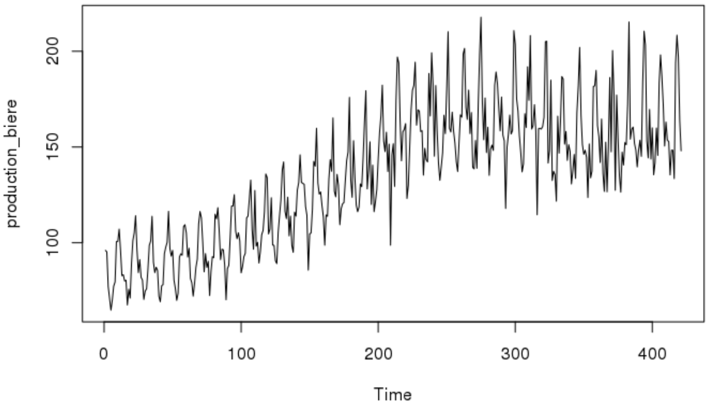
<u>DM - Série Temporelle</u>

KATHIR Ranujan - GS

I. <u>Exercice 1 : Modélisation de la série "beer" sous forme de SARIMA</u>

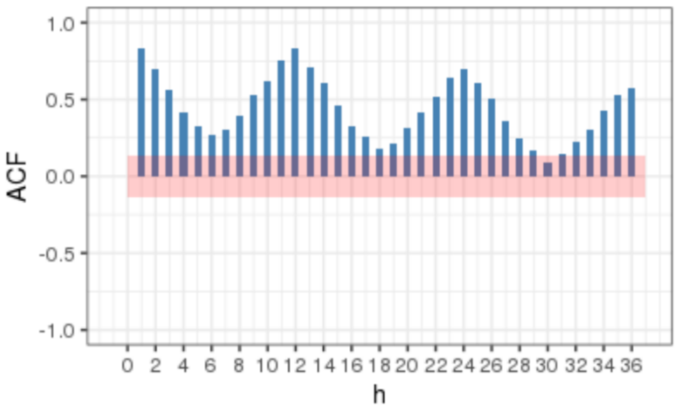
Le jeu de données “beer” contient 422 observations (mensuelles, de janvier 1956 et février 1991) de la production de bière en moyenne. On observe que, le plus de bière produite en un mois s'élève à 217.8, \textit{a contrario}, le moins de bière produite en un mois est de 64.8.

On désigne par X_t la série brute “beer”. Visualisons cette série.

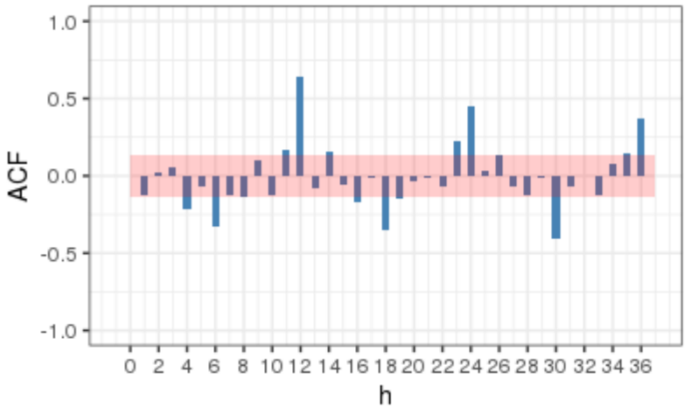


On remarque que sur le graphe, il y a 3 phases, une phase où les valeurs de production oscillent autour de 50, puis, une phase de croissance, et enfin, une phase finale avec des valeurs qui oscillent autour de 175.

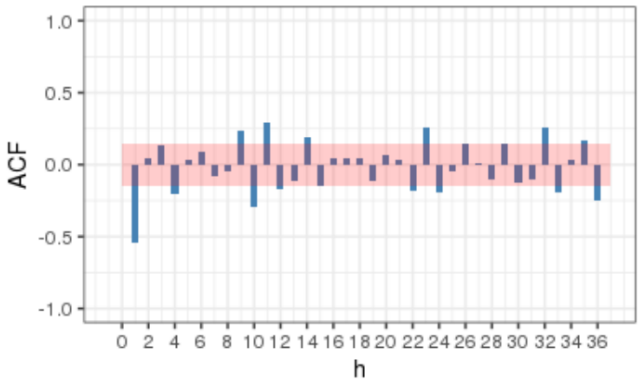
Afin de pallier l’accroissement de la saisonnalité et passer ainsi d’un modèle multiplicatif à un modèle additif, on travaille sur le logarithme de la série. On désigne par Y_t le logarithme de la série brute : $Y_t = log(X_t)$.



