Série temporelle Projet

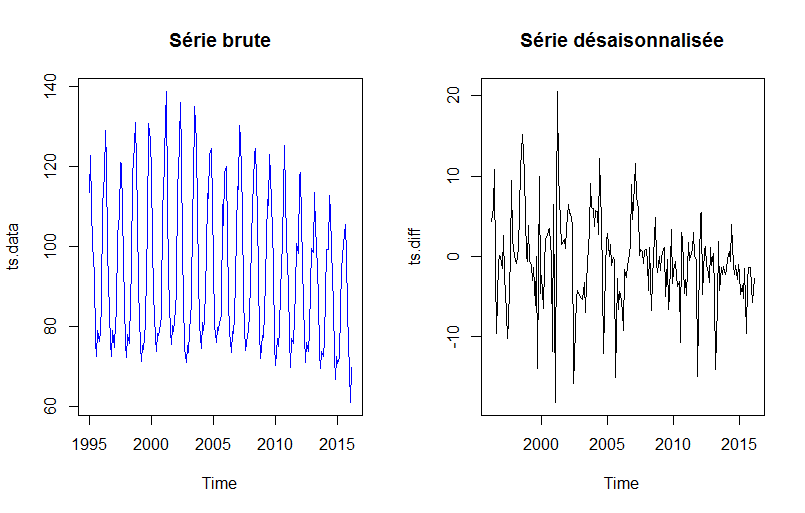
**Partie 1 : Les données**

1. **Que représente la série choisie ? (secteur, périmètre, traitements éventuels, transformation logarithmique, …)**

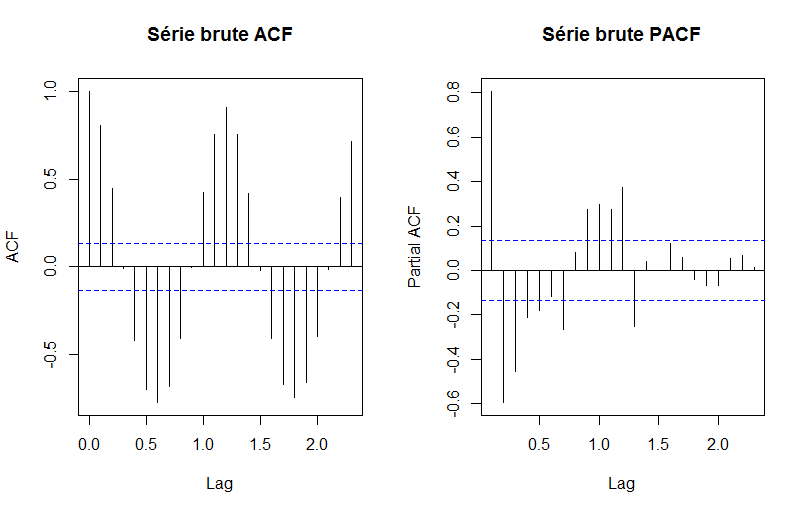
On choisit la série  « Production et distribution d'électricité, de gaz, de vapeur et d'air conditionn », qui concerne la plupart des habitants français. On a trouvé judicieux d’étudier un indicateur relatif à la vie quaditienne.

1. **Transformer si besoin la série pour la rendre stationnaire (désaisonnalisation, différentiation, suppression de la tendance déterministe, ..). Justifier soigneusement vos choix. + 3) Représenter graphiquement la série choisie avant et après transformation.**

La courbe de la série brute met en évident que la série brute est saisonnière (T=12). Afin de la rendre stationnaire, on a besoin de la désaisonnaliser par la différence saisonnière (Lag=12) et puis de retirer la tendance déterministe si besoin. Voici les deux courbes des série avant (ts.data) et après le traitement(ts.diff).



On produit l’ACF et la PACF pour vérifier la saison de la série bruite comme suivant :



On en apprend que les deux indicateurs s’explosent autour de Lag=12 qui confirme notre conclusion précédente. Selon la graphe, il n’y a pas de tendance déterministe sur la série ts.diff et on va faire la test ADF et PP pour vérifier la stationnarité de la série « ts.diff » telle que :

|  |
| --- |
| Augmented Dickey-Fuller Test  data: ts.diff  Dickey-Fuller = -4.7646, Lag order = 5, p-value = 0.01  alternative hypothesis: stationary |

