de Schwarz		-52,83458			
Hannan-Quinn		-62,29201			

TABLE 8 – Modèle AR(1) en rajoutant le coefficient 12 sur AR et le coefficient 2 sur MA

Nous pouvons donc remarquer que le coefficient d'ordre 2,  $\theta_2$  est significatif pour la moyenne mobile MA. Pour avoir plus de détails sur le modèle, regardons le corrélogramme des résidus suivant :

Fonction d'auto-corrélation résiduelle  
 \*\*\*, \*\*, \* indicate significance at the 1%, 5%, 10% levels  
 using standard error  $1/T^{0.5}$

RETARD	ACF	PACF	Q	[p. crit.]
1	0,0442	0,0442		
2	-0,0116	-0,0136		
3	0,0261	0,0273		
4	0,0660	0,0636	1,3043	[0,253]
5	-0,0269	-0,0321	1,4378	[0,487]
6	0,0215	0,0253	1,5236	[0,677]
7	-0,0140	-0,0204	1,5606	[0,816]
8	0,0858	0,0860	2,9495	[0,708]
9	-0,0381	-0,0447	3,2244	[0,780]
10	-0,0404	-0,0375	3,5363	[0,831]
11	-0,1405 *	-0,1413 *	7,3219	[0,502]
12	0,0900	0,0956	8,8851	[0,448]
13	-0,0761	-0,0809	10,0106	[0,440]
14	-0,0946	-0,0791	11,7584	[0,382]
15	-0,0509	-0,0336	12,2680	[0,424]
16	0,1121	0,1000	14,7543	[0,323]
17	0,0447	0,0671	15,1528	[0,368]
18	0,0088	0,0104	15,1684	[0,439]
19	-0,0238	-0,0042	15,2822	[0,504]
20	0,0981	0,0642	17,2340	[0,439]
21	0,0297	0,0349	17,4138	[0,495]
22	-0,0165	-0,0303	17,4694	[0,558]

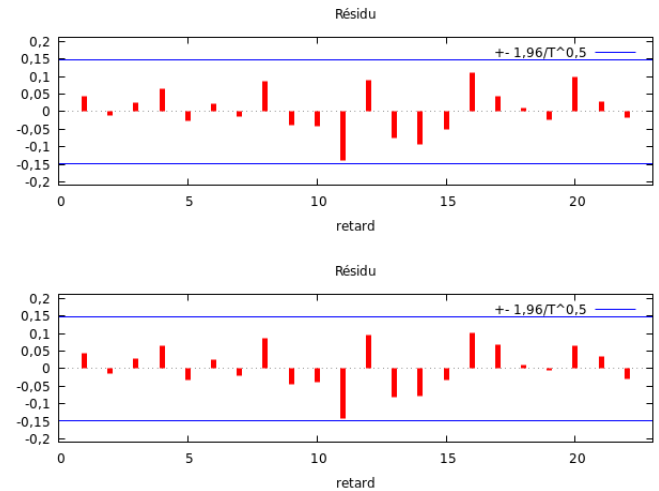


FIGURE 17 – Corrélogramme des résidus du modèle AR(1) en rajoutant le coefficient 12 sur AR et le coefficient 2 sur MA

On peut remarquer qu'il y'a une dépendance à l'ordre 11. On propose du coup de l'ajouter sur la moyenne mobile MA. On obtient les résultats suivants après cette transformation :

Évaluations de la fonction : 60				
Évaluations du gradient : 15				
Modèle 9: ARMA, utilisant les observations 2005:01-2019:10 (T = 178)				
Estimé à l'aide du filtre de Kalman (MV exacte)				
Variable dépendante: IPC				
Écarts type basés sur la matrice hessienne				
	coefficient	erreur std.	z	p. critique
const	99,3678	1,20446	82,50	0,0000 ***
phi_1	1,05835	0,0124328	85,13	0,0000 ***
phi_12	-0,0673534	0,0130674	-5,154	2,55e-07 ***
theta_2	0,0861662	0,0831491	1,036	0,3001
theta_11	-0,286974	0,0851041	-3,372	0,0007 ***
Moy. var. dép.		100,0312		
Éc. type var. dép.		2,226065		
Moyenne des innovations		-0,005577		
Éc. type des innovations		0,185551		
Log de vraisemblance		44,29653		
Critère d'Akaike		-76,59306		
Critère de Schwarz		-57,50236		
Hannan-Quinn		-68,85127		

