Projet séries temporelles linéaires

Antoine Joubrel et Quentin Menner

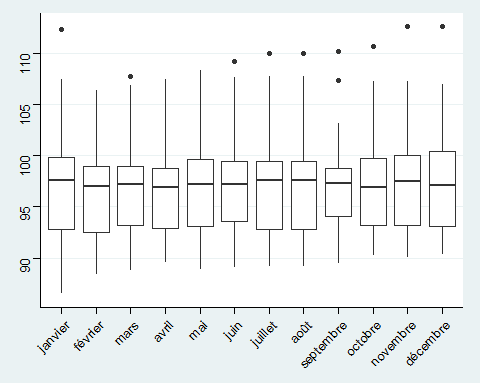
## Partie I : Les données

### Question 1

La série étudiée représente la production mensuelle de biens manufacturés en France métropolitaine entre janvier 1985 et janvier 2000. C’est une série agrégée corrigée des variations saisonnières et des jours ouvrés (CVS-CJO) contenant 181 observations. L’unité de mesure est un indice de base 100 en 1990. La série est de la forme suivante :

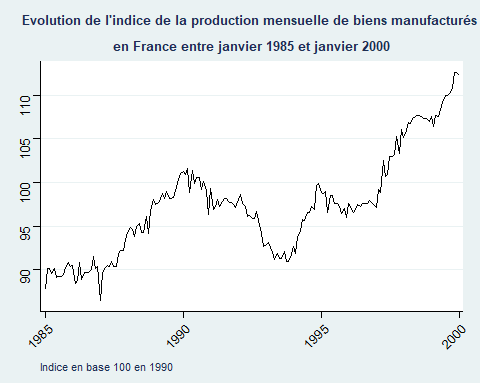
### Question 2

1. Nous sommes d’abord demandés si la série avait une saisonnalité. Une façon de le voir est de regarder la répartition de l’indice pour chaque mois :



Nous avons conclu qu’aucune saisonnalité semble se dégager.

1. Toutefois la série possède une tendance nette à la hausse malgré une baisse de l’indice entre 1990 et 1994.

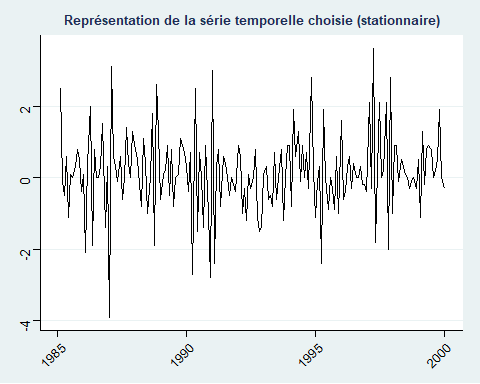
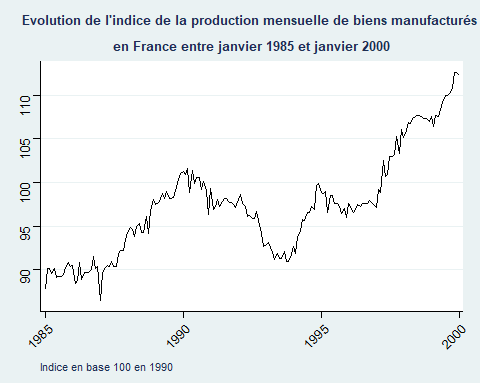


Elle est nette quand on réalise une régression linéaire de la série sur le temps (coefficent de 0.087).

Nous avons donc différencié la série puis réalisé des tests de stationnarité sur cette nouvelle série.

Les test ADF et PP rejette l’hypothèse de racine unité et le test KPSS ne rejette pas l’hypothèse de stationnarité de la série. Nous considérons donc que la série différenciée est stationnaire.

