

Séries temporelles

Correction examen 2018-2019

1) Choix du modèle

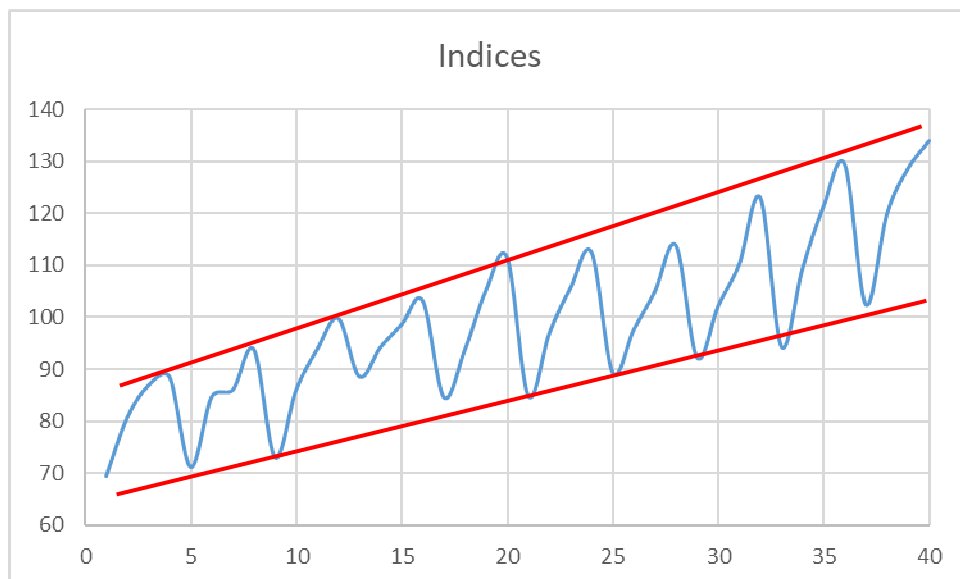
Pour représenter la série (X_t) nous avons choisi le modèle

$X_t = \dots$

Utilisez la représentation graphique de la série pour répondre aux questions suivantes.

- 1) Justifiez le choix d'un modèle additif.
- 2) Justifiez la tendance linéaire $at+b$. Que pouvez-vous dire sur les coefficients a et b ?
- 3) Justifiez l'ajout de 4 composantes saisonnières.
- 4) Quelles sont les hypothèses sur le bruit et sur les coefficients de la saisonnalité. Pourquoi ?

Correction



- 1) Les deux droites reliant les max et les min de la série sont presque parallèles, d'où le choix d'un modèle additif, $x_t = m_t + S_t + \varepsilon_t$.
- 2) En moyenne, on constate que les indices croissent avec le temps. On aura donc $a > 0$. Au temps 0, l'indice n'est pas nul donc on aura aussi $b > 0$.
- 3) On remarque une périodicité sur les trimestres, d'où l'ajout des 4 composantes saisonnières.
- 4) Somme des erreurs = 0, $\text{var}(\varepsilon_t) = \text{constante}$ et $\text{cov}(\varepsilon_t, \varepsilon_s) = 0$ pour que ε_t soit un bruit sans plus aucune information.
 $c_1 + c_2 + c_3 + c_4 = 0$ pour unicité de l'écriture

II) Estimation par moindres carrés (avec R)

- 1) Notons X la matrice du modèle, β le vecteur des coefficients et Y le vecteur des observations. Donnez la formule permettant d'estimer les coefficients. Ecrivez le vecteur β ainsi que les 5^{èmes} lignes de X et Y .
- 2) Donnez les coefficients estimés par R.
- 3) Pourquoi n'y-a-t-il pas de composante S_1 ?
- 4) Quelle(s) hypothèse(s) doit-on ajouter sur le bruit pour pouvoir établir des tests statistiques ?
- 5) Commentez les résultats obtenus par R.
- 6) Représentez graphiquement les erreurs et commentez.
- 7) Représentez sur un même graphique la série observée (`plot`) et la série prédite (`points`) entre 1995 et 2004.

P.S. Pour obtenir une ligne et non des points dans les instructions graphiques `plot` et `points`, il faut préciser `type="l"`.