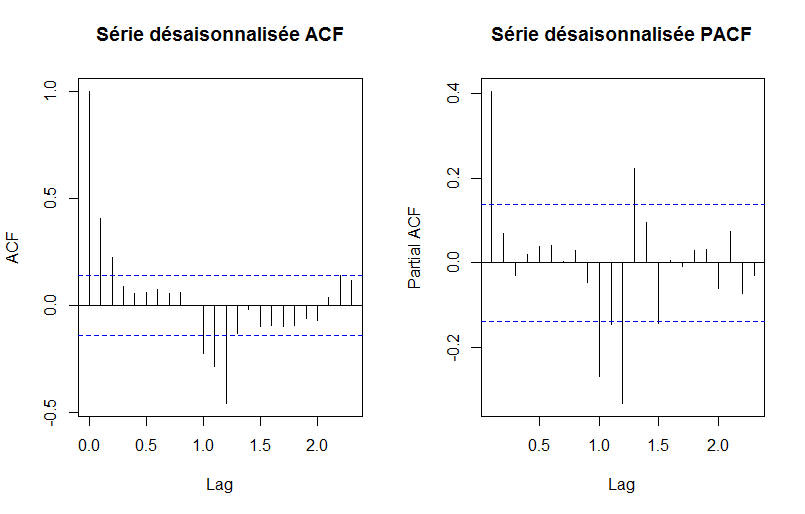
|  |
| --- |
| Phillips-Perron Unit Root Test  data: ts.diff  Dickey-Fuller Z(alpha) = -124.62, Truncation lag parameter = 4, p-value = 0.01  alternative hypothesis: stationary |

Les résultats de tests nous rendent compte que la série « ts.diff » est stationnaire et par conséquent on a trouvé judicieux de l’estimer par le modèle ARMA

**Partie 2 : Modèles ARMA**

1. **Choisir, en le justifiant, un modèle ARMA(p,q) (avec éventuellement une composante saisonnière) our votre série corrigée Xt. Estimer les paramètres du modèle et vérifier sa validité.**
2. **Détermination des ordres du modèle**

On analyse tout d’abord l’ACF et la PACF pour déterminer les ordres du modèle comme suivant :



Bien que la désaisonnalisation améliore l’ACF et la PACF à certain niveau, on constate encore les pics autour de Lag=12, ce qui s’explique par le fait que :

Cela nous fait penser au modèle SARIMA((p,d,q),(P,D,Q)) pour éliminer cette corrélation dans la série ts.diff. Vu que ts.diff est déjà stationnaire, d=D=1. Selon l’analyse ACF et PACF, on met pour l’instant que p=1 et q=2 pour déterminer P et Q. Par la méthode d’exhaustion, on trouve que P=0 et Q=1 sont pertinents. Pour ce vérifier, on a simulé la série ts.diff par (P,Q)=(1,1) et (P,Q)=(0,2) tandis qu’au final les paramètres estimés ne sont plus significatifs. Voici les résultats des simulations concernées.

|  |
| --- |
| Call:  arima(x = ts.diff, order = c(1, 0, 2), seasonal = list(order = c(0, 0, 1), period = 12))  Coefficients:  ar1 ma1 ma2 sma1 intercept  0.9903 -0.6093 -0.2788 -0.7305 -0.5206  s.e. 0.0154 0.0673 0.0636 0.0556 0.7315  ar1 ma1 ma2 sma1 intercept  64.2012003 -9.0488626 -4.3875429 -13.1327847 -0.7116312 |

|  |
| --- |
| Call:  arima(x = ts.diff, order = c(1, 0, 2), seasonal = list(order = c(1, 0, 1), period = 12))  Coefficients:  ar1 ma1 ma2 sar1 sma1 intercept  0.9912 -0.6185 -0.2737 -0.0937 -0.6818 -0.5245  s.e. 0.0137 0.0674 0.0630 0.0967 0.0779 0.7931  ar1 ma1 ma2 sar1 sma1 intercept  72.492637 -9.181550 -4.346413 -0.968601 -8.750800 -0.661396 |

|  |
| --- |
| Call:  arima(x = ts.diff, order = c(1, 0, 2), seasonal = list(order = c(0, 0, 2), period = 12))  Coefficients:  ar1 ma1 ma2 sma1 sma2 intercept  0.9915 -0.6217 -0.2722 -0.7920 0.0949 -0.5297  s.e. 0.0131 0.0674 0.0629 0.0795 0.0835 0.8282  ar1 ma1 ma2 sma1 sma2 intercept  75.7863532 -9.2287307 -4.3284445 -9.9623597 1.1366152 -0.6395858 |

