

Séries temporelles

Jean Gaudart

Aix-Marseille Université

UMR 912, SESSTIM (AMU, INSERM, IRD)

plan

1. Introduction
2. Analyse graphique
3. Décomposition d'une série
4. Corrélation
5. Cofacteur
6. Lissage exponentiel
7. Modèles Stochastiques
8. Régression
9. Séries stationnaires = ARMA
10. Séries non-stationnaires = ARIMA

VII. Modèles stochastiques

VII.1 Bruit Blanc

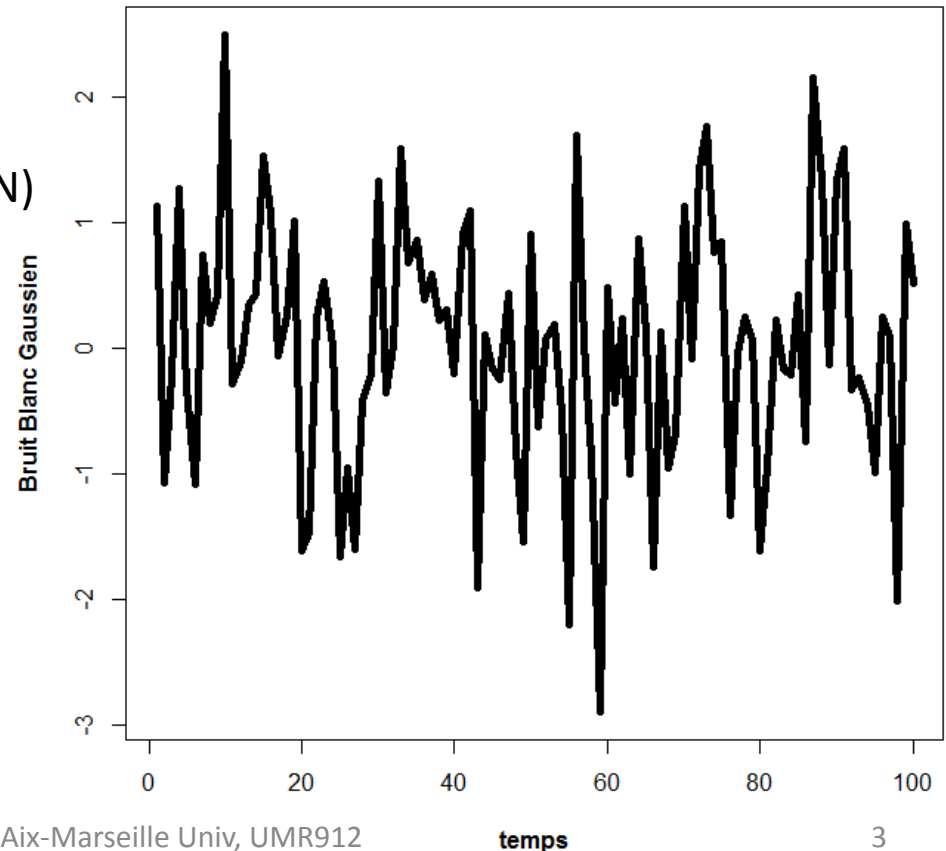
Une série temporelle $\{\varepsilon_t, t=1\dots n\}$ est un bruit blanc discret (Discret White Noise) si les variables $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$, sont Indépendantes, Identiquement Distribuées (iid) de moyenne 0.

\Rightarrow variance constante σ^2

\Rightarrow Correlation $\rho(\varepsilon_i, \varepsilon_j)=0, \forall i \neq j$

Si $\varepsilon_t \rightarrow N(0, \sigma^2) \Rightarrow$ Bruit Blanc Gaussien (GWN)

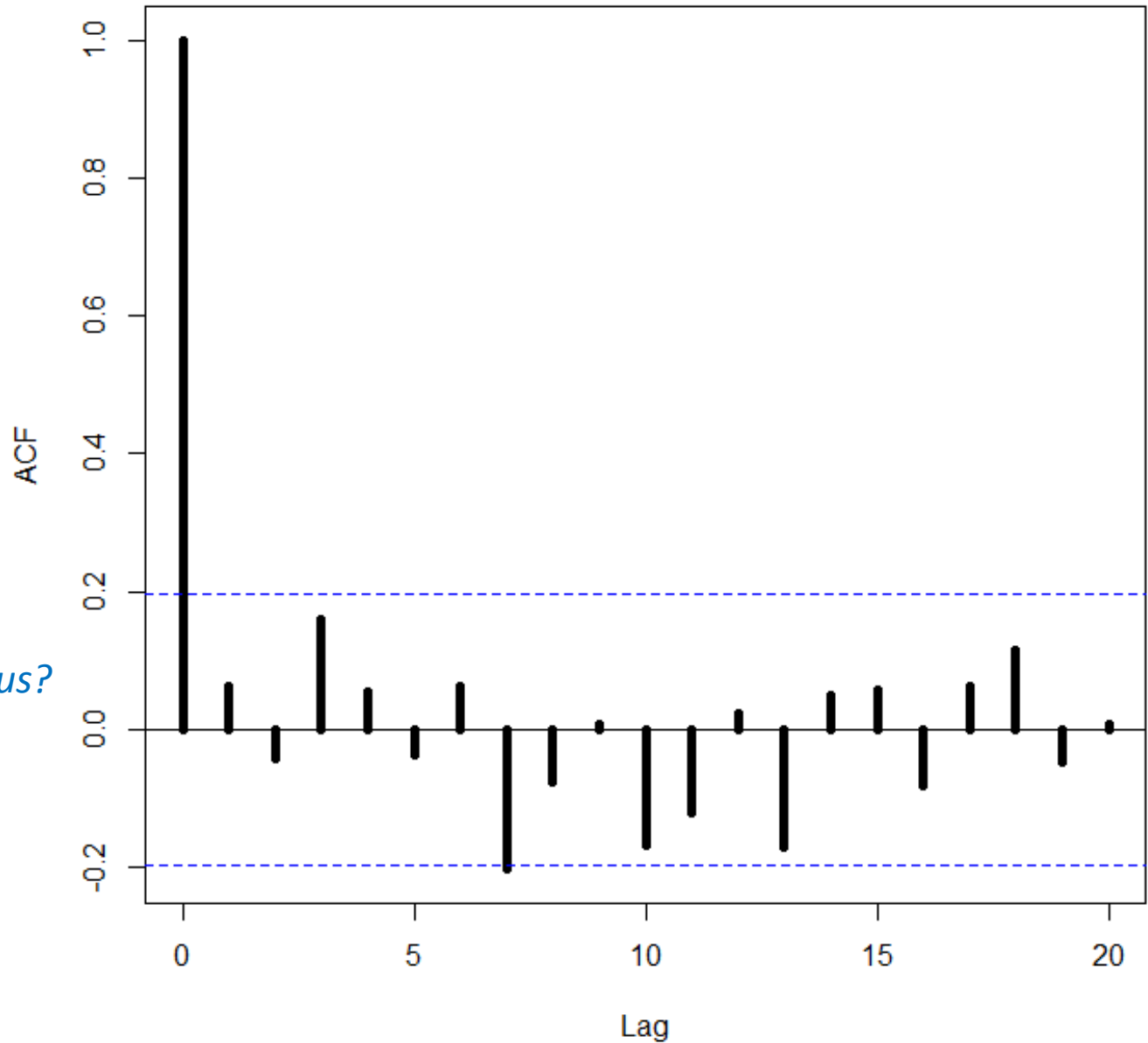
```
e<-rnorm(100)
e.ts<-ts(e)
plot(e.ts)
```



Series e.ts

acf(e.ts)

?Que remarquez vous?