TABLE 9 – Modèle AR(1) en rajoutant le coefficient 12 sur AR et les coefficients 2 et 11 sur MA

On remarque que le coefficient  $\theta_2$  n'est pas significatif, on peut penser à l'appliquer sur la moyenne mobile (MA). Nous avons les résultats suivants :



```

Évaluations de la fonction : 75
Évaluations du gradient : 20

Modèle 10: ARMA, utilisant les observations 2005:01-2019:10 (T = 178)
Estimé à l'aide du filtre de Kalman (MV exacte)
Variable dépendante: IPC
Écart type basés sur la matrice hessienne

```

	coefficient	erreur std.	z	p. critique	
const	99,4234	1,18993	83,55	0,0000	***
phi_1	1,18418	0,0743594	15,93	4,25e-57	***
phi_2	-0,132881	0,0800513	-1,660	0,0969	*
phi_12	-0,0593667	0,0131222	-4,524	6,06e-06	***
theta_11	-0,299098	0,0795569	-3,760	0,0002	***

```

Moy. var. dép.      100,0312
Ec. type var. dép.  2,226065
Moyenne des innovations -0,006438
Ec. type des innovations 0,184629
Log de vraisemblance 45,12587
Critère d'Akaike     -78,25173
Critère de Schwarz    -59,16103
Hannan-Quinn         -70,50994

```

TABLE 10 – Modèle AR(1) en rajoutant le coefficient 2 et 12 sur AR et le coefficient 11 sur MA

Ce dernier modèle tous les coefficients sont significatifs, et pour avoir plus de détails sur le modèle, nous allons regarder le corrélogramme des résidus suivant :

Fonction d'auto-corrélation résiduelle  
 \*\*\*, \*\*, \* indicate significance at the 1%, 5%, 10% levels  
 using standard error  $1/T^{0,5}$

RETARD	ACF	PACF	Q [p. crit.]
1	-0,0078	-0,0078	
2	0,0063	0,0062	
3	0,0216	0,0217	
4	0,0234	0,0238	
5	-0,0311	-0,0311	0,3847 [0,535]
6	0,0213	0,0201	0,4696 [0,791]
7	-0,0498	-0,0503	0,9352 [0,817]
8	0,0640	0,0642	1,7061 [0,790]
9	-0,0462	-0,0450	2,1116 [0,834]
10	-0,0118	-0,0123	2,1380 [0,907]
11	-0,0471	-0,0468	2,5643 [0,922]
12	0,0676	0,0642	3,4474 [0,903]
13	-0,0829	-0,0755	4,7829 [0,853]
14	-0,0704	-0,0765	5,7499 [0,836]
15	-0,0582	-0,0537	6,4163 [0,844]
16	0,0828	0,0766	7,7735 [0,803]
17	0,0300	0,0487	7,9521 [0,847]
18	0,0116	0,0048	7,9790 [0,890]
19	-0,0285	-0,0236	8,1424 [0,918]
20	0,0713	0,0486	9,1747 [0,906]
21	0,0186	0,0337	9,2456 [0,932]
22	-0,0142	-0,0187	9,2871 [0,953]

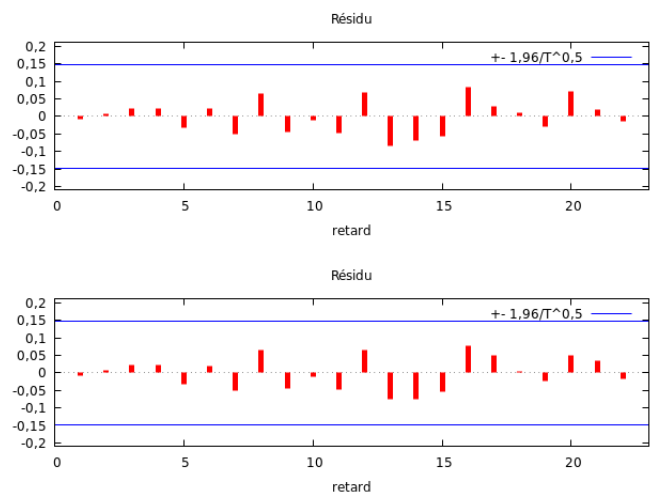


FIGURE 18 – Corrélogramme des résidus du modèle AR(1) en rajoutant les coefficients 2 et 12 sur AR et le coefficient 11 sur MA

Sur les résultats du corrélogramme des résidus, on peut noter une dépendance pour tous les coefficients. Nous pouvons donc terminer notre étude avec ce modèle.