Correction

1) $\beta^{\wedge} = ({}^tXX)^{-1}XY$

	b	a	S1	S2	S3	S4		Y	
	1	1	1	0	1	0		69,4	b
	1	2	0	1	0	0		80,73	a
X=	1	3	0	0	1	0	Y=	86,87	$\beta =$ C ₁
	1	4	0	0	0	1		88,5	C ₂
	1	5	1	0	0	0		71,17	C ₃
	C ₄

2)

```
>
> modele=lm(Indices~t+Trimestres,data=serie[1:32,])
> summary(modele)

Call:
lm(formula = Indices ~ t + Trimestres, data = serie[1:32, ])

Residuals:
    Min       1Q   Median       3Q      Max
-4.5759 -2.1324 -0.5256  1.3251  8.7479

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  68.05199    1.42212   47.853 < 2e-16 ***
t              0.90770    0.06004   15.118 1.06e-14 ***
TrimestresT2   9.46230    1.55762    6.075 1.74e-06 ***
TrimestresT3  15.45960    1.56109    9.903 1.75e-10 ***
TrimestresT4  21.39315    1.56685   13.654 1.23e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

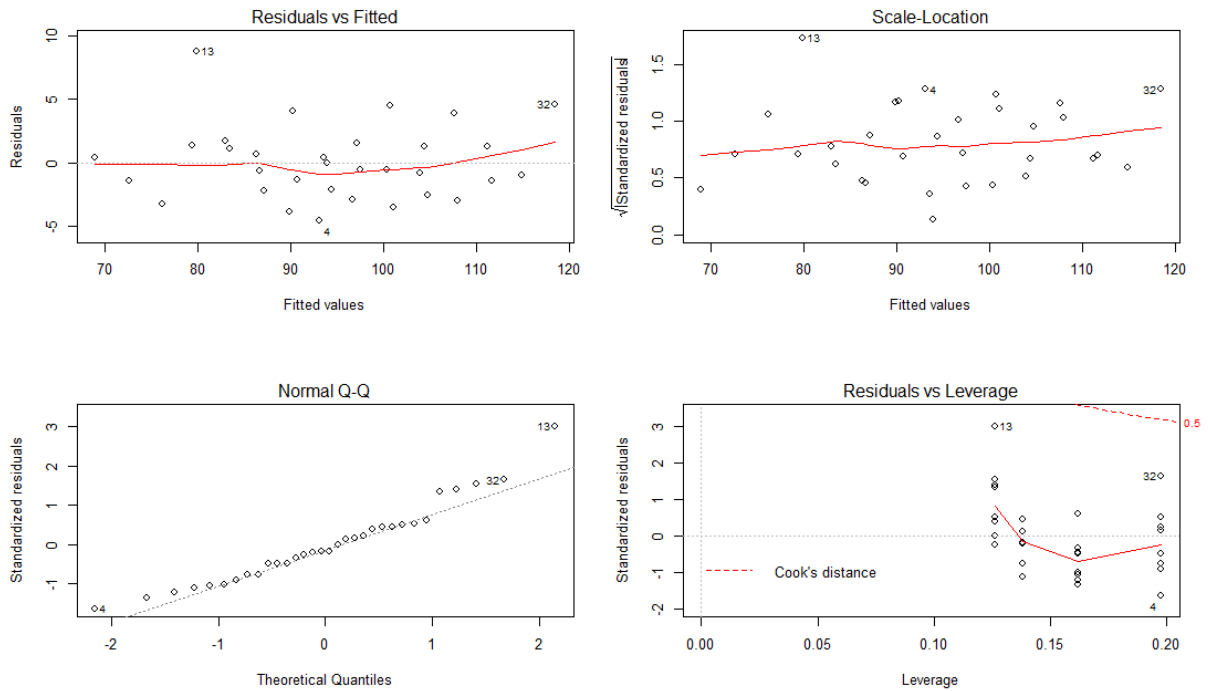
Residual standard error: 3.113 on 27 degrees of freedom
Multiple R-squared:  0.9479,    Adjusted R-squared:  0.9401
F-statistic: 122.7 on 4 and 27 DF,  p-value: < 2.2e-16
```

- 3) Etant donné que la somme des colonnes $S_1+S_2+S_3+S_4 = 1$, la matrice X n'est pas de plein rang et la matrice $({}^tXX)$ n'est pas inversible. Donc on supprime une des 5 colonnes, ici la colonne b.
- 4) Pour pouvoir établir les tests d'hypothèses il faut supposer en plus que les résidus sont gaussiens.
- 5) Le modèle est bon car 94,79% de la variabilité de la série est expliquée par ce modèle. De plus la p-valeur du test de Fisher confirme qu'il y a au moins un coefficient non nul dans le