VII.2 Marche Aléatoire (Random Walk)

Une série temporelle $\{y_t, t=1\dots n\}$ est une marche aléatoire si

$$y_t = y_{t-1} + \varepsilon_t$$

En substituant

$$y_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \dots$$

On définit l'opérateur retard \mathbf{B} tel que $\mathbf{B}y_t = y_{t-1}$ ainsi $\mathbf{B}^n y_t = y_{t-n}$

$$\Leftrightarrow y_t = \mathbf{B}y_t + \varepsilon_t$$

Propriétés:

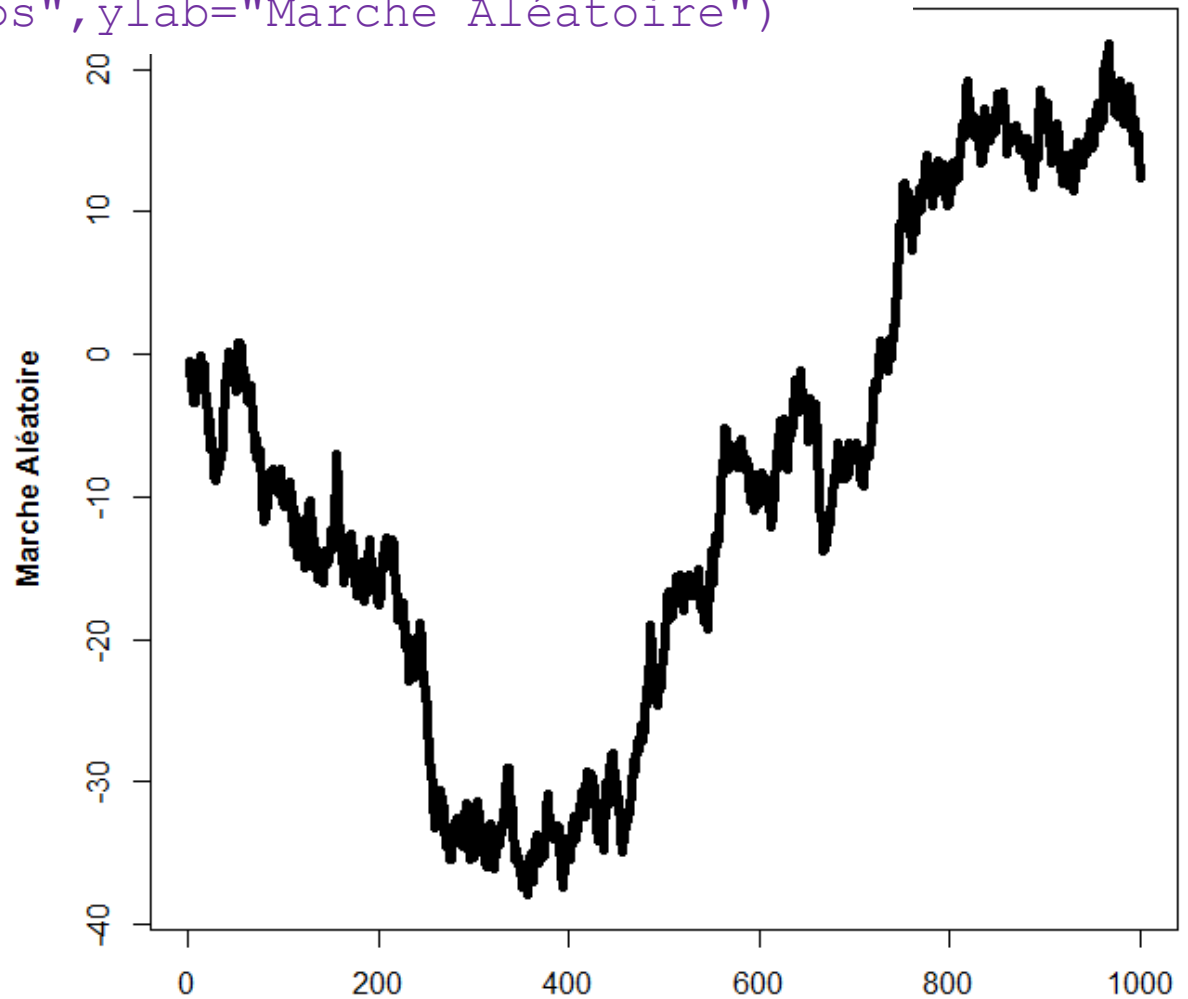
$$\begin{cases} \mu_y = 0 \\ \gamma_k(t) = \text{Cov}(y_t, y_{t+k}) = t\sigma^2 \Leftrightarrow \rho_k(t) = \frac{1}{\sqrt{1+k/t}} \end{cases}$$

? Le processus est-il stationnaire?

Simulation d'une Marche Aléatoire:

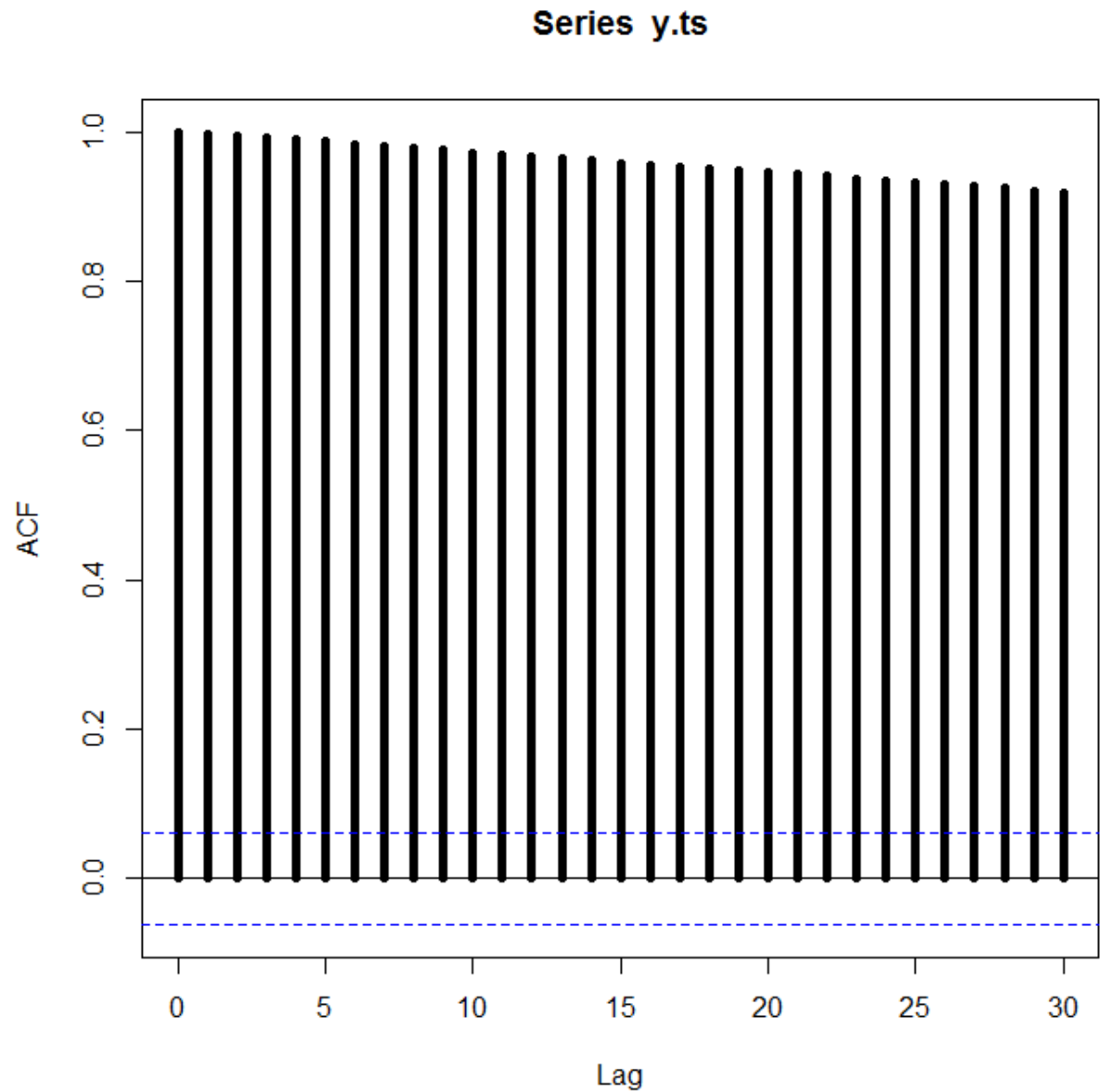
```
y<-e<-rnorm(1000)
for(t in 2:1000){y[t]<-y[t-1]+e[t]}
y.ts<-ts(y)
plot(y.ts,xlab="Temps",ylab="Marche Aléatoire")
```