### 3 Validation du modèle

Dans cette partie on va appliquer des tests statistiques pour vérifier si le modèle est bon. Lorsque le processus est bien estimé, les résidus se comportent comme un bruit blanc. Il ne doit pas ainsi exister d'autocorrélation dans la série.

Pour vérifier que  $\epsilon$  est de moyenne nulle, on fera le test suivant :

$$H_0 : E(\epsilon) = 0 \quad \text{sous } H_0 \quad \sqrt{T} \frac{\bar{e}}{\hat{\sigma}} \approx N(0, 1)$$

On rejette  $H_0$  avec un risque de 5% lorsque la statistique  $t = \sqrt{T} \frac{\bar{e}}{\hat{\sigma}}$  est supérieure à 1.96 en valeur absolue.

Sur notre modèle  $T = 178$  taille de l'échantillon d'estimation et :

- Moyenne des innovations  $\bar{e} = -0,006438$
- Ec. type des innovations  $\hat{\sigma} = 0,184629$

$$t = \sqrt{178} \frac{0.006438}{0.184629} = 0.4652 < 1.96$$

Donc, on ne rejette pas  $H_0$  l'espérance est nulle. On représente ensuite le graphique des résidus suivant :

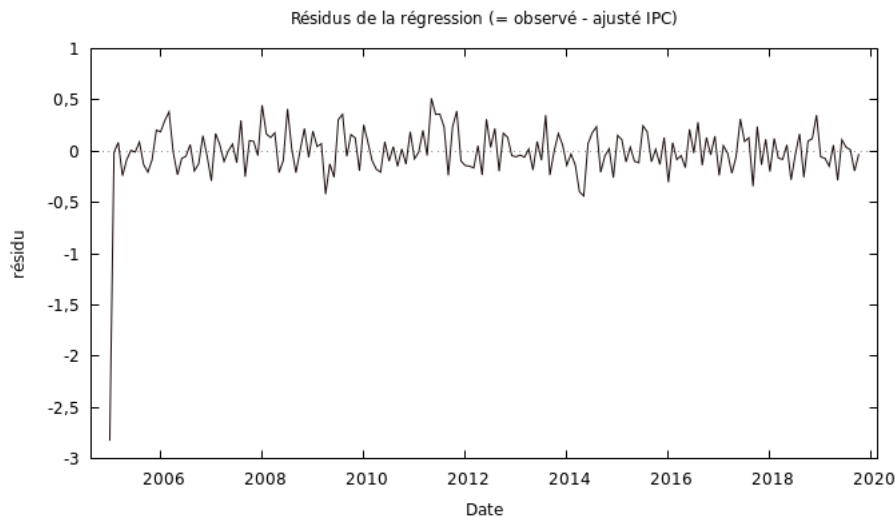


FIGURE 19 – Graphique des résidus

On remarque également que l'espérance et la variance semblent constantes au cours du temps.

On va s'intéresser maintenant au test de Portmanteau. On considère l'hypothèse nulle suivante :

$$H_0 : p_k = 0 \quad \forall k$$

On rejette  $H_0$  avec un risque de 5% lorsque l'autocorrélation estimée  $\hat{p}_k(e)$  est supérieure à  $1.96/\sqrt{T}$  en valeur absolue.

En regardant le corrélogramme des résidus dans le tableau 18, on peut voir que toutes les p-critiques sont supérieures à 5% et aucune autocorrélation ne sort de l'intervalle de confiance. Donc, on ne rejette pas l'hypothèse  $H_0$ .

Notre modèle est donc bon et nous pouvons passer à l'étape suivante.

## 4 Prévision

Maintenant que notre modèle est validé, nous pouvons faire des prévisions et étudier les performances de notre modèle en se basant sur le MAPE. À l'aide du logiciel Gretl, on obtient les prévisions ci-dessous :