Par conséquent, on fixe (P,Q)=(0,1) dans les tests suivants et puis détermine p et q par la matrix AIC et BIC telles que ci-après :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| AIC(p,q) | 0 | 1 | 2 | 3 | 4 |
| 0 | 1209.087 | 1176.282 | 1167.657 | 1165.523 | 1166.423 |
| 1 | 1161.521 | 1160.607 | 1160.565 | 1158.53 | 1159.611 |
| 2 | 1161.016 | 1156.273 | 1158.1 | 1159.857 | 1157.452 |
| 3 | 1162.736 | 1158.065 | 1156.737 | 1158.168 | 1158.71 |
| 4 | 1164.51 | 1159.855 | 1157.911 | 1160.254 | 1159.616 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| BIC(p,q) | 0 | 1 | 2 | 3 | 4 |
| 0 | 1211.004 | 1178.839 | 1170.852 | 1169.357 | 1170.896 |
| 1 | 1164.077 | 1163.802 | 1164.399 | 1163.003 | 1164.724 |
| 2 | 1164.211 | 1160.107 | 1162.573 | 1164.969 | 1163.204 |
| 3 | 1166.57 | 1162.538 | 1161.85 | 1163.919 | 1165.101 |
| 4 | 1168.983 | 1164.968 | 1163.663 | 1166.645 | 1166.646 |

En synthétisant les deux matrices, on en conclut que p=2 et q=1. Ainsi, SARIMA((2,0,1),(0,0,1)) pour simuler la série désaisonnalisée. On refait la simulation pour calibrer ce modèle comme suivant :

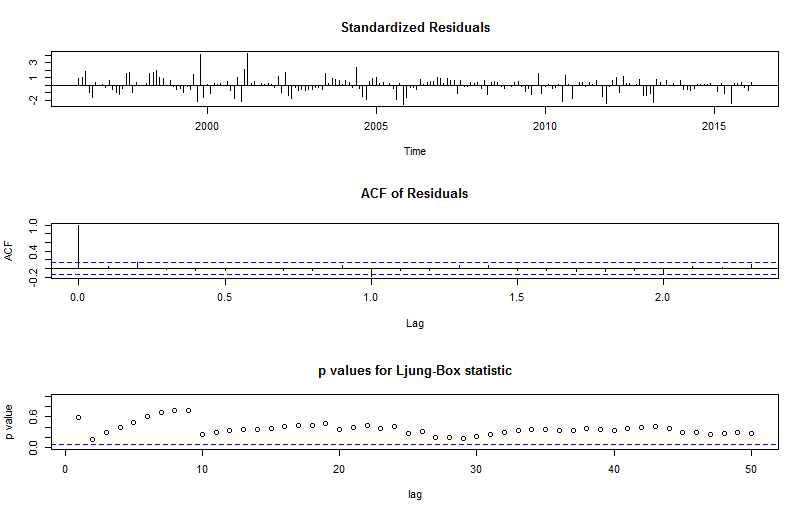
|  |
| --- |
| Call:  arima(x = ts.diff, order = c(2, 0, 1), seasonal = list(order = c(0, 0, 1), period = 12))  Coefficients:  ar1 ar2 ma1 sma1 intercept  1.3654 -0.3703 -0.9393 -0.7298 -0.5164  s.e. 0.0734 0.0720 0.0281 0.0569 0.7262  ar1 ar2 ma1 sma1 intercept  18.5962436 -5.1421469 -33.4778337 -12.8294320 -0.7111624 |

On voit bien que tous les params sont significatifs et on réécrit le modèle explicit par la formule suivante:

C’est aussi un modèle ARMA(p=2, q=13).