? *Que remarquez vous?*



`acf(y.ts)`

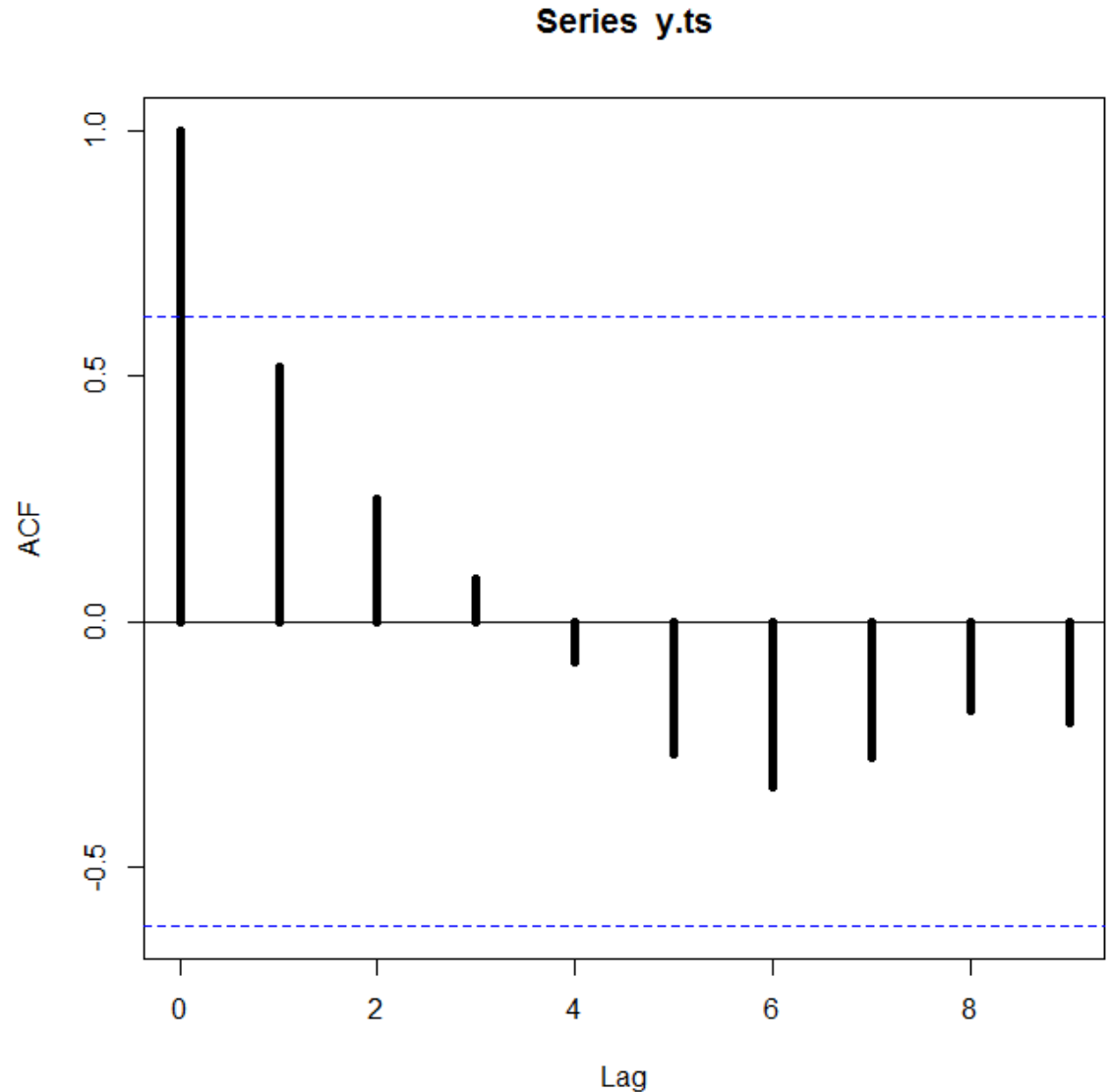
? *Interprétez?*



? Essayez sur une série plus courte $t=10$?

? Essayez sur une série plus courte $t=10$?

?pourquoi une telle différence
séries longue/série courte?



Opérateur différence $\nabla^{(1)} y_t = y_t - y_{t-1}$

$$\Leftrightarrow \nabla^{(1)} y_t = (1 - B)y_t$$

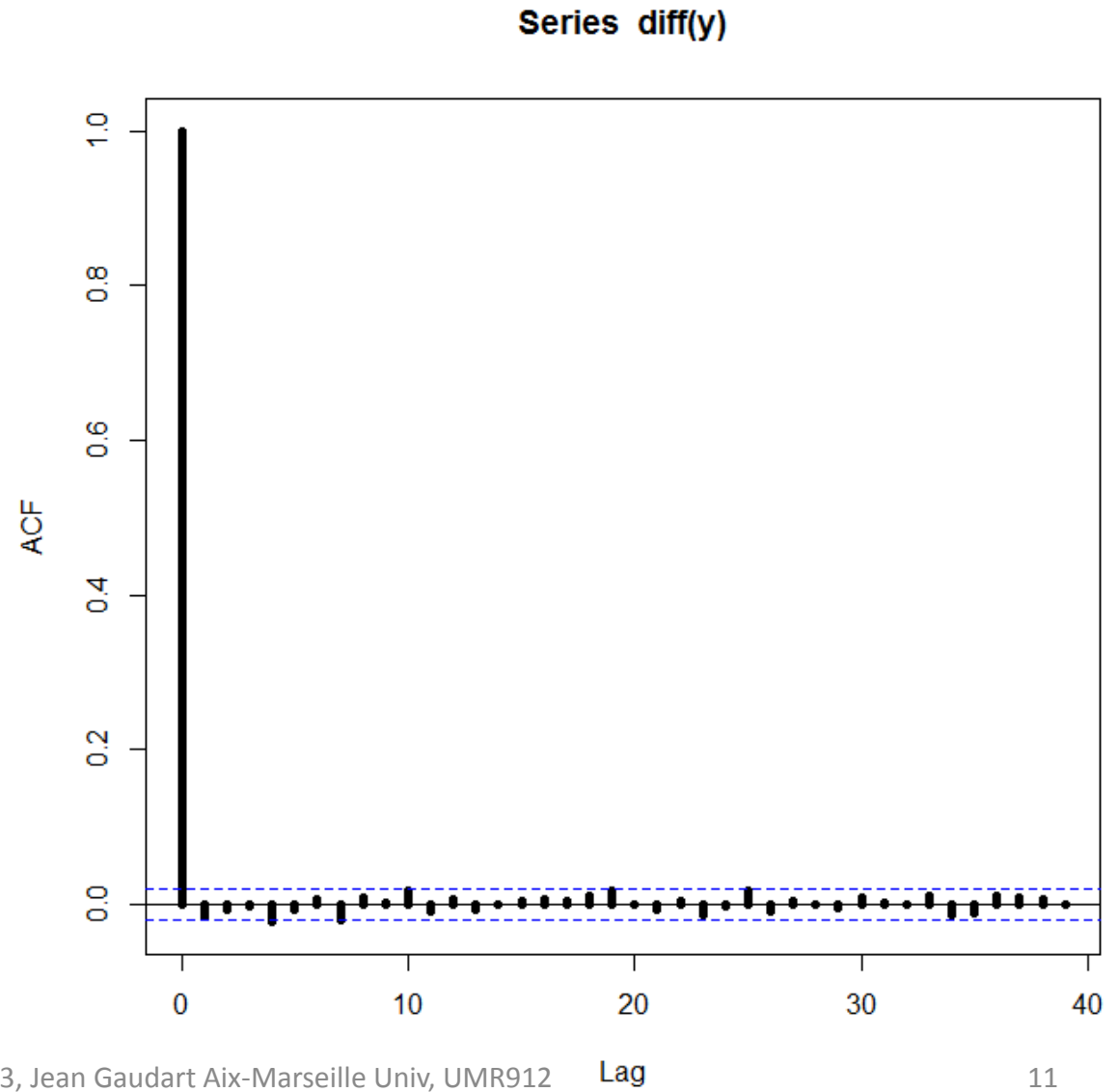
$$\Leftrightarrow \nabla^{(n)} = (1 - B)^n$$

Différence du 1^{er} ordre appliquée à la marche aléatoire précédente

$$\nabla^{(1)} y_t = y_t - y_{t-1} = \varepsilon_t = \text{Bruit Blanc}$$

```
acf(diff(y))
```

? *Que remarquez vous?*



VII.3 Modèle Auto-régressif

Une série temporelle $\{y_t, t=1\dots n\}$ est un processus autorégressif AR(p) si:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \varepsilon_t$$

$$\Leftrightarrow \theta_p(\mathbf{B})y_t = \left(1 - \alpha_1 \mathbf{B} - \alpha_2 \mathbf{B}^2 - \dots - \alpha_p \mathbf{B}^p\right)y_t = \varepsilon_t$$

Remarques:

- Une marche aléatoire est un AR(1) avec $\alpha_1=1$
- Le modèle de lissage exponentiel est un cas particulier avec $\alpha_i = \alpha(1-\alpha)^i$ et $p \rightarrow \infty$
- Le modèle est une régression de y_t sur le passé
- Prédiction à t : $\hat{y}_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p}$
- Estimation des paramètres : minimiser SSE

Propriétés d'un AR(1):

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t$$

$$\begin{cases} \mu_y = 0 \\ \gamma_k = \frac{\alpha^k \sigma^2}{1 - \alpha^2} \end{cases} \Leftrightarrow \rho_k = \alpha^k \quad \text{avec } |\alpha| < 1 \text{ et } k \geq 0$$

Corrélogramme: décroissance rapide à 0, d'autant plus rapide que α petit

Mais par définition: ρ_k jamais nul $\forall k$,

Or, y_t ne dépend que de y_{t-1} !