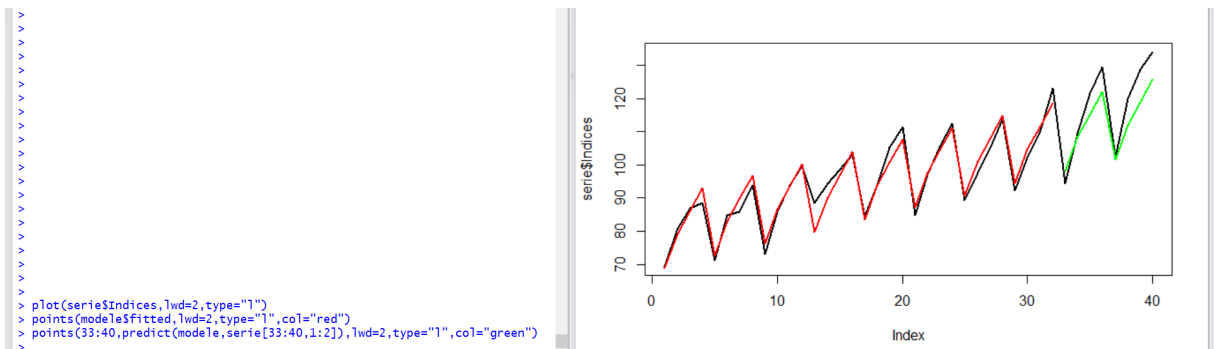
modèle et les tests de Student montrent que tous les coefficients sont significativement non nuls.

On note le coefficient $a > 0$.

- 6) Il n'y a pas de structure dans les résidus. Donc visuellement on peut valider les hypothèses que les résidus sont centrés, de variance constante et non corrélés. Le QQ-plot confirme que les résidus suivent une loi normale (exception aux extrémités). Les résidus standardisés montrent une valeur atypique N°13 mais qui reste tout de même proche de $\sqrt{2}$.



7)



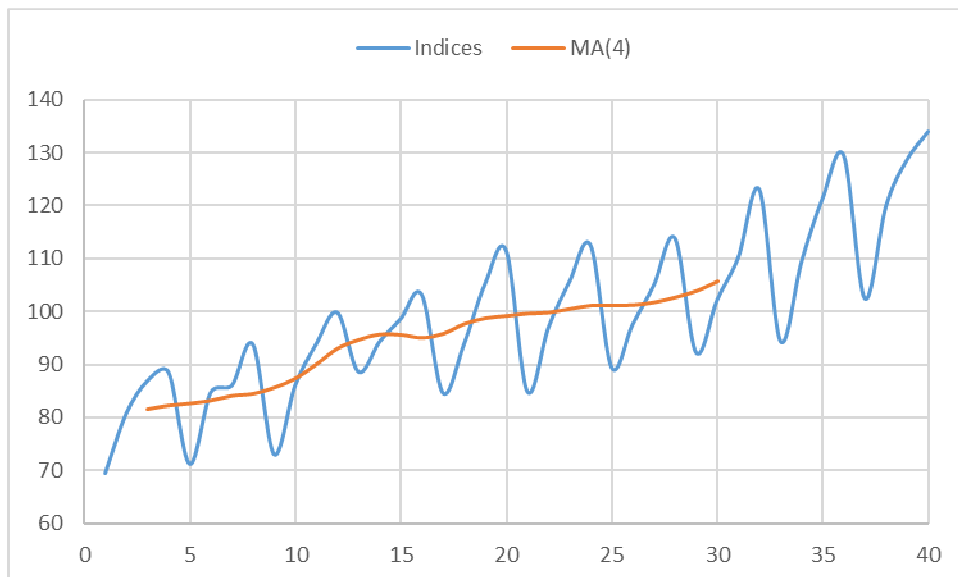
Le modèle reproduit bien la tendance et les saisons sur la base d'apprentissage (jusqu'à $t=32$). En ce qui concerne la base test, on voit que sur la dernière année, il a tendance à sous-estimer les ventes.