On observe une décroissance lente vers 0. Cela soulève un problème de non-stationnarité. On effectue alors une différenciation *(I-B).*



On observe que, la sortie ACF de la série ainsi différenciée présente encore une décroissance lente vers 0 pour les multiples de 12. On effectue cette fois la différenciation $I-B^{12}$.



La sortie ACF de la série différenciée semble être interprétable comme un autocorrélogramme simple empirique.

On identifiera donc un modèle ARMA sur la série : $(I-B)(I-B^{12})log(X_t)$

1) <u>Identification, estimation et validation du modèle</u>

On estime en premier lieu un modèle $SARIMA(1,1,1)(1,1,1)_{12}$ au vu des autocorrélogrammes empiriques simples et partiels :

$$(I-\phi_1(B))(I-\phi_1'(B^{12}))(I-B)(I-B^{12})log(X_t) = (I+\Theta_1(B))(I+\Theta_1'(B^{12}))\varepsilon_t$$

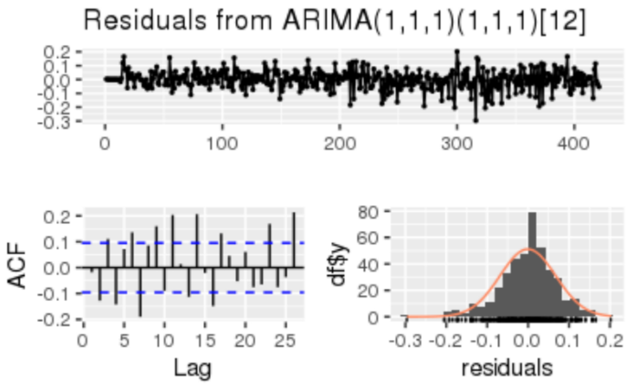
On observe les tests de significativité au niveau des paramètres.

```
[1] "Test statistic"
       ar1        ma1       sar1       sma1
 -3.201871 -45.279891   2.564482 -26.931135
[1] "p-value"
       ar1        ma1       sar1       sma1
0.00136538 0.00000000 0.01033299 0.00000000
```

Toutes les p-valeurs sont significatives, donc le modèle a des paramètres significatifs.

On réalise maintenant des tests de Ljung-Box pour observer si les résidus sont des bruits blancs ou non.

k	p_valeur
<dbl>	<dbl>
6	0
12	0
18	0
24	0
30	0
36	0



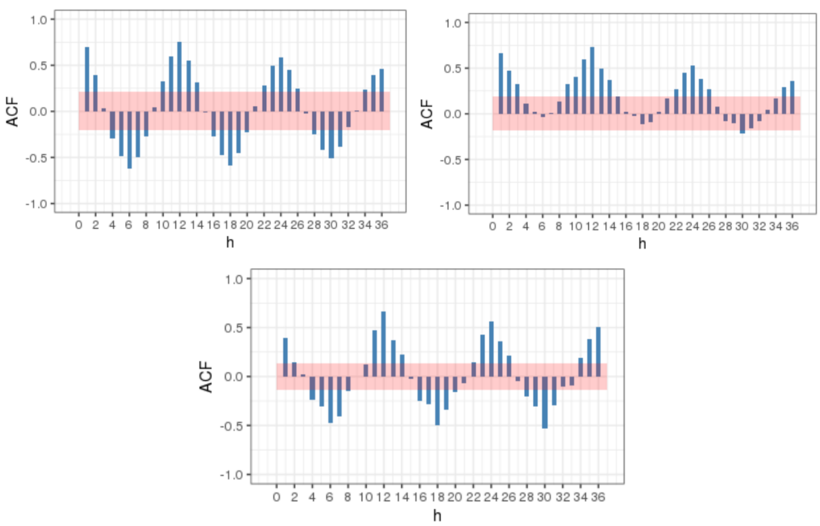
Nous sommes satisfait des résultats obtenus étant donné que les p-valeurs sont nulles.