\Rightarrow Autocorrelation partielle (pacf):

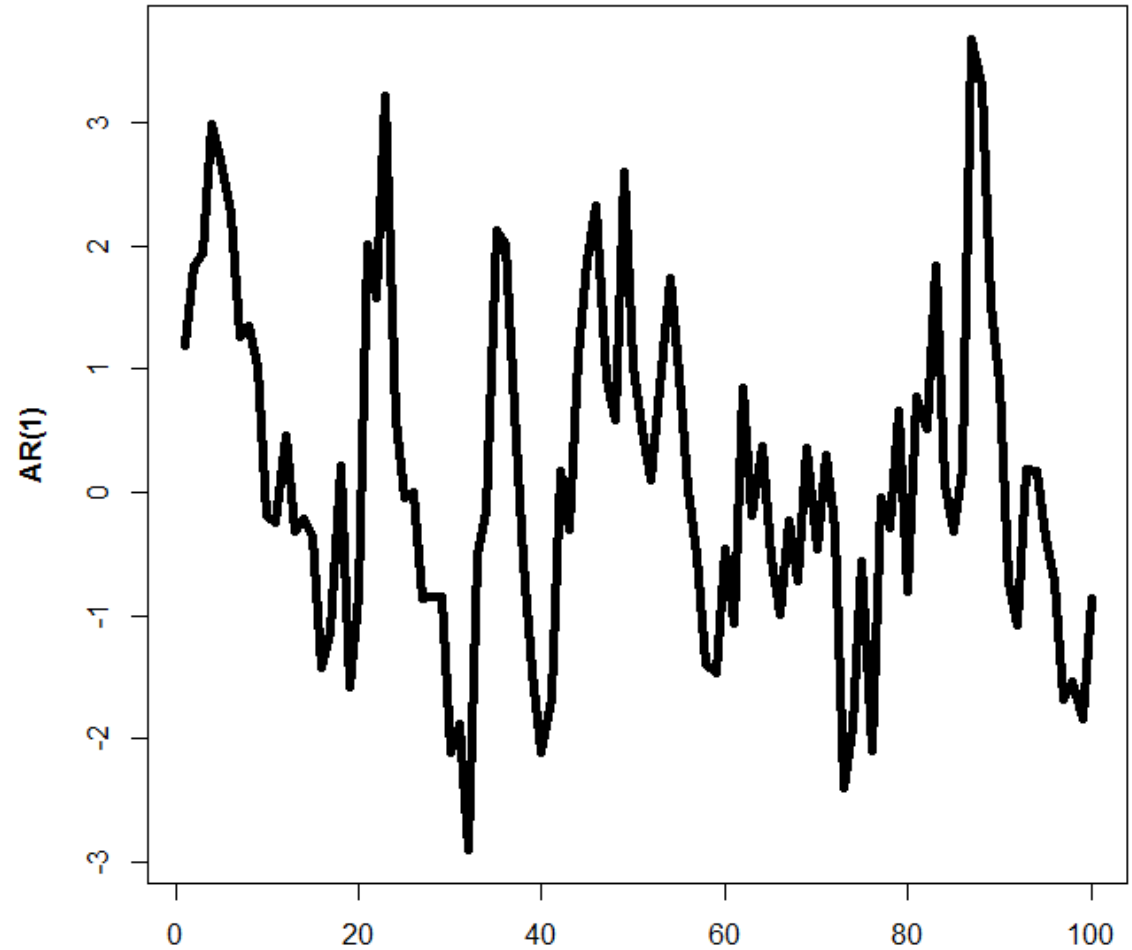
Corrélation après avoir enlevé les corrélations à des lags plus petits

\Rightarrow pacf(AR(p), $k > p$) = 0

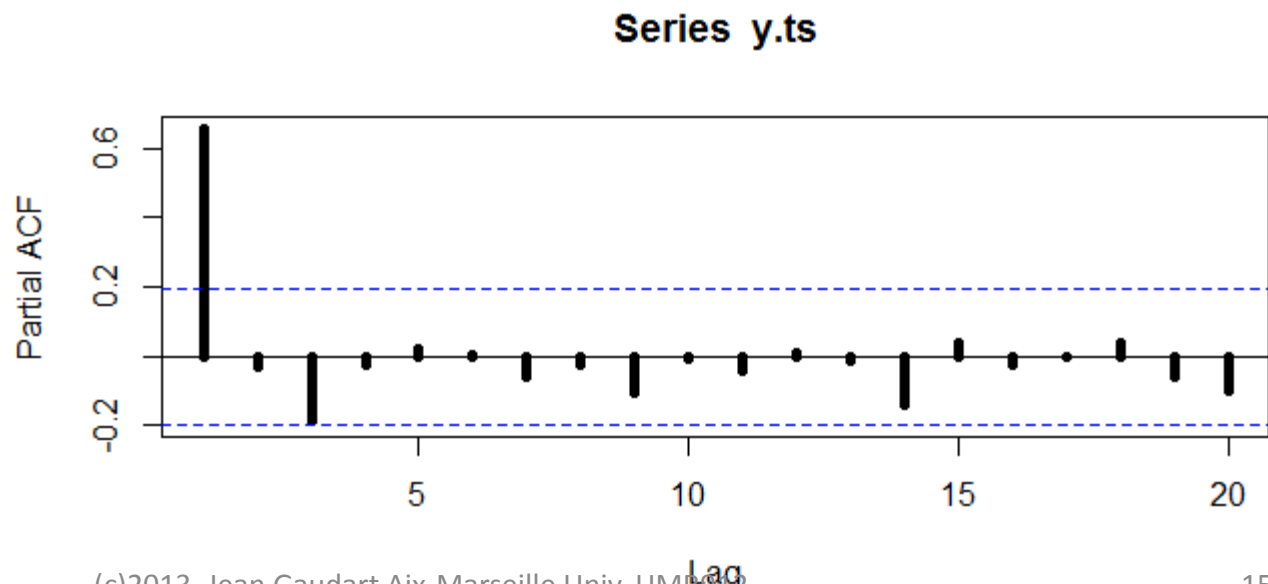
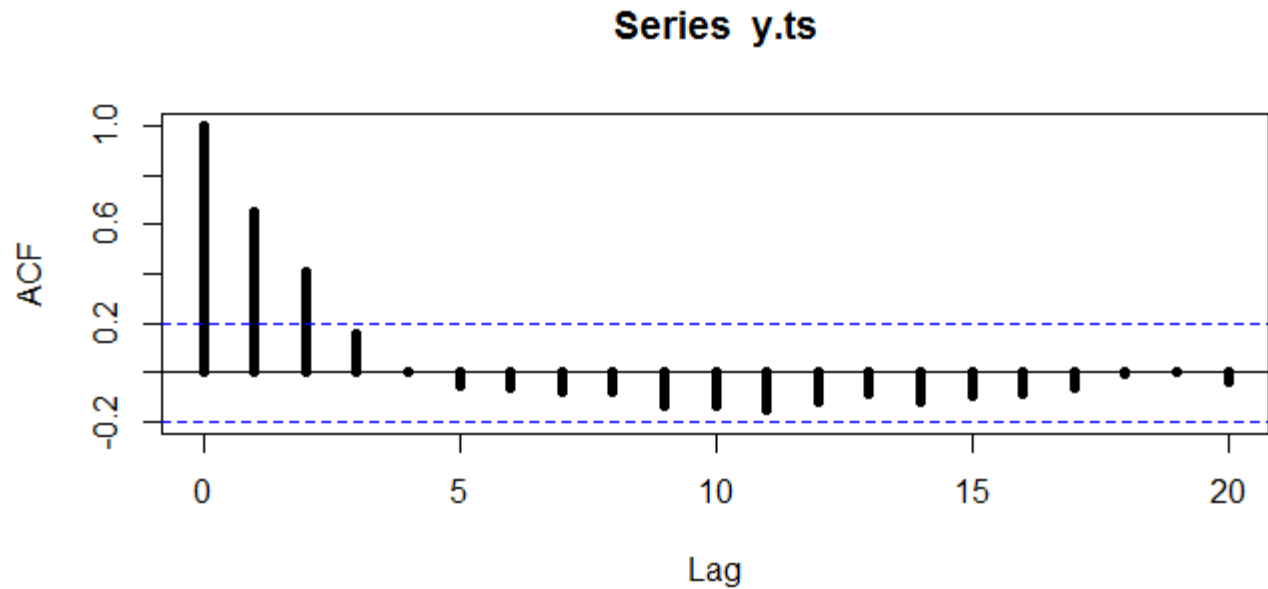
\Rightarrow Autocorrélogramme partiel utile pour déterminer l'ordre de l'AR(p)