### Question 3

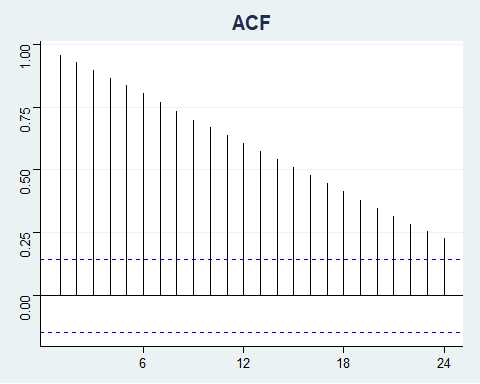


## Partie II : Modèles ARMA

### Question 4

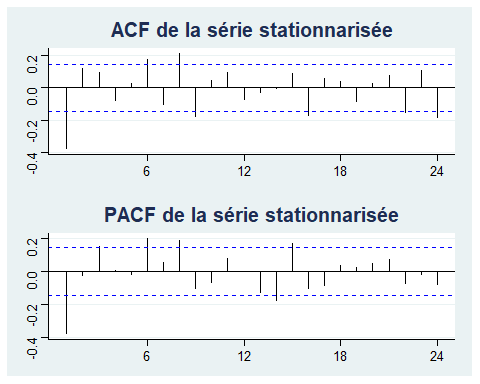
Une caractéristique importante de notre série temporelle (non stationnarisée) est l’autocorrélation des résidus. On observe en effet qu’ils sont linéairement décroissants et en dehors de l’interval de confiance comme c’est le cas dans un AR pure. Nous avons donc fait l’hypothèse que la série est un AR pure.

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo



Continuous tout de même la méthode de Box-Jenkins pour voir si notre hypothèse se vérifie.

Les graphiques de l’ACF et du PACF de la série stationnaire donnent comme valeur maximale des coeffecients AR et MA : 1.



On évalue donc 4 modèles : ARMA(0,0), ARMA(1,0) (ou AR(1)), ARMA(0,1) (ou MA(1)) et ARMA(1,1) en regardant d’abord l’autocorrélation des résidus et la significativité des coefficients.

* Pour le modèle ARMA(0,0) (bruit blanc), on estime le coefficient de la constante:

## Warning: package 'knitr' was built under R version 4.0.5

## # A tibble: 1 x 5  
## term estimate std.error statistic p.value  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 intercept 0.137 0.0836 1.64 0.102

Ce paramètre n’est pas significative.

Puis nous avons étudié l’autocorrélation des résidus par des tests de Ljung-Box :

## lag b.statistic b.p.value   
## [1,] 1 NA NA   
## [2,] 2 29.05626 7.030669e-08  
## [3,] 3 30.76253 2.089307e-07  
## [4,] 4 32.15385 4.856948e-07  
## [5,] 5 32.31955 1.645929e-06  
## [6,] 6 38.24597 3.367374e-07  
## [7,] 7 40.44434 3.725091e-07  
## [8,] 8 48.98071 2.288786e-08  
## [9,] 9 55.46514 3.584941e-09  
## [10,] 10 55.90873 8.170861e-09

Les résidus ne sont pas du tout autocorrélés. Mais**on ne retient pas ce modèle** car le coefficient n’est pas significatif.

* On procède de même pour le modèle AR(1) :

## # A tibble: 2 x 5  
## term estimate std.error statistic p.value  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 ar1 -0.388 0.0694 -5.59 0.0000000222  
## 2 intercept 0.134 0.0556 2.40 0.0163

Les coefficients sont significatifs aux seuils usuels.

## lag b.statistic b.p.value   
## [1,] 1 NA NA   
## [2,] 2 NA NA   
## [3,] 3 4.091048 0.04311091   
## [4,] 4 4.720079 0.09441648   
## [5,] 5 5.985802 0.112303   
## [6,] 6 13.86031 0.007754497  
## [7,] 7 14.06167 0.0152224   
## [8,] 8 18.53717 0.005021067  
## [9,] 9 21.7933 0.002757393  
## [10,] 10 21.91809 0.005069882  
## [11,] 11 24.59587 0.0034523   
## [12,] 12 25.77409 0.004056201  
## [13,] 13 27.38773 0.004014003  
## [14,] 14 27.40036 0.006763869  
## [15,] 15 27.70803 0.009937022