Nous vérifierons les résultats obtenus via la sélection automatique avec le critère d'information bayésienne (BIC). On obtient un modèle *ARIMA(1,1,1)*.

2) <u>Analyse a posteriori par troncature en 3 temps de la série</u>

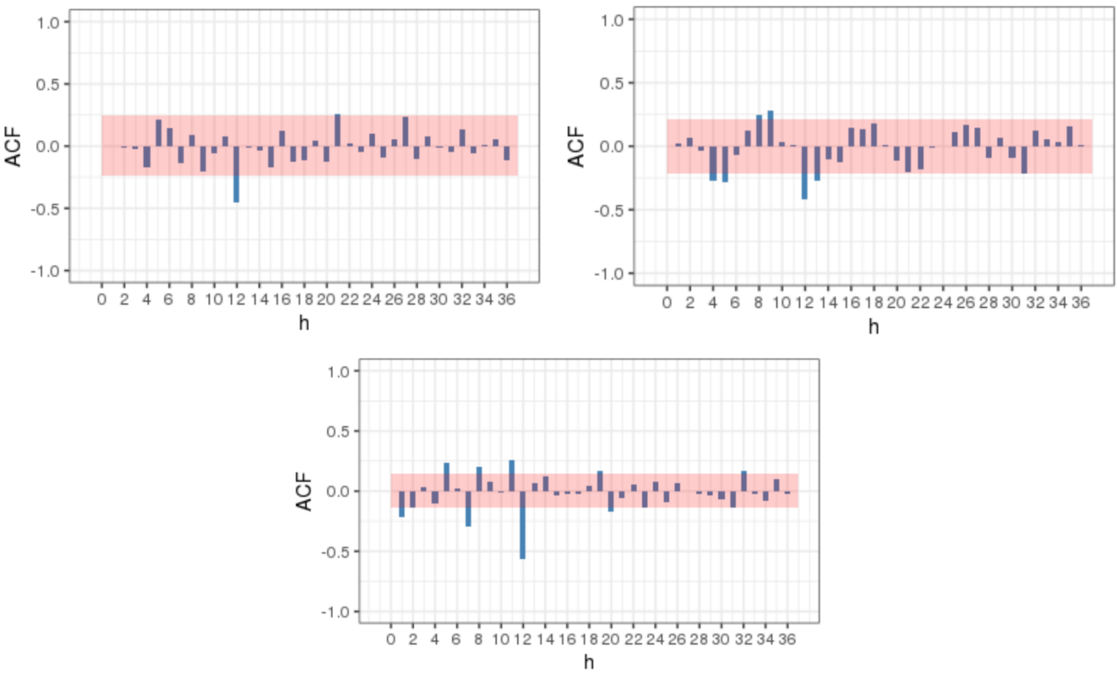
Étant donné que nous trouvons 2 modèles différents, nous allons essayer de tronquer les données. La troncature est réalisée avec des intervalles que nous avons choisis nous-même en observant le début et la fin des phases via la visualisation de la série.

A l'aide de la fonction *window*, on coupe la série en 3 temps. Pour la première, nous coupons la première partie du début à 89 (série 1ère temps), puis la seconde partie de 90 à 200 (série 2ème temps), et enfin, la dernière de 201 jusqu'à la fin (série 3ème temps).



<u>Note de lecture</u> : la première en haut à gauche représente l'ACF de la série du 1er temps; la deuxième en haut à droite représente l'ACF de la série du 2nd temps; la troisième en bas au milieu représente l'ACF du dernier temps

On constate que le PACF et le ACF sont significatifs aux ordres 12, 24 et 36, mais aussi 6, 18 et 30. Nous allons ajouter un terme autorégressif saisonnier. On effectue une différenciation $(I\text{-}B^{12})^2$.