Simulation AR(1):

```
y<-e<-rnorm(100)
for (t in 2:100){y[t]<-0.7*y[t-1]+e[t]}
y.ts<-ts(y)
plot(y.ts)
```



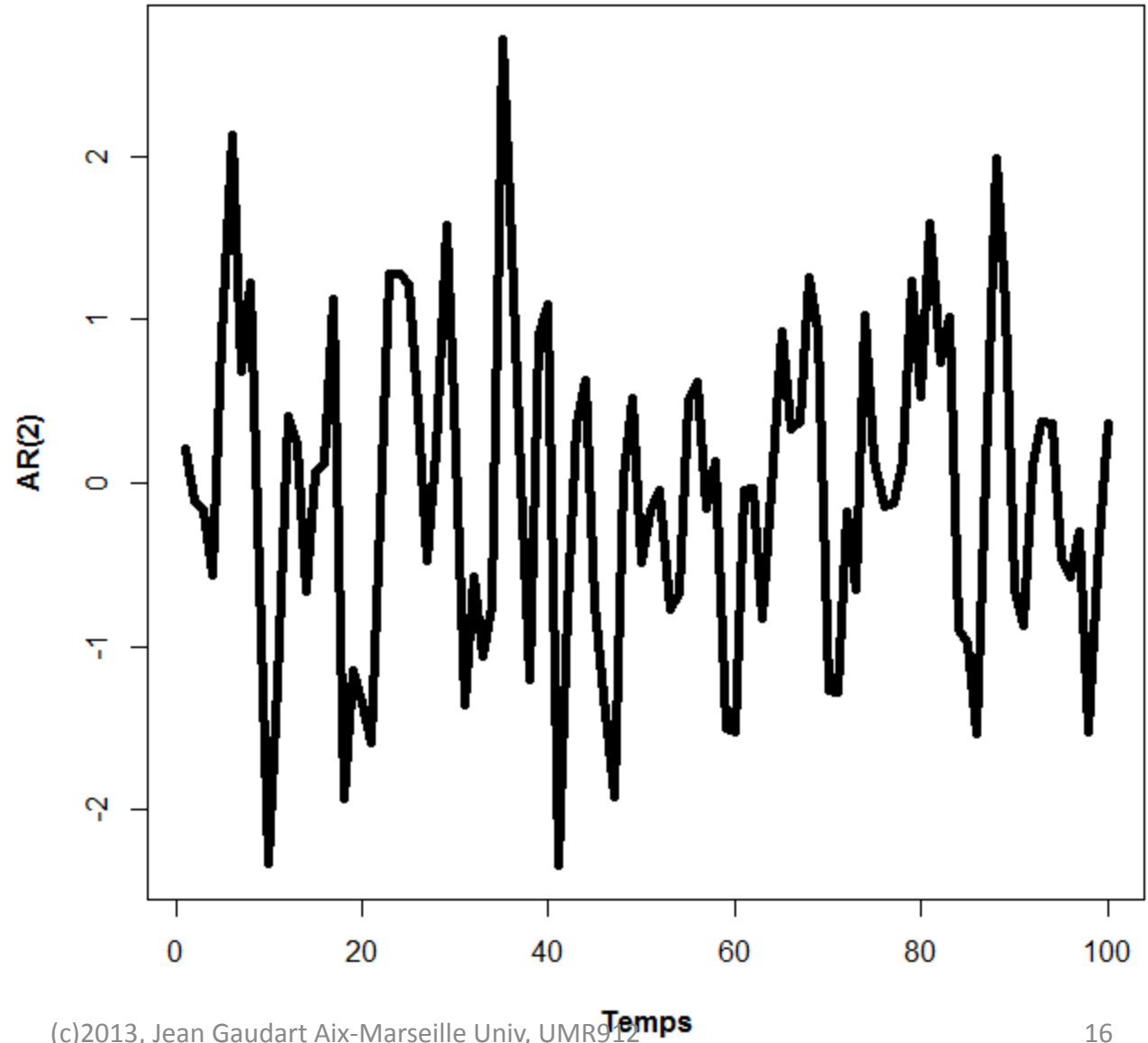
```
layout(1:2)
acf(y.ts)
pacf(y.ts)
```



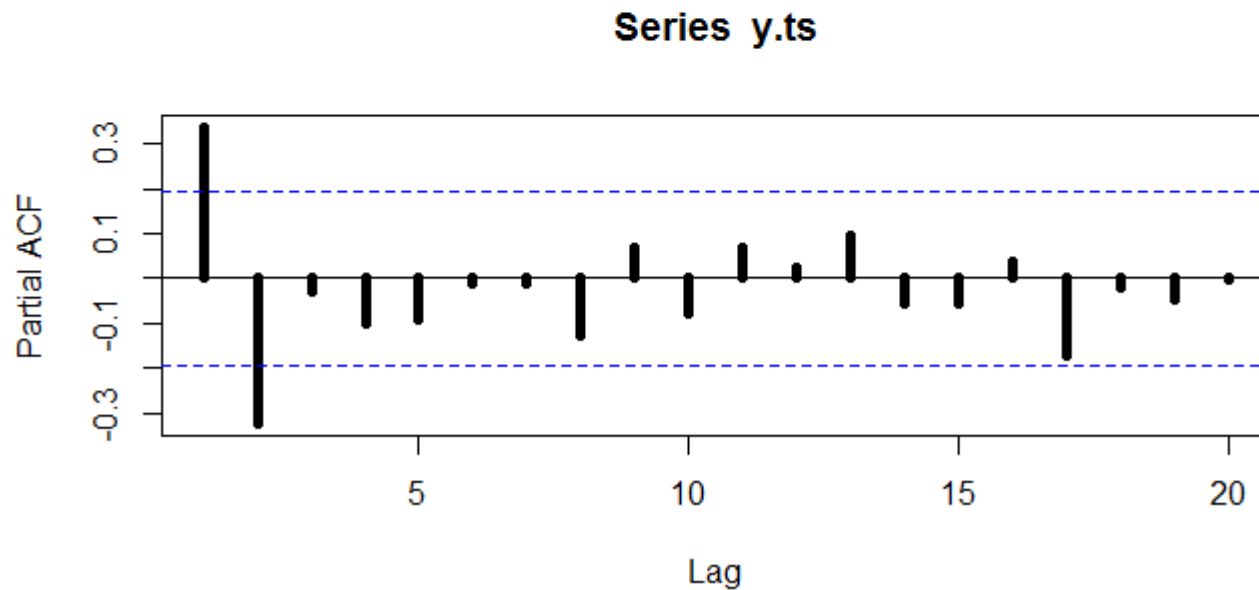
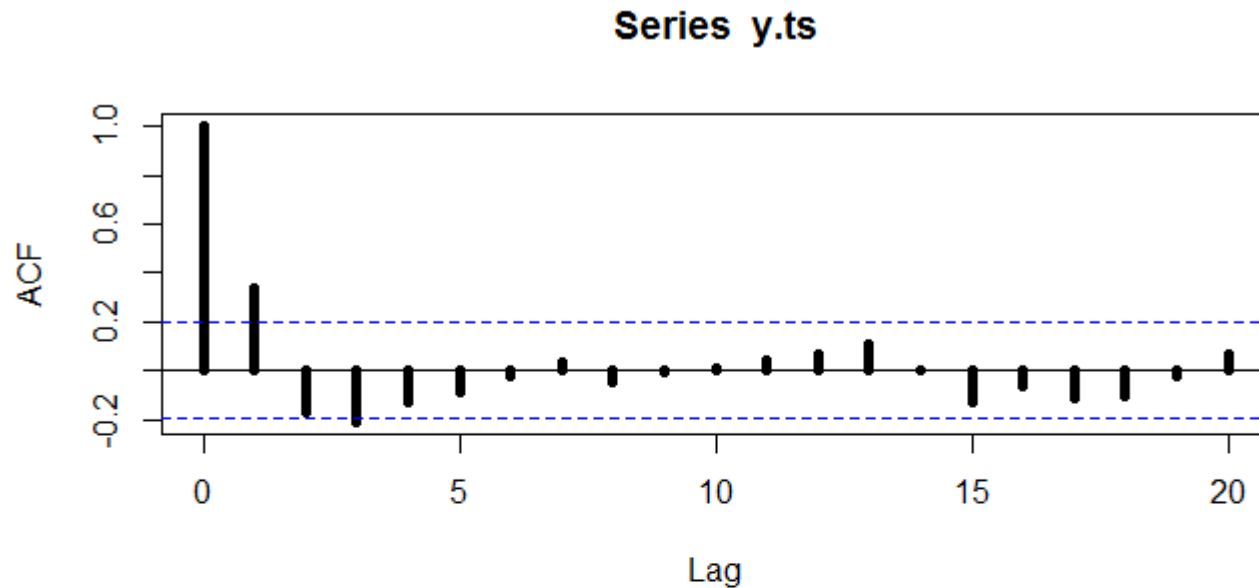
ATTENTION:
Acf débute à k=0
Pacf débute à k=1