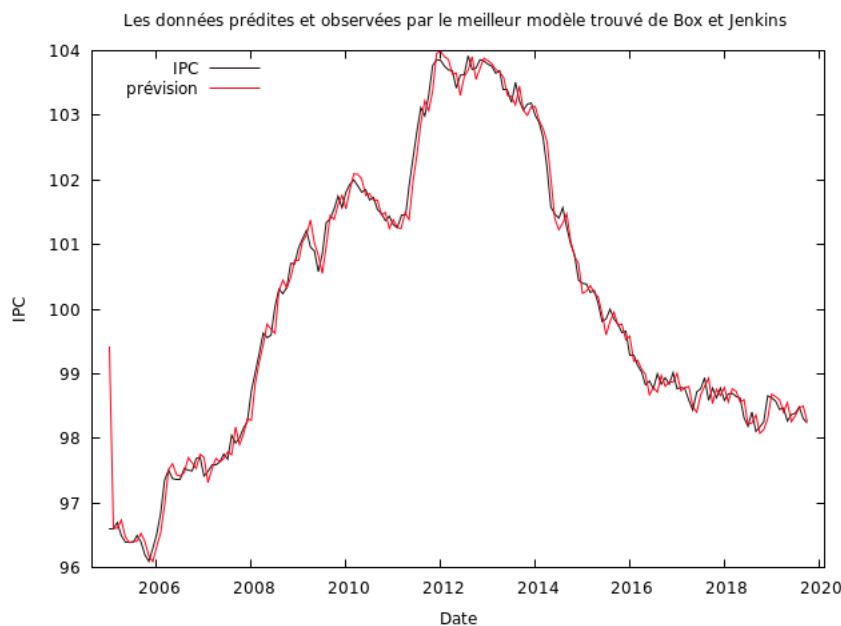


FIGURE 20 – Les valeurs prédites et observées par le meilleur modèle trouvé

Pour ce modèle nous avons obtenu un MAPE égale à 0.16%. nous utiliserons donc ce modèle pour la prévision des 12 dernières observations. Nous avons obtenu les résultats suivants :

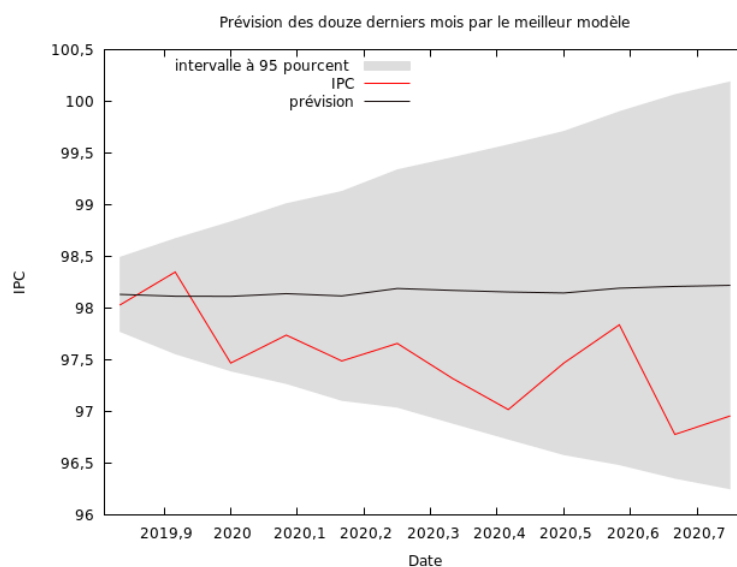


FIGURE 21 – La prédiction des douze dernières données par le meilleur modèle de Box et Jenkins

L'erreur moyenne en pourcentage obtenue avec ces 12 dernières observations est de : 0.71%. Nous allons ainsi afficher ci-dessous le graphique sur l'ensemble de l'échantillon comportant les observations et les valeurs prédites.

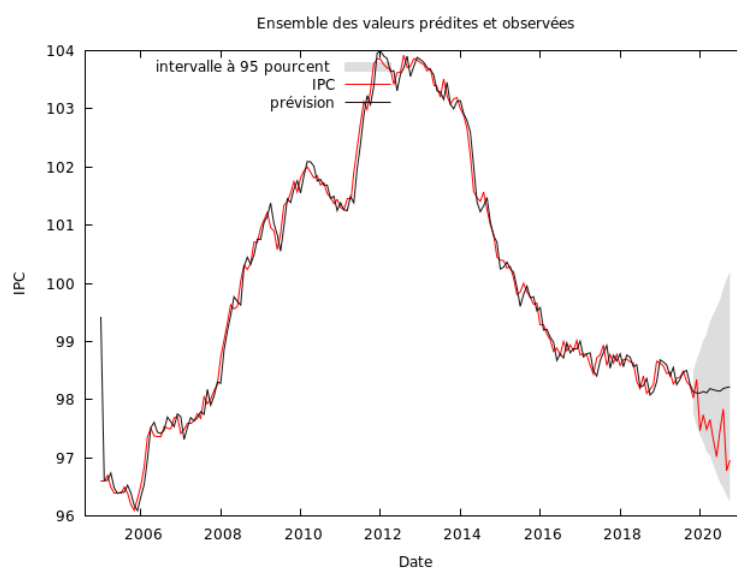


FIGURE 22 – Ensemble de valeurs prédites et observées par le meilleur modèle de Box et Jenkins