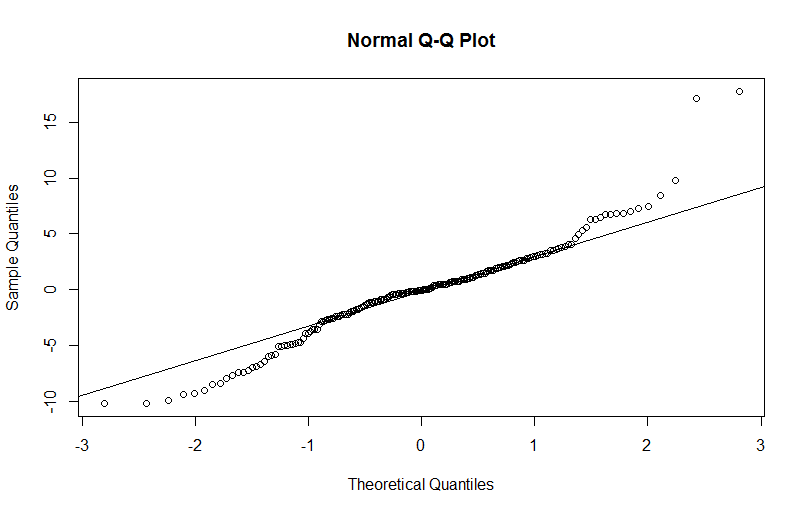
1. **Test des résidus**

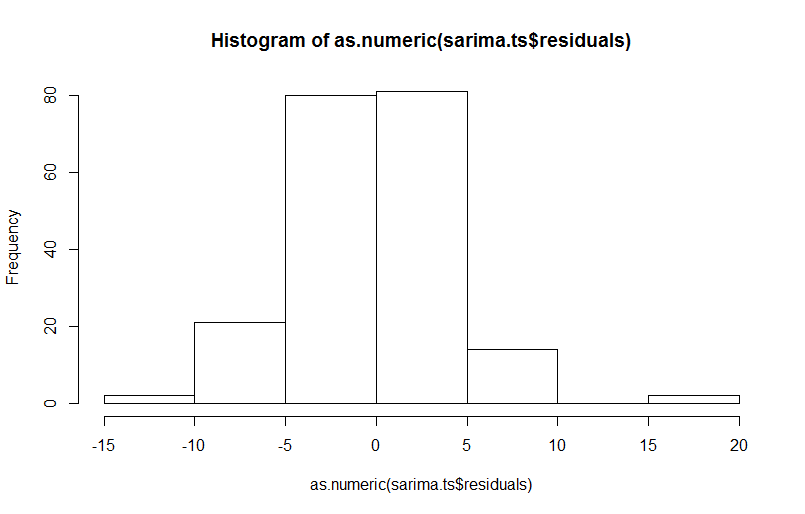
On choisit d’utiliser la fonction tsdiag pour tester si les résidus sont des bruits blancs.



Le 2ième graphe nous dit que les résidus ne sont pas corrélés sauf la point 0 qui correspont seulement à la variance des résidus. En plus, les p-values du test Liung-Box justifie qu’on ne rejette pas H0 :les résidus suivent la loi normale. Ces arguments nous faire parvenir à la conclusion que les résidus du modèle sont des bruits blancs.

D’autre part, on met en place le test qqnorm pour vérifier la normalité des résidus. Selon la graphe dessous, la plupart de point se trouve sur la droite, ce qui rejustifie la normalité des résidus du modèle.





**Partie 3 : Prévision**

**On note T la longueur de la série. Ici, T=312 (12 mois sur 26 ans). On suppose que les résidus de la série sont gaussiens.**

**5) Ecrire l’équation vérifiée par la région de confiance de niveau α sur les valeurs futures (XT+1,XT+2).**

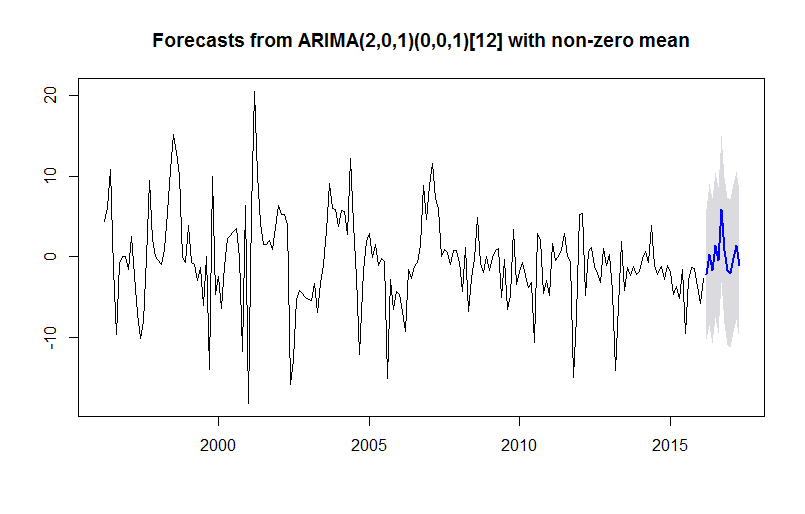
**6) Préciser les hypothèses utilisées pour obtenir cette région.**

**7) Déterminer graphiquement cette région pour α=95%. Commenter.**

Le fonction R « forecast » s’emploie afin de faire la prédiction. Voici les résultats pour XT+1,XT+2

|  |
| --- |
| Point Forecast Lo 95 Hi 95  2016.2 -2.1791297 -10.283208 5.924949  2016.3 0.3292976 -8.479642 9.138237 |

Voici les courbes prédites dans 12 mois avec l’intervalle de confiance pour 95%.



8) Question ouverte : soit Yt une série stationnaire disponible de t=1 à T. On suppose que YT+1 est disponible plus rapidement que XT+1. A quelles conditions cette information permet-elle d’améliorer la prévision de XT+1 ?