<u>Note de lecture</u> : la première en haut à gauche représente l'ACF de la série du 1er temps différenciée; la deuxième en haut à droite représente l'ACF de la série du 2nd temps différenciée; la troisième en bas au milieu représente l'ACF du dernier temps différenciée

On modélisera les données par : $(I-B^{12})log(X_t)$.

3) <u>Identification, estimation et validation de modèles obtenus lors de l'analyse a posteriori</u>

On estime en premier lieu un modèle $SARIMA(1,0,1)(1,2,1)_{12}$ au vu des autocorrélogrammes empiriques simples et partiels :

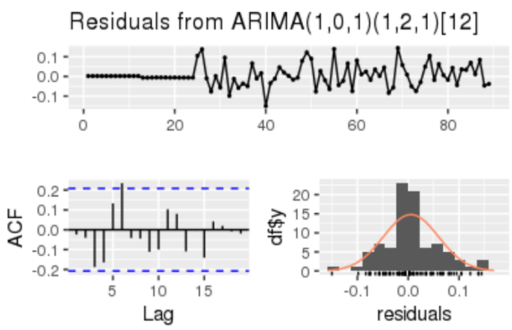
$$(I-\phi_1(B))(I-\phi_1'(B^{12}))(I-B^{12})^2 log(X_t) = (I+\Theta_1(B))(I+\Theta_1'(B^{12}))\varepsilon_t$$

On observe les tests de significativité au niveau des paramètres.

```
[1] "Test statistic"
       ar1        ma1       sar1       sma1 
-12.153295  11.533206  -1.143650  -3.358653 
[1] "p-value"
         ar1          ma1         sar1         sma1 
0.0000000000 0.0000000000 0.2527688124 0.0007832344 
[1] "Test statistic"
       ar1        ma1       sar1       sma1 
-0.1664841 -0.2192700 -1.3941280 -7.2311132 
[1] "p-value"
         ar1          ma1         sar1         sma1 
8.677760e-01 8.264397e-01 1.632790e-01 4.789502e-13 
[1] "Test statistic"
       ar1        ma1       sar1       sma1 
  0.590870  -1.479371  -4.300526 -16.498147 
[1] "p-value"
         ar1          ma1         sar1         sma1 
5.546076e-01 1.390411e-01 1.703932e-05 0.000000e+00 
```

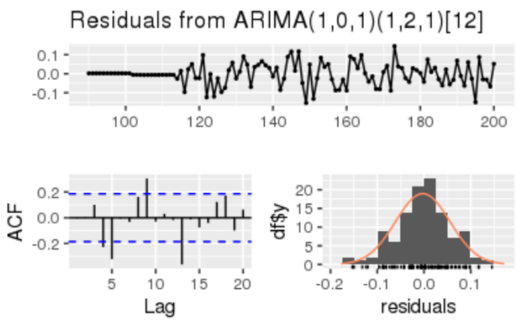
Les paramètres ne sont pas tous significatifs pour le modèle 1. En revanche, pour les autres, elles sont toutes bonnes.

k <dbl>	p_valeur <dbl>
6	0.001
12	0.025
18	0.096
24	0.094
30	0.124
36	0.194



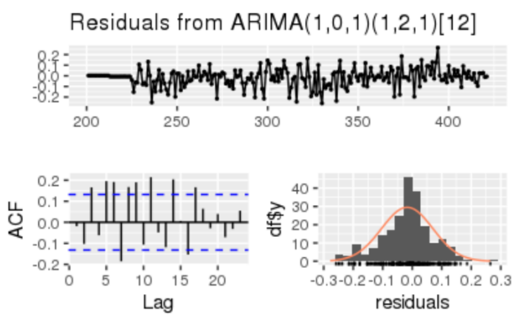
Ci-dessus, on a effectué le test de Ljung-Box pour le 1er modèle. On constate que les p-valeurs ne sont pas significatives. Donc, le test de blancheur n'est pas satisfaisant.

k <dbl>	p_valeur <dbl>
6	0
12	0
18	0
24	0
30	0
36	0



Ci-dessus, on a effectué le test de Ljung-Box pour le 2ème modèle. On constate que le test de blancheur est satisfaisant. De plus, on observe pas de tendance particulière dans le graphe de l'ACF.