---

## *Conclusion*

Dans ce projet, on s'est intéressé à deux différentes méthodes de prévision : la méthode de lissage et la méthode de Box-Jenkins. Nous avons obtenu un meilleur MAPE de 0.15% pour la méthode de lissage(Holt-Winters cas multiplicatif) contre un MAPE de 0.16% pour la méthode de Box-Jenkins. En comparant le MAPE, on a tendance à conclure que la méthode de lissage est meilleur dans notre étude. On présente ci-dessous les valeurs de prévisions des douze derniers mois par les deux méthodes.

Date	IPC	Lissage	Box-Jenkins
2019-11	98,03	98,454	98,13
2019-12	98,35	98,759	89,12
2020-01	97,47	98,609	98,11
2020-02	97,74	98,527	98,14
2020-03	97,49	98,409	98,12
2020-04	97,66	98,393	98,19
2020-05	97,32	98,223	98,17
2020-06	97,02	98,245	98,16
2020-07	97,47	98,215	98,15
2020-08	97,84	98,346	98,19
2020-09	96,78	98,153	98,21
2020-10	96,96	98,160	98,22
MAPE		0,15 %	0,16 %

TABLE 11 – Valeurs de prévisions des douze derniers mois par les deux méthodes

---

---

## Sommaire des figures

1	Évolution de l'IPC des produits d'entretien . . . . .	4
2	Coubres joignant les maxima et minima de la série temporelle . . . . .	5
3	Graphique associé au tableau de Buys-Ballot . . . . .	7
4	Courbe de la tendance globale par moyenne mobile . . . . .	8
5	Série corrigée des variations saisonnières . . . . .	10
6	Valeurs prédites et observées par la méthode de Holt . . . . .	12
7	Valeurs prédites et observées par la méthode de Holt-Winters additive . . .	13
8	Valeurs prédites et observées par la méthode de Holt-Winters multiplicative	15
9	La prévision des douze dernières données . . . . .	16
10	Ensemble de valeurs prédites et observées par Holt-Winters multiplicative .	16
11	Corrélogramme de la série brute . . . . .	18
12	Première différence sur la série brute . . . . .	19
13	Corrélogramme de la série de la première différence . . . . .	20
14	Différence saisonnière appliquée à la première différence . . . . .	20
15	Corrélogramme de la série de la différence saisonnière appliquée à la première différence . . . . .	21
16	Corrélogramme des résidus du modèle AR(1) . . . . .	22
17	Corrélogramme des résidus du modèle AR(1) en rajoutant le coefficient 12 sur AR et le coefficient 2 sur MA . . . . .	24
18	Corrélogramme des résidus du modèle AR(1) en rajoutant les coefficients 2 et 12 sur AR et le coefficient 11 sur MA . . . . .	25
19	Graphique des résidus . . . . .	26
20	Les valeurs prédites et observées par le meilleur modèle trouvé . . . . .	27
21	La prévision des douze dernières données par le meilleur modèle de Box et Jenkins . . . . .	28
22	Ensemble de valeurs prédites et observées par le meilleur modèle de Box et Jenkins . . . . .	28

---

---

## Sommaire des tableaux

1	Tableau de Buys-Ballot contient les valeurs $X_t$ . . . . .	6
2	Tableau de Buys-Ballot contient les valeurs $T_t$ . . . . .	8
3	Tableau de Buys-Ballot pour la série CVS . . . . .	10
4	Valeur de MAPE pour différentes méthodes de lissage . . . . .	15
5	Test de Dickey-Fuller . . . . .	18
6	Modèle AR(1) appliqué à la série . . . . .	22
7	Modèle AR(1) en rajoutant les coefficients 2 et 12 . . . . .	23
8	Modèle AR(1) en rajoutant le coefficient 12 sur AR et le coefficient 2 sur MA	23
9	Modèle AR(1) en rajoutant le coefficient 12 sur AR et les coefficients 2 et 11 sur MA . . . . .	24
10	Modèle AR(1) en rajoutant le coefficient 2 et 12 sur AR et le coefficient 11 sur MA . . . . .	25
11	Valeurs de prévisions des douze derniers mois par les deux méthodes . . . .	29

---