Simulation AR(2):

```
y<-e<-rnorm(100)
for (t in 3:100){y[t]<-0.5*y[t-1]-0.2*y[t-2]+e[t]}
y.ts<-ts(y)
plot(y.ts)
```




```
layout(1:2)
acf(y.ts)
pacf(y.ts)
```



ATTENTION:
Acf débute à k=0
Pacf débute à k=1

Modélisation d'une série AR(p)

Simulation AR(1):

```
y<-e<-rnorm(100)
for (t in 2:100){y[t]<-0.7*y[t-1]+e[t]}
y.ts<-ts(y)
```

```
y.ar<-ar(y.ts,method="mle")
```

ordre

```
y.ar$order [1] 1
```

paramètre

```
y.ar$ar [1] 0.7025421
```

Intervalle de confiance à 95%

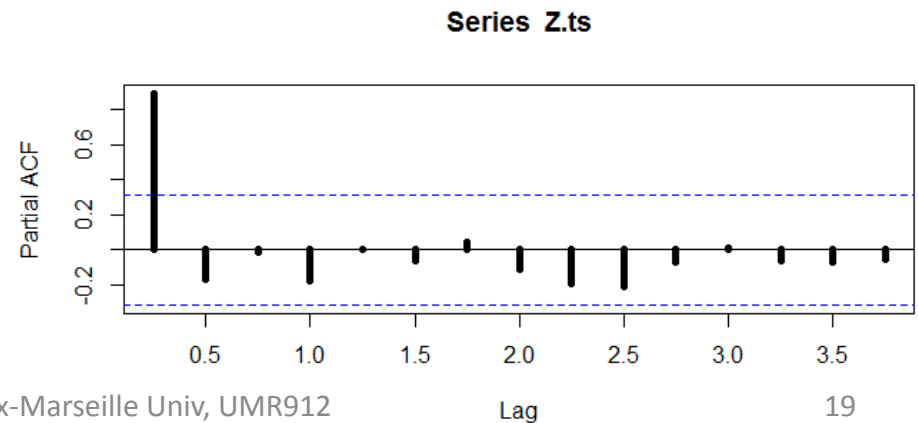
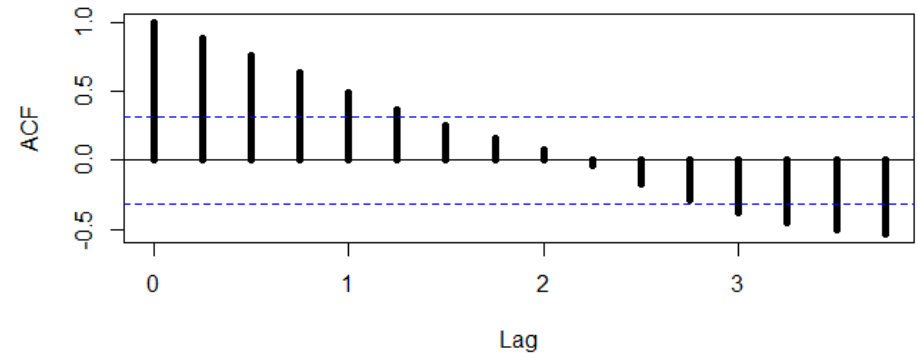
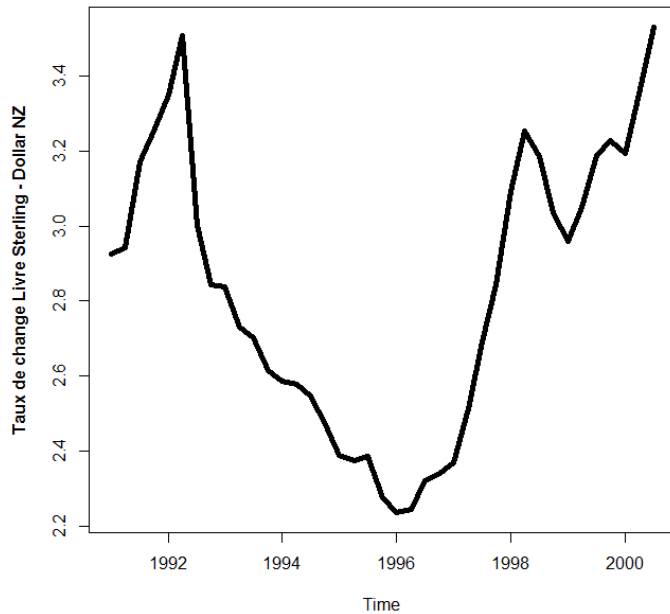
```
y.ar$ar+c(-1.96,1.96)*sqrt(y.ar$asy.var) [1] 0.564 0.841
```

Reprise Ex. 5: Taux de change £GBP - \$NZ

```
Z<- read.table("C:/Mon_dossier/pounds_nz.dat",header=TRUE)
Z.ts<-ts(Z, start=1991,freq=4)
plot(Z.ts)
```

```
mean(Z.ts)      [1] 2.823251
```

```
layout(1:2)
acf(Z.ts)
pacf(Z.ts)      xrate
```



```

Z.ar<-ar(Z.ts)
Z.ar$order                                [1] 1
Z.ar$ar                                    [1] 0.890261
Z.ar$ar+c(-1.96,1.96)*sqrt(Z.ar$asy.var) [1] 0.7435 1.037

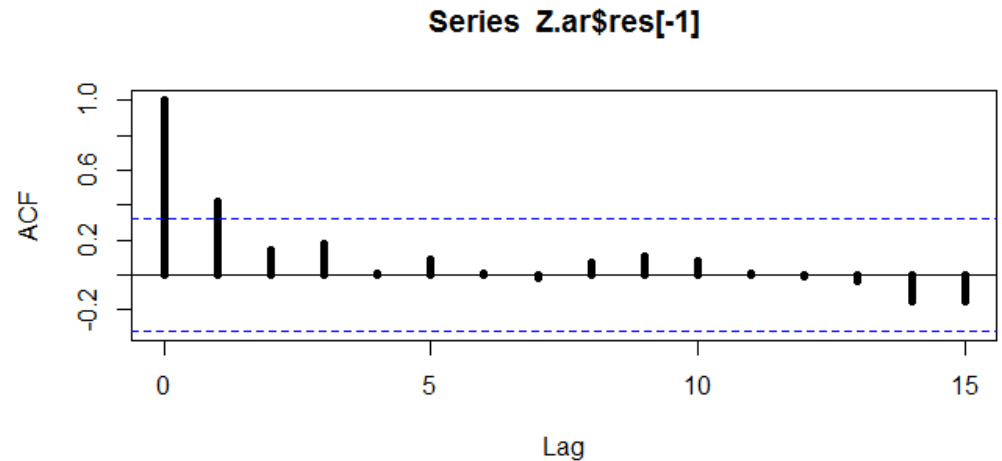
```