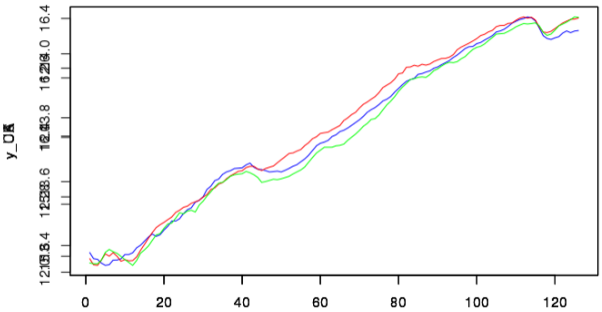
k <dbl>	p_valeur <dbl>
6	0
12	0
18	0
24	0
30	0
36	0



Ci-dessus, on a effectué le test de Ljung-Box pour le 3ème modèle. On constate que le test de blancheur est satisfaisant. De plus, on observe pas de tendance particulière dans le graphe de l'ACF.

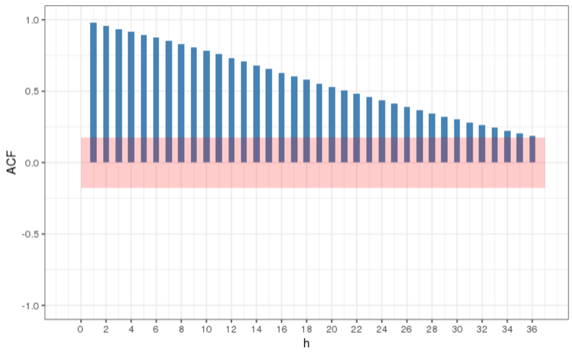
II. Exercice 3 : Modélisation du jeux de données PIB trimestriels du Royaume-Uni, du Canada et des USA, entre 1980 et 2011 sous ARMA et VAR

On importe le jeux de données PIB, puis on le sépare en 3 parties en fonction des pays.
On visualise les séries obtenues. Pour la légende, en rouge, le PIB des Etats-Unis, en vert, le PIB du Canada et en bleu, le PIB du Royaume-Unis.

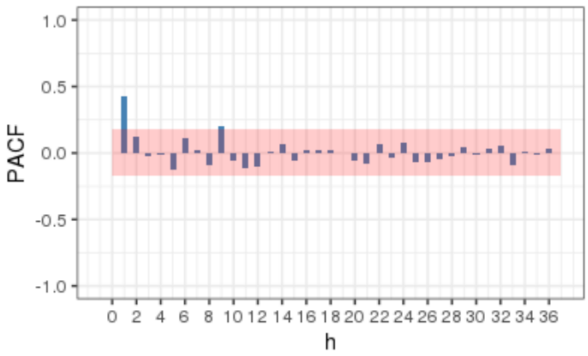


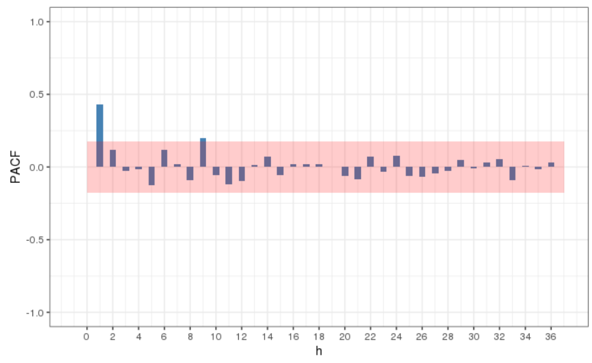
Afin de pallier l'accroissement de la saisonnalité et passer ainsi d'un modèle multiplicatif à un modèle additif, on travaille sur le logarithme de la série. On désigne par Y_t le logarithme de la série brute : $Y_t = log(X_t)$. On réalise cela sur les trois séries.

