Université Paris I, Panthéon - Sorbonne

Deuxième Année Master T.I.D.E. 2021 – 2022

Econométrie des séries chronologiques

Examen final, janvier 2022

Examen de 2h00. Tout document ou calculatrice est interdit.

Avec le logiciel R, on simule le nombre de voyageurs urbains (en milliers) d'une agglomération entre janvier 2000 et décembre 2019.

1. On commence par simuler une tendance et une saisonnalité:

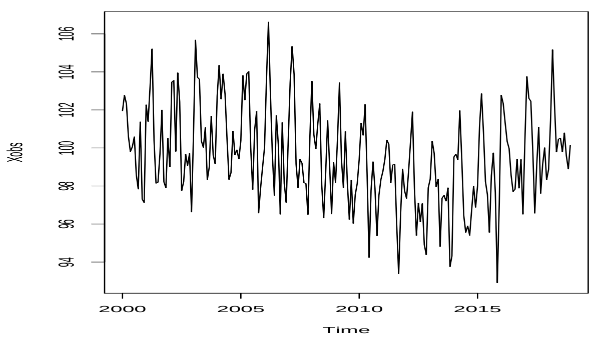
```
n=240; a=0; t=c(1:n); t1=60; t2=162
a[1:t1]=100+0.04*c(1:t1)
a[(t1+1):t2]=a[t1]-0.02*(c((t1+1):t2)-t1)
a[(t2+1):n]=a[t2]+0.06*(c((t2+1):n)-t2)
s=2*sin(t*pi/6)-cos(t*pi/3)
```

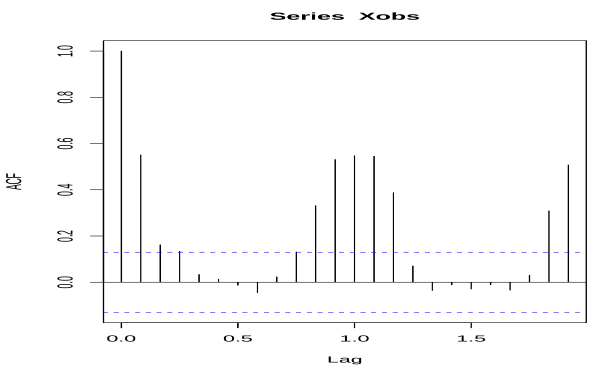
Question 1: *Vérifier que* `s` *est bien une saisonnalité et déterminer sa période. Décrire comment se comporte la tendance* `a` *entre 2010 et 2019.*

2. On définit ensuite les valeurs observées et on suppose inconnue les données de 2019. Ainsi on a ensuite tapé les commandes:

```
u=rep(0,n); epsi=rt((n+2),8)
u=epsi[3:(n+2)]-0.8*epsi[1:(n)]
X=ts(a+s+u,2000,frequency = 12)
Xobs=ts(X[1:228],2000,frequency = 12);
Xnonobs=ts(X[229:240],2019,frequency=12)
ts.plot(Xobs); acf(Xobs)
```

On a ainsi obtenu:





Question 2: *Montrer que* `u[1:n]` *est une trajectoire d'un processus que l'on précisera. Donner l'expression de la tendance et de la saisonnalité de* `Xobs` *en juillet 2015.* `Xobs` *est-il stationnaire? Expliquer son correlogramme.*

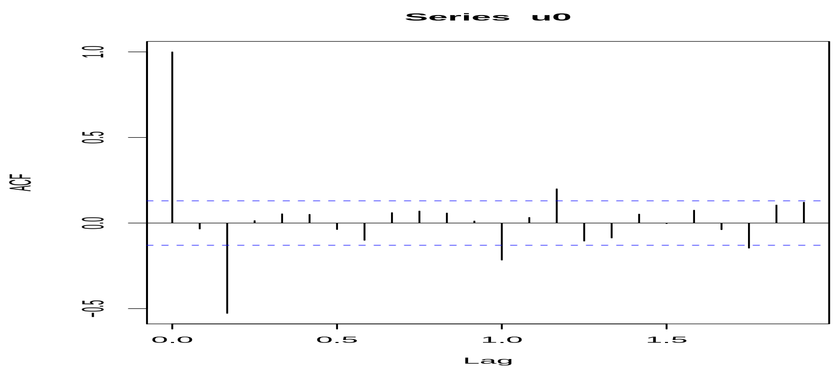
3. On suppose connaître exactement la tendance et la saisonnalité de cette série, ce qui servira de référence pour mesurer l'identification et la prédiction du processus:

```
u0=Xobs-(a[1:228]+s[1:228])
M0=(mean(u0^2))^0.5; M0
library(forecast); acf(u0)
```

```
uest0=auto.arima(u0,max.p=5,max.q=5,ic="bic"); uest0
upred0=predict(uest0,n.ahead=12)
Xpred0=ts(a[229:240]+s[229:240],2019,frequency=12)+upred0$pred
Mpred0=mean((Xnonobs-Xpred0)^2)^0.5; Mpred0
```