III) Lissage par moyenne mobile (sur tableur)

- 1) Lissage de la série **entre 1995 et 2002**
 - a) Quel est l'objectif de lisser la série ?
 - b) Quel ordre allez-vous choisir pour calculer les moyennes mobiles centrées ?
 - c) Sur les données entre 1995 et 2002, à partir de quel temps la série lissée commence et à quel temps finit-elle ?
 - d) Représentez sur un même graphique la série observée et la série lissée. Commentez.
- 2) Dessaisonnalisation de la série **entre 1995 et 2002**
 - a) Calculez les coefficients saisonniers en utilisant les observations entre 1995 et 2002. On calculera les coefficients modifiés de façon à ce qu'ils respectent la condition d'unicité de l'écriture.
 - b) Calculez la série corrigée des variations saisonnières (CSV). Représentez la série observée et la série CSV sur un même graphique entre 1995 et 2002.
 - c) Ajustez la série CVS par une droite en utilisant la méthode des moindres carrés sur le temps pour $t=1$ à $t=32$. Représentez cette droite sur le graphique précédent pour $t=1$ à $t=40$.
- 3) Prévisions **entre 1995 et 2004**
 - a) Ajoutez la composante saisonnière à la droite et représentez la série obtenue pour $t=1$ à $t=40$.
 - b) Commentez et comparez avec les résultats obtenus à la question II.

Correction

- 1)
 - a) Faire disparaître les composantes saisonnières
 - b) On choisit un ordre 4 car les tendances saisonnières sont trimestrielles
 - c) $t=3$ et $t=30$
 - d) Il n'y a plus de composante saisonnière. Il reste une tendance quasiment linéaire croissante.

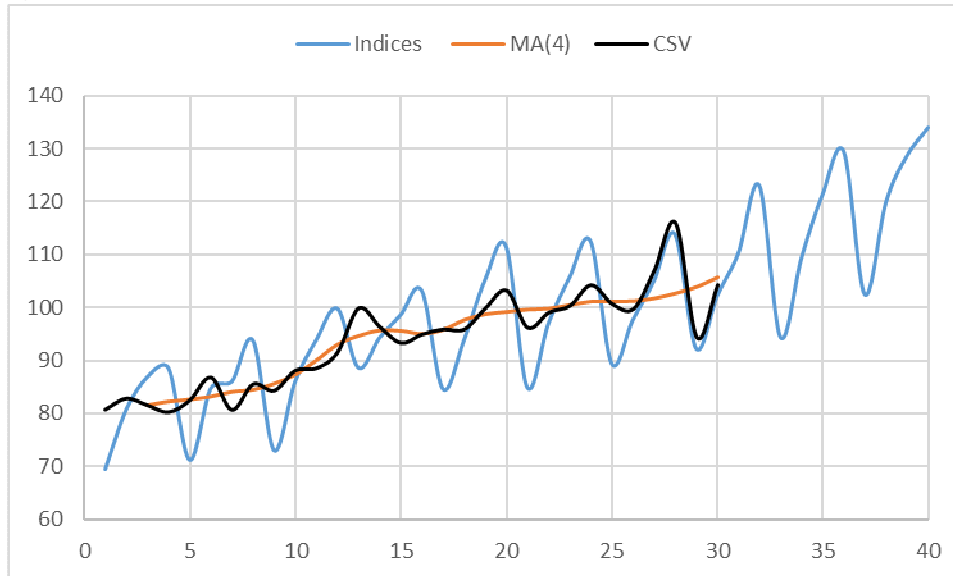


- 2)
 - a)

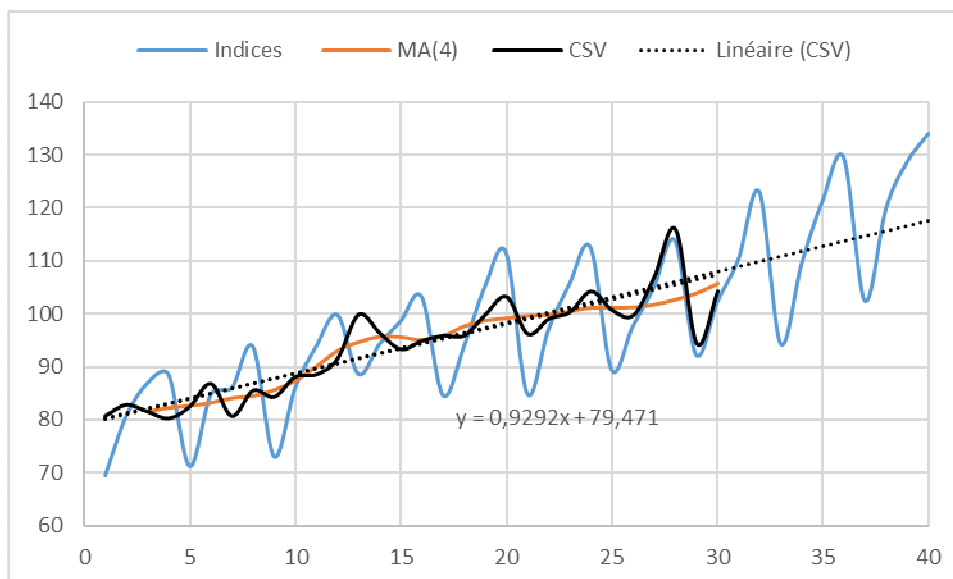
	brut	modifié
c1	-11,43	-11,35
c2	-2,23	-2,15