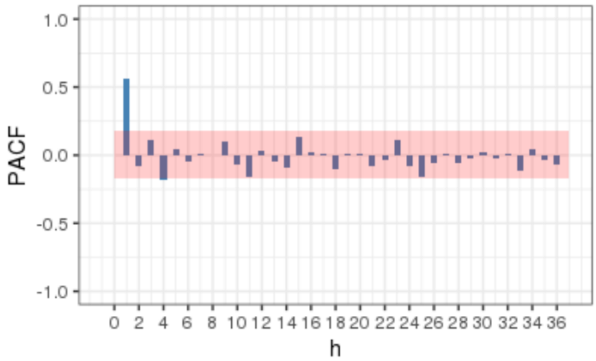
1) <u>Etude de la saisonnalité des séries</u>



Le graphe de l'ACF montre une décroissance lente vers 0. Cela soulève un problème de non-stationnarité. On effectue une différenciation *(I-B)*. Sur les PACF, il n'y a aucun pic significatif.







<u>Note de lecture</u> : la première en haut à gauche représente la PACF de la série du 1er temps différenciée; la deuxième en haut à droite représente la PACF de la série du 2nd temps différenciée; la troisième en bas au milieu représente la PACF du dernier temps différenciée

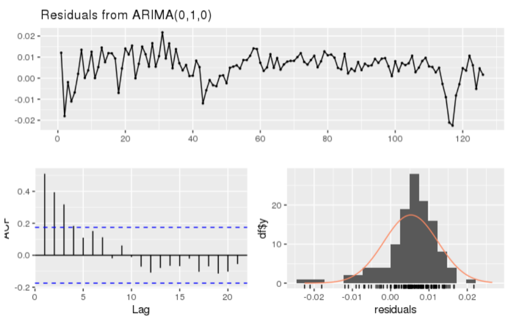
Pour les deux premières PACF, on observe un pic au niveau du 9 et pour la dernière, on observe un pic au niveau 4.

2) <u>Modélisation en AR(1)</u>

On estime en premier lieu un modèle *AR(1)* au vu des autocorrélogrammes empiriques simples et partiels :

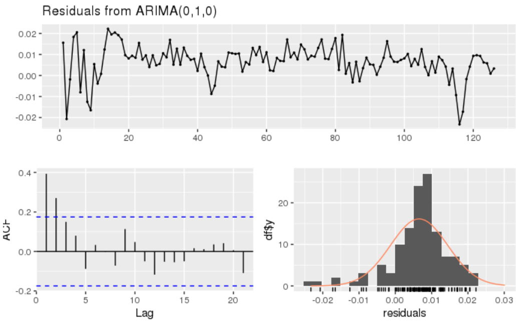
$$(I-\phi_1(B))Y_t = \varepsilon_t$$

k <dbl>	p_valeur <dbl>
6	0
12	0
18	0
24	0
30	0
36	0



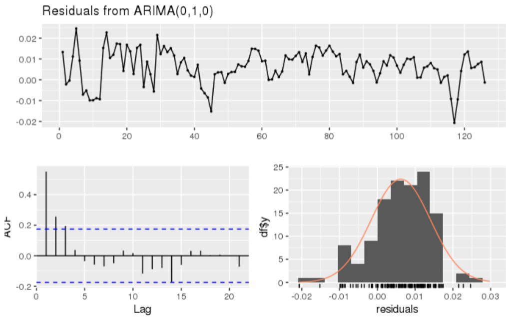
Ci-dessus, on a effectué le test de Ljung-Box pour le 1er modèle. On constate que le test de blancheur est satisfaisant. De plus, on observe pas de tendance particulière dans le graphe de l'ACF.

k	p_valeur
<dbl>	<dbl>
6	0.000
12	0.000
18	0.002
24	0.008
30	0.032



Ci-dessus, on a effectué le test de Ljung-Box pour le 2ème modèle. On constate que les p-valeurs ne sont pas significatives. Donc, le test de blancheur n’est pas satisfaisant.

k	p_valeur
<dbl>	<dbl>
6	0
12	0
18	0
24	0
30	0
36	0



Ci-dessus, on a effectué le test de Ljung-Box pour le 3ème modèle. On constate que le test de blancheur est satisfaisant. De plus, on observe pas de tendance particulière dans le graphe de l’ACF.

Nous vérifierons les résultats obtenus via la sélection automatique avec le critère d’information bayésienne (BIC). On obtient un modèle *ARIMA(1,1,0)*.