$$\hat{z}_t = 2,8 + 0,89(z_{t-1} - 2,8)$$

```

layout(1:2)
acf(Z.ar$res[-1])
pacf(Z.ar$res[-1])

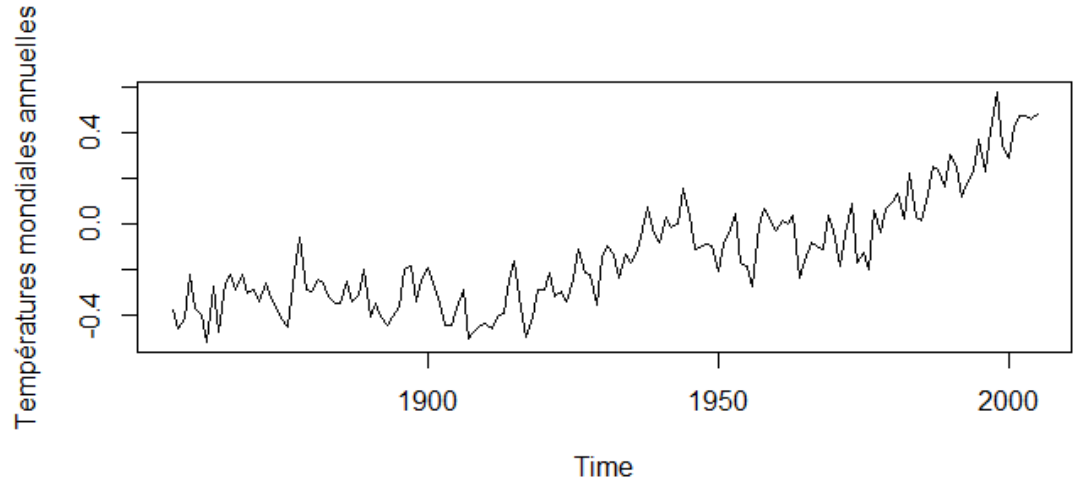
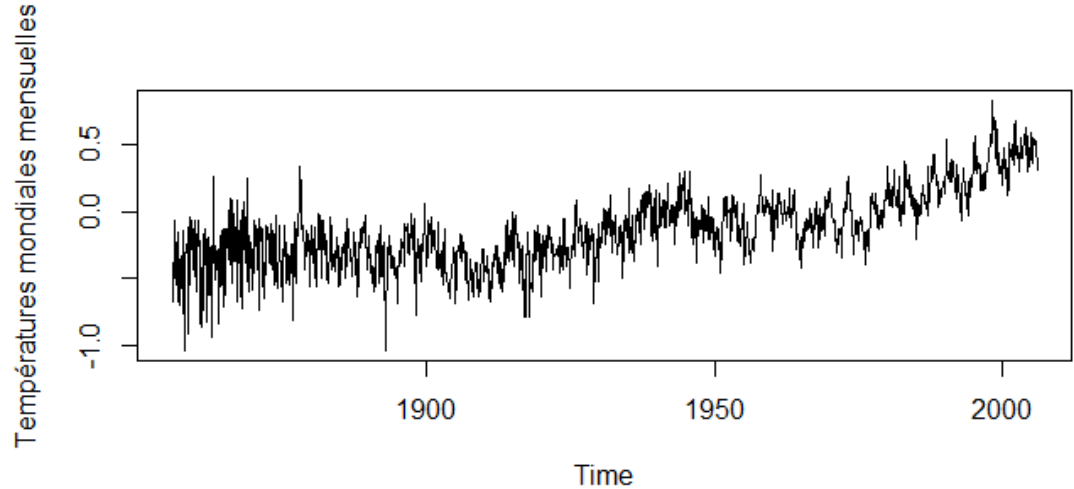
```



Exemple 10: Température Globale

```
Global<- scan("C:/Mon_dossier/global.dat")
Global.ts<-ts(Global,start=c(1856,1),end=c(2005,12),freq=12)
Global.an<-aggregate(Global.ts, FUN=mean)
```

```
layout(1:2)
plot(Global.ts)
plot(Global.an)
```



```
mean(Global.an)
```

```
[1] -0.14
```

```
Global.ar<-ar(Global.an, method="mle")
```

```
Global.ar$order
```

```
Global.ar$ar
```

```
[1] 4
```

```
[1] 0.59 0.01 0.11 0.27
```

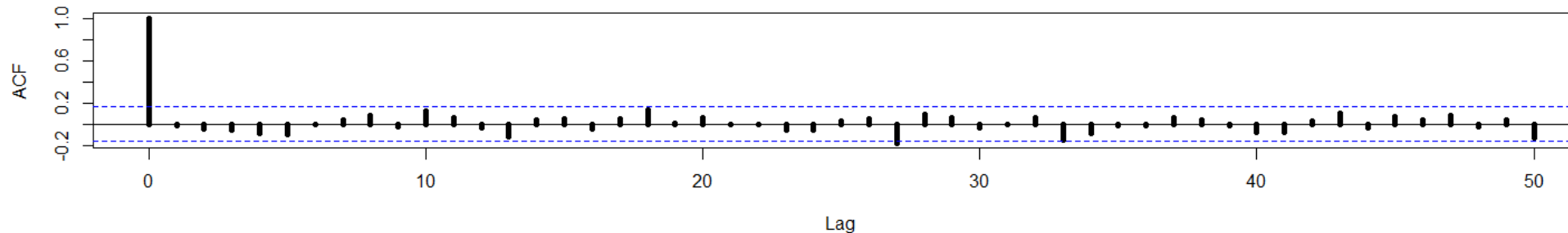
$$\hat{y}_t = -0,14 + 0,59(y_{t-1} + 0,14) + 0,01(y_{t-2} + 0,14) + 0,11(y_{t-3} + 0,14) + 0,27(y_{t-4} + 0,14)$$

```
layout(1:2)
```

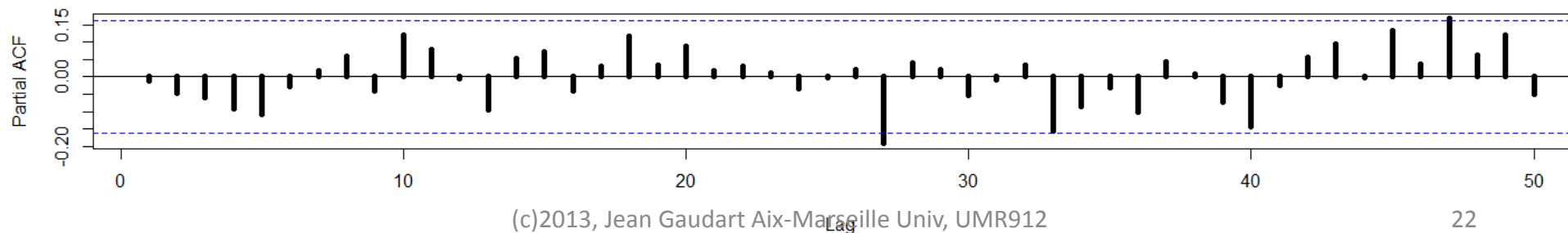
```
acf(Global.ar$res[-(1:Global.ar$order)], lag=50)
```

```
pacf(Global.ar$res[-(1:Global.ar$order)], lag=50)
```

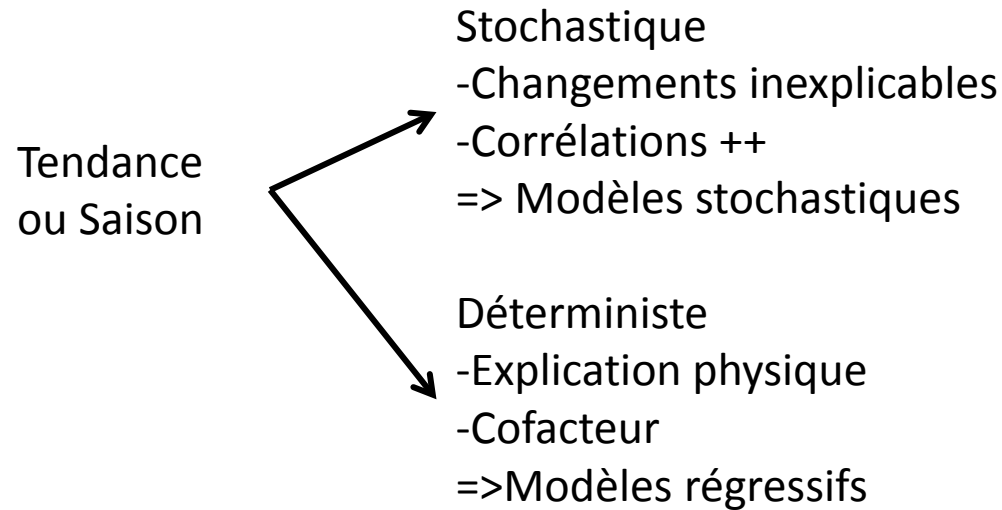
Series Global.ar\$res[-(1:Global.ar\$order)]



Series Global.ar\$res[-(1:Global.ar\$order)]

