Modèles de Prévision.TD1-TP1.

Exercice 1. Soit $\{\epsilon_t\}$ une suite de variables aléatoires indépendantes distribuées selon une loi normale de moyenne 0 et de variance σ^2, et soit a, b et c des constantes. Déterminer lequel ou lesquels des processus ci-dessous sont stationnaires. Pour chaque processus stationnaire, calculer la moyenne et la fonction d'autocovariance.

a) $X_t = a + b\epsilon_t + c\epsilon_{t-2}$

d) $X_t = a + b\epsilon_0$

e) $X_t = \epsilon_t \epsilon_{t-1}$

Exercice 2. Soit $\{X_t\}$ une série chronologique stationnaire de moyenne nulle et a, b, des constantes. :Si $Y_t = a + bt + s_t + X_t$, où s_t est une composante de saisonnalité de période 12, démontrer que $\Delta\Delta_{12}Y_t = (1-B)(1-B^{12})Y_t$ est stationnaire.

Exercice 3 (Données du niveau du Lac Huron). Les données lake.dat donnent le niveau du Lac Huron moins 570 pieds entre les années 1875 et 1972. L'ensemble contient donc 98 données. Faire une modélisation préliminaire (à être complétée plus tard) de cette série en suivant les étapes suivantes.

i) Tracer le graphique de la série et identifier visuellement une tendance et/ou de la saisonnalité.

ii) Si nécessaire, estimer la tendance par régression et éliminer la saisonnalité à l'aide de différences.

iii) Proposer un modèle pour les résidus obtenus en ii). Justifier votre réponse à l'aide du corrélogramme des résidus et des résultats des tests de détection du bruit blanc.

Exercice 4 (Données de varicelle). Le fichier varicelle.dat contenant le nombre de cas de varicelle relevés à New-York de janvier 1931 à juin 1972

(1) Créer un objet de type série temporelle contenant cette série. Représenter graphiquement la série.

(2) Analyser qualitativement cette série, c'est-à-dire repérer d'éventuelles tendances et/ou saisonnalités (changer d'échelle si besoin).

(3) Quel est le nombre de cas de varicelle mensuel moyen?

(4) Tracer les 25 premières auto-corrélations. Interpréter ces résultats.

(5) Tracer sur un graphique l'évolution annuelle du nombre de cas de varicelle. Que Constatez-vous ?

```
#Code pour l'évolution annuelle du nombre de cas de varicelle :
v=c();
for  (i  in  1:41)
{
```

```
cp<-window(serie,start=c(i,1),end=c(i,12))
v=c(v,sum(cp))
v=ts(v)
}
plot(1:41,v)
plot.ts(1:41,v)
```

Exercice 5 (Simulations de séries temporelles).

(1) Quelle est la fonction d'auto-corrélation d'un bruit blanc?

(2) Simuler un bruit blanc gaussien de taille 100, et représenter le graphiquement.

(3) Tracer la fonction d'auto-corrélation.

(4) Recommencer les deux questions précédentes et observer la variabilité des résultats. Jouer sur la longueur de la série.

(5) Simuler maintenant la série temporelle $X(t) = 0.5t + 2\epsilon_t$ avec ϵ_t suit $\mathcal{N}(0, 1)$ (taille 100).

(6) Représenter graphiquement la série et interpréter-la qualitativement.

(7) Faites de même pour $X(t) = 0.5t + \epsilon_t + 3\cos(t\pi)$ avec ϵ_t suit $\mathcal{N}(0, 1)$.

Exercice 6. (Op erateur différence).

Téléchargez le fichier de données

http://w3.mi.parisdescartes.fr/ vperduca/data/PNBTrimestrielUSenMillionsDol.dat

On y trouve les chiffres trimestriels du PNB des USA, en milliers de milliards de dollars, de 1954 à 1987. On considère, comme souvent en économie, non le processus du PNB mais celui de son log, que l'on note X_t, pour $t = 1, \dots, T$ (X_1 correspond au premier trimestre, X_2 au second, ..., etc)

1. Visualiser la trajectoire de X_t

2. Pour chaque $h \in \{1, \dots, 20\}$, calculer la corrélation empirique $\widehat{\gamma}_{X,T}(h)$ de X_t et X_{t+h} sur les données $X_1, \dots, X_T$. Expliquer pourquoi on ne peut pas modéliser les X_t par des v.a. indépendantes

3. On introduit alors $Y_t = X_t - X_{t-1}$. Visualiser une trajectoire de Y, ainsi que sa corrélation empirique $\widehat{\gamma}_Y(h)$, en fonction de $h \in \{1, \dots, 20\}$. Expliquer pourquoi on ne peut pas modéliser les Y_t par des v.a. indépendantes. Quel avantage présente néanmoins le processus Y sur X?

Exercice 7 (Danone test de blancheur)

Test de la blancheur du rendement de l'action Danone et de celle de son carré. On suivra les étapes suivantes :

1. Calculer le carré du rendement centré.

2. Tester la blancheur du rendement sur toute la série (on pourra tester la blancheur aux retards 3, 6, 9 et 12). On obtient un résultat inattendu. Lequel?

3. Après examen du chronogramme du rendement (fig. 1.4, chap. 1) on décide de se limiter à l'étude de la série des 600 premières valeurs. Pourquoi ce choix? Qu'a-t-on observé sur le chronogramme?

4. Tester la blancheur du rendement et du rendement centré au carré sur la série de ces 600 valeurs. Conclusion?

5. Au vu de ces résultats, le rendement peut-il être un bruit blanc gaussien?

Code pour calculer rendement:

```
> require("caschrono")
> require("timeSeries")
> data("csdl")
> aa = returns(csdl, percentage = TRUE)
> aab = aa[complete.cases(aa) == TRUE,]
> # in previous version we use package its which will not be maintained anymo
> # r.csdl = its(aab, as.POSIXct(row.names(aab)))
> r.csdl = zoo(aab, as.POSIXct(row.names(aab)))
> r.danone= r.csdl[,3]
> rdt2 = (r.danone-mean(r.danone))^2
> a0=Box.test.2(r.danone, nlag=c(3,6,9,12),
+  type="Ljung-Box",decim=4)
> a1=Box.test.2(r.danone[1:600], nlag=c(3,6,9,12),
+  type="Ljung-Box",decim=4)
> a2=Box.test.2(rdt2[1:600], nlag=c(3,6,9,12),
+  type="Ljung-Box",decim=4)
> a12 = cbind(a0,a1[,2],a2[,2])
> colnames(a12)= c("Retard","p-val. serie compl.",
+  "p-val. 600 obs.", "p-val. rdt carre")
> require("xtable")
> xtable(a12, caption="Table a12 : test de blancheur du rendement de Danone et de son carre.",
+  label="danoblanc",
+ digits=4)
```