On a ainsi obtenu:

```
> M0
[1] 1.453572

> uest0
Series: u0
ARIMA(0,0,2) with zero mean

Coefficients:
          ma1      ma2
      -0.0598  -0.7800
s.e.   0.0488   0.0504

> Mpred0
[1] 0.8524758
```



Question 3: *Expliquer ce que sont les valeurs* `1.453572` *et* `0.8524758`*. Pouvait-on s'attendre à une telle figure? Expliquez en détail et formellement ce qui a été fait pour obtenir* `uest0`*? A quoi correspondent et comment sont calculés* `upred0` *et* `Xpred0`*? Donner les 4 premières valeurs de* `upred0` *et de* `Xpred0`*.*

4. On suppose maintenant que l'on ne connaît pas la tendance et la saisonnalité, mais que l'on connaît les dates de changement et la forme de la tendance:

```
tt1=c(1:t1); tt2=c((t1+1):t2); tt3=c((t2+1):228);
a1=lm(X[tt1]~tt1); a2=lm(X[tt2]~tt2); a3=lm(X[tt3]~tt3)
aest1=c(a1$fit,a2$fit,a3$fit)
Xdet1=Xobs-aest1
sest1=stl(Xdet1,s.win="perio",t.win=1)$time[,"seasonal"]
u1=Xobs-(aest1[1:228]+sest1[1:228])
M1=mean(u1^2)^0.5; M1
uest1=auto.arima(u1,max.p=5,max.q=5,ic="bic"); uest1
upred1=predict(uest1,n.ahead=12)
new=data.frame(tt3=c(229:240))
apred1=predict.lm(a3,newdata=new)
Xpred1=ts(apred1+sest1[1:12],2019,frequency=12)+upred1$pred
Mpred1=mean((Xnonobs-Xpred1)^2)^0.5; Mpred1
```

On a ainsi obtenu:

```
> M1
[1] 1.4425
> uest1
```

```
Series: u1
ARIMA(0,0,2) with zero mean

Coefficients:
         ma1      ma2
      0.1066  -0.6622
s.e.  0.0488   0.0484

> Mpred1
[1] 1.195657
```

Question 4: *Expliquer en détail comment a été obtenu* `aest1`*. Peut-on s'attendre à ce que ce soit un bon estimateur de* `a`*? Dire en détail ce qu'est* `sest1`*. Pourquoi a-t-on choisi la valeur* `t.win=1`*? Expliquer avec des formules comment est calculé et ce qu'est* `apred1`*. Pourquoi utilise-t-on* `sest1[1:12]`*? En argumentant, pouvait-on s'attendre à ce que* `M1<M0`*? Et* `Mpred1>Mpred0`*?*

5. Désormais on ne suppose plus connus les instants de changement, d'où les commandes:

```
Q=matrix(1E20,228,228)
for (k1 in c(2:224))
{ for (k2 in c((k1+2):226))
     {reg1=lm(Xobs[1:k1]~c(1:k1))
      reg2=lm(Xobs[(k1+1):k2]~c((k1+1):k2))
      reg3=lm(Xobs[(k2+1):228]~c((k2+1):228))
      Q[k1,k2]=sum(reg1$res^2)+sum(reg2$res^2)+sum(reg3$res^2)
} }
Indice=which(Q==min(Q),arr.ind=TRUE); Indice
test1=c(1:Indice[1]); test2=c((Indice[1]+1):(Indice[2])); test3=c((Indice[2]+1):228)
ar1=lm(Xobs[test1]~test1); ar2=lm(Xobs[test2]~test2); ar3=lm(Xobs[test3]~test3);
a1est=ar1$fit; a2est=ar2$fit; a3est=ar3$fit; aest2=c(a1est,a2est,a3est)
sest2=stl(Xobs-aest2,s.win="perio",t.win=1)$time[,"seasonal"]
u2=Xobs-(aest2[1:228]+sest2[1:228])
M2=mean(u2^2)^0.5; M2
uest2=auto.arima(u2,max.p=5,max.q=5,ic="bic"); uest2
upred2=predict(uest2,n.ahead=12)
new=data.frame(test3=c(229:240))
apred2=predict.lm(ar3,newdata=new); ar3$coefficients
Xpred2=ts(apred2+sest2[1:12],2019,frequency=12)+upred2$pred
Mpred2=mean((Xnonobs-Xpred2)^2)^0.5; Mpred2
```