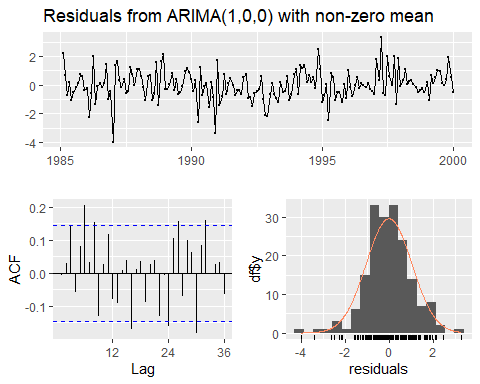
Globalement, les résidus ne sont pas corrélés entre eux (à part pour l’ordre 5). Donc **le modèle AR(1) est retenu.**

De même **le modèle MA(1) est retenu** car les résidus sont encore moins autocorrélés et les coefficients sont significatifs. Toutefois **le modèles ARMA(1,1) n’est pas retenu** car les coefficients ne sont pas tous significatifs.

Pour départager le modèle AR(1) du modèle MA(1) nous avons comparé les critères d’informations (AIC et BIC) qui montre que le modèle AR(1) est meilleur.

Est-il valide ? Pour cela, nous avons observé la répartition des résidus.

## Warning: package 'forecast' was built under R version 4.0.5



##   
## Ljung-Box test  
##   
## data: Residuals from ARIMA(1,0,0) with non-zero mean  
## Q\* = 44.817, df = 22, p-value = 0.002799  
##   
## Model df: 2. Total lags used: 24

### Question 5

Nous supposons que donc la production mensuelle de biens manufacturés suit un **ARIMA(1,1,0)** car la série avait besoin d’être différencié une fois pour être stationnaire. Le coefficients lié à l’AR(1) est estimé à **-0.388** et la constante est de **0.134**. La formule de la série est donc :

avec un bruit blanc gaussien.

## Partie III : Prévision

### Question 6

Comme nous avons réalisé une différentiation d’un mois à l’autre dans la question 2, nous prenons ici le modèle AR(1) de la forme suivante :

Avec $(\epsilon\_{t})\_{t} \sim {\sf Norm}(0,1)$. Avec une probabilité de , . Ce qui donne : $$ $$ D’après la partie II, . Donc on a comme région de confiance de niveau :

Et :

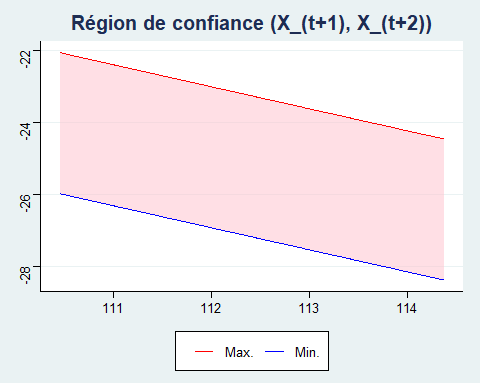
Ce qui donne comme interval de confiance pour :

### Question 7

Pour obtenir cette région de confiance, nous utilisons les hypothèses que les résidus suivent une loi normale centrée réduite, que la série temporelle est considérée comme stationnaire une fois la différenciation effectuée et qu’elle peut être modélisée par un modèle ARIMA de paramètres (1,1,0).

### Question 8

On prend ici une région de confiance de niveau 95%. Donc et . On considère T et T-1 comme étant respectivement la dernière et l’avant-dernière période de notre série temporelle ce qui donne et . Donc et

Cela nous donne la région de confiance suivante pour en fonction de : 

### Question 9

On cherche à déterminer une méthode permettant de prédire à à partir d’une autre série et plus précisément de sa valeur à la T+1ème période .

On sait que est une série stationnaire. De même dans notre cas, est aussi stationnaire. Comme on considère que l’on a avant , on peut donc réaliser la régression suivante : Pour pouvoir réaliser cette régression il faut s’assurer que la condition suivante est respectée : On doit en plus considérer comme condition que les observations sont comme un échantillon donc que l’écart type estimé de notée est juste ce qui ne serait pas le cas normalement. Dans ce cas, on peut réaliser un test de contre . Pour tester au niveau , nous considérons la t-statistique et la région critique : où est le quantile d’ordre de .