c3	5,21	5,28
c4	8,14	8,21
moyenne	-0,08	0,00

b)

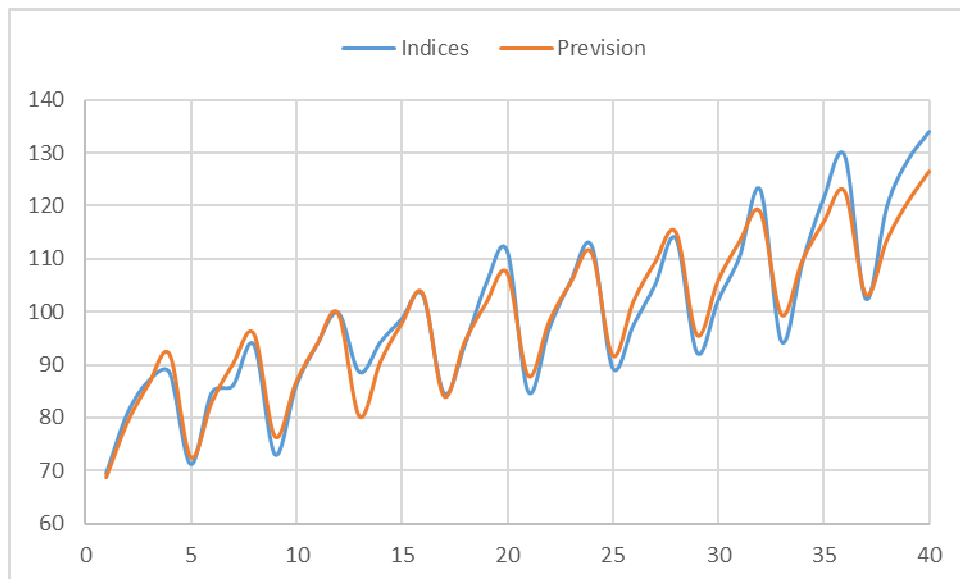


c) On obtient la droite d'équation $y=0.9577x+79,15$



3)

a)



b) On obtient la même prévision que dans la question II

IV) Prévision par la méthode du lissage exponentiel (avec R)

- 1) Transformez la série Indices dans la classe `ts` avec une fréquence correspondant à une période trimestrielle. On pourra créer trois séries, une avec les données entre 1995 et 2002, l'autre entre 2003 et 2004 et une avec toutes les observations.
- 2) Appliquez la fonction `HoltWinters` (avec tendance et composantes saisonnières) sur les données entre 1995 et 2002.
- 3) Comparer les coefficients trouvés ici avec ceux calculés dans la méthode précédente.
- 4) Représentez sur un même graphique la série observée (dans la classe `ts`) et la série obtenue (`fitted(modele)`).
- 5) Calculez les prévisions obtenues avec le modèle de `HoltWinters` entre 2003 et 2004. Ajoutez-les sur le graphique.
- 6) Comparer avec les deux méthodes précédentes.

1)

```
>
>
>
> strain=ts(data = tab$Indices[1:32], start = c(1995,1), frequency = 4)
> strain
      Qtr1  Qtr2  Qtr3  Qtr4
1995 69.40 80.73 86.87 88.50
1996 71.17 84.70 86.00 93.80
1997 73.00 86.00 93.87 99.80
1998 88.60 94.27 98.63 103.20
1999 84.60 93.80 105.23 111.50
2000 84.90 96.97 105.67 112.53
2001 89.40 97.63 105.00 113.87
2002 92.27 102.20 110.27 123.07
> stest=ts(data = tab$Indices[33:40], start = c(2003,1), frequency = 4)
> stest
      Qtr1  Qtr2  Qtr3  Qtr4
2003 94.37 109.47 121.50 129.47
2004 102.47 119.97 128.73 134.07
> |
```

2)