On a ainsi obtenu:

```
> Indice
     row col
[1,]  37 168

> M2
[1] 1.448701

> ar3$coefficients
(Intercept)       test3
90.68275409  0.04338298

> Mpred2
[1] 1.35401
```

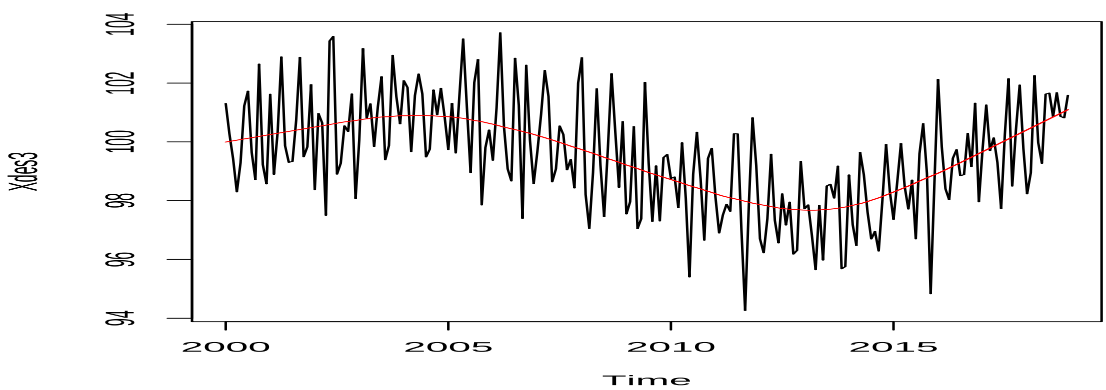
Question 5: *Expliquer ce qui a été fait lors des 9 premières lignes, notamment en écrivant de façon formelle ce qu'est* `Q`*. Pourquoi avait-on choisi* `1E20` *à la première ligne? Que pouvait-on espérer comme valeur pour* `Indice`*? Pouvait-on s'attendre aux valeurs* `90.68275409 0.04338298`*? Commentez les valeurs obtenues pour* `M2` *et* `Mpred2`*.*

6. Pour finir on ne fait plus aucune hypothèses sur la tendance et la saisonnalité:

```
sest3=stl(Xobs,s.win="perio")$time[,"seasonal"]
Xdes3=Xobs-sest3
library(lokern)
kerada3=glkerns(time(Xobs),Xdes3)
ts.plot(Xdes3)
lines(kerada3,col='red')
aest3=ksmooth(time(Xobs),Xdes3,"normal",bandwidth=kerada3$bandwidth)
sest3=stl(Xobs-aest3$y,s.win="perio")$time[,"seasonal"]
u3=Xobs-(aest3$y[1:228]+sest3[1:228])
M3=mean(u3^2)^0.5; M3
uest3=auto.arima(u3,max.p=5,max.q=5,ic="bic"); uest3
Box.test(uest3$res^2,lag=5)
upred3=predict(uest3,n.ahead=12)
Holt3=HoltWinters(aest3$y[1:228],gamma=FALSE)
apred3=predict(Holt3,n.ahead = 12)
Xpred3=ts(apred3+sest3[1:12],2019,frequency=12)+upred3$pred
Mpred3=mean((Xnonobs-Xpred3)^2)^0.5; Mpred3
```

et voici les résultats numériques obtenus:

```
> M3
[1] 1.43731
> uest3
Series: u3
ARIMA(0,0,2) with zero mean

Coefficients:
         ma1      ma2
      0.0701  -0.6884
s.e.  0.0496   0.0476

> Box.test(uest3$res^2,lag=5)

Box-Pierce test

data:  uest3$res^2
X-squared = 4.4501, df = 5, p-value = 0.4866

> Mpred3
[1] 1.281921
```



Question 6: *Expliquer en détail ce qui a été fait lors des 8 premières lignes (méthodes utilisées notamment). Pourquoi calcule-t-on 2 fois* `sest3`*? Expliquer en détail la commande* `Box.test...` *et les conclusions que l'on en peut tirer. Décrire ce qu'est* `Holt3`*. Pourquoi ne pas avoir utilisé* `aest3`*? Au final, quelle méthode utiliseriez vous pour prédire le comportement en 2019?*