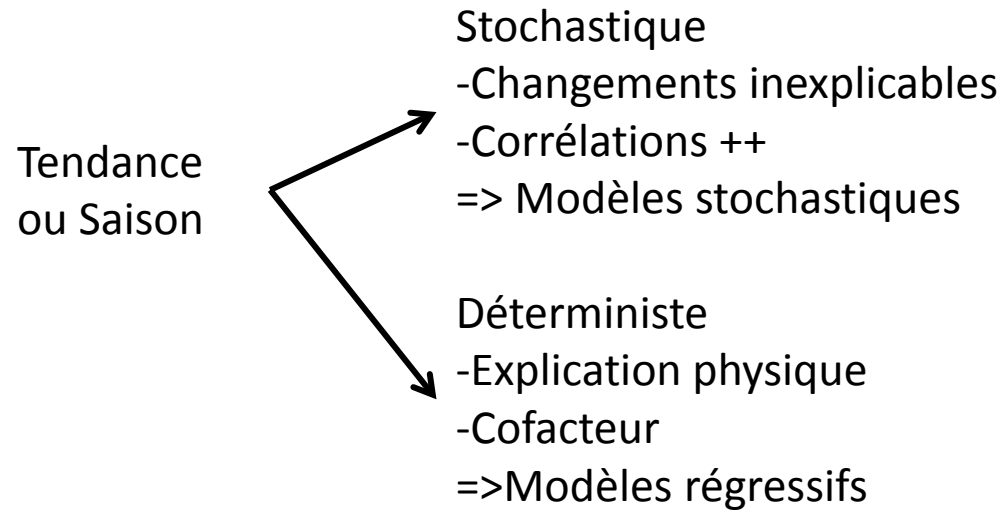


VIII. Régression



NB: les résidus d'un tel modèle régressif => série temporelle, autocorrélées
=> ? *Quel impact sur l'estimation des paramètres ?*

VIII. Régression



NB: les résidus d'un tel modèle régressif => série temporelle, autocorrélées
=> ? *Quel impact sur l'estimation des paramètres ?*

⇒ Sous-estimation des variances des paramètres estimés
⇒ Sous-estimation de la p-value

VIII.1 Modèle linéaire avec tendance

Une série temporelle $\{y_t : t=1\dots n\}$ est linéaire si

$$y_t = \alpha_0 + \alpha_1 X_{1,t} + \alpha_2 X_{2,t} + \dots + \alpha_m X_{m,t} + \varepsilon_t$$

X_i VA explicatives aux temps t , ($i=1\dots m$, $t=1\dots n$)

NB: le modèle suivant est aussi linéaire (par rapport aux covariables)

$$y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_m t^m + \varepsilon_t$$

le modèle suivant n'est pas linéaire (par rapport aux covariables)

$$y_t = \alpha_0 + \alpha_1^{X_t} + \varepsilon_t$$

Certains modèles non-linéaires peuvent être transformés

$$y_t = \exp\{\alpha_0 + \alpha_1 x_t + \varepsilon_t\} \quad \Leftrightarrow \quad \log(y_t) = \alpha_0 + \alpha_1 x_t + \varepsilon_t$$

Mais risque de biais de prévision

A partir du modèle linéaire

$$y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_m t^m + \varepsilon_t$$

On estime la tendance

$$m_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_m t^m$$

Une série linéaire dépendant du temps n'est pas stationnaire

Elle peut-être stationnarisée par différentiation

$$y_t = \alpha_0 + \alpha_1 t + \varepsilon_t$$

$$\Leftrightarrow \nabla^1 y_t = y_t - y_{t-1} = \varepsilon_t - \varepsilon_{t-1} + \alpha_1$$

la série $\{\varepsilon_t\}$ étant stationnaire \Rightarrow la série $\{\nabla^1 y_t\}$ est stationnaire

Si le modèle linéaire est un polynôme de degré $p \Rightarrow$ différentiation d'ordre p

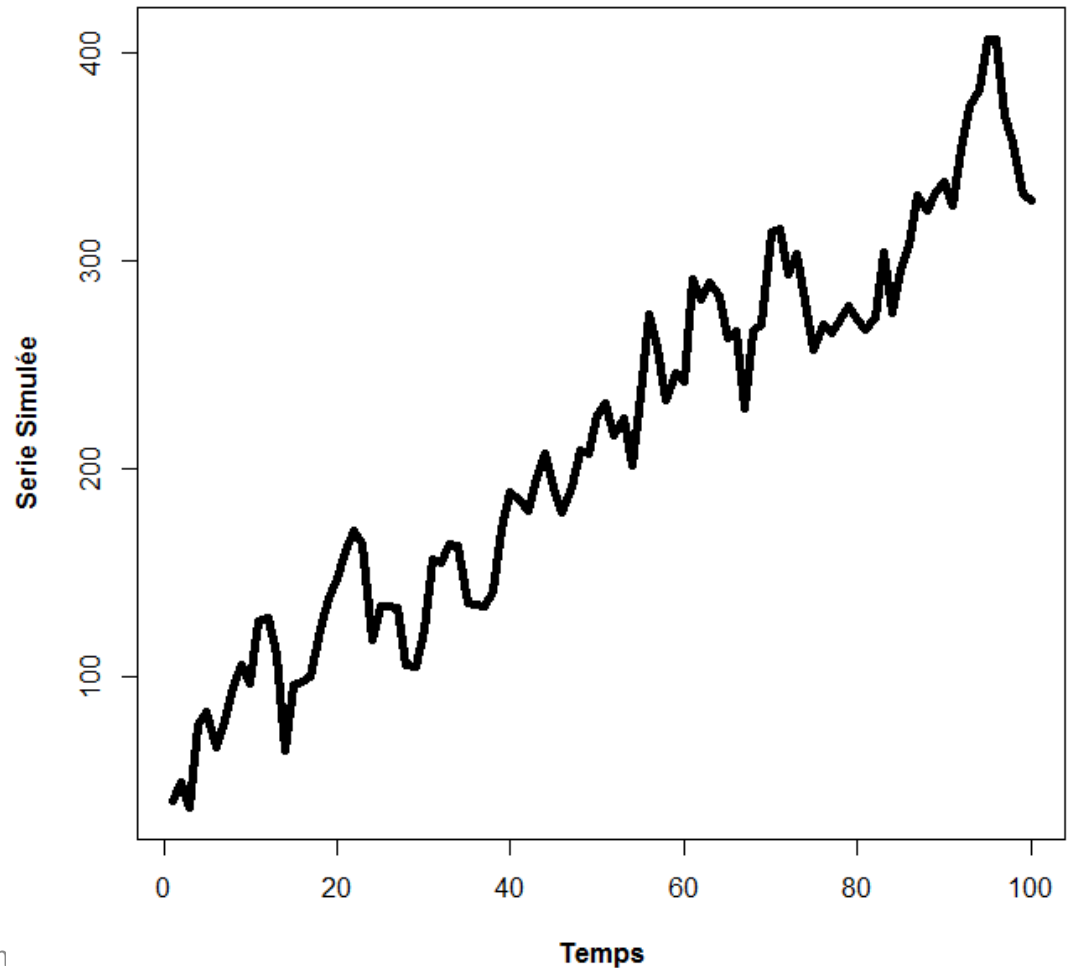
Simulations

```
ar1<-e<-rnorm(100,sd=20)
for(t in 2:100){ar1[t]<-0.8*ar1[t-1]+e[t]}
```

```
Temps<-1:100
y<-50+3*Temps+ar1
y.ts<-ts(y)

plot(y.ts)
```

? Expliquez ce modèle simulé?



```
ar1<-e<-rnorm(100,sd=20)
for(t in 2:100){ar1[t]<-0.8*ar1[t-1]+e[t]}
```

```
Temps<-1:100
y<-50+3*Temps+ar1
y.ts<-ts(y)
```

$$\mathbf{ar}_t = 0,8\mathbf{ar}_{t-1} + \boldsymbol{\varepsilon}_t$$

$$\mathbf{y}_t = 50 + 3t + \mathbf{ar}_t$$

$$\mathbf{y}_t = 50 + 3t + 0,8\mathbf{ar}_{t-1} + \boldsymbol{\varepsilon}_t$$

Estimation du modèle linéaire

Habituellement => minimisation du SSE, avec la fonction `lm()`

```
y.lm<-lm(y~Temps)  
summary(y.lm)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	58.5512	4.8801	12.00	<2e-16	***
Temps	3.0633	0.0839	36.51	<2e-16	***

?Interprétez la p-value?

Estimation du modèle linéaire : série simulée

Habituellement => minimisation du SSE, avec la fonction `lm()`

```
y.lm<-lm(y~Temps)  
summary(y.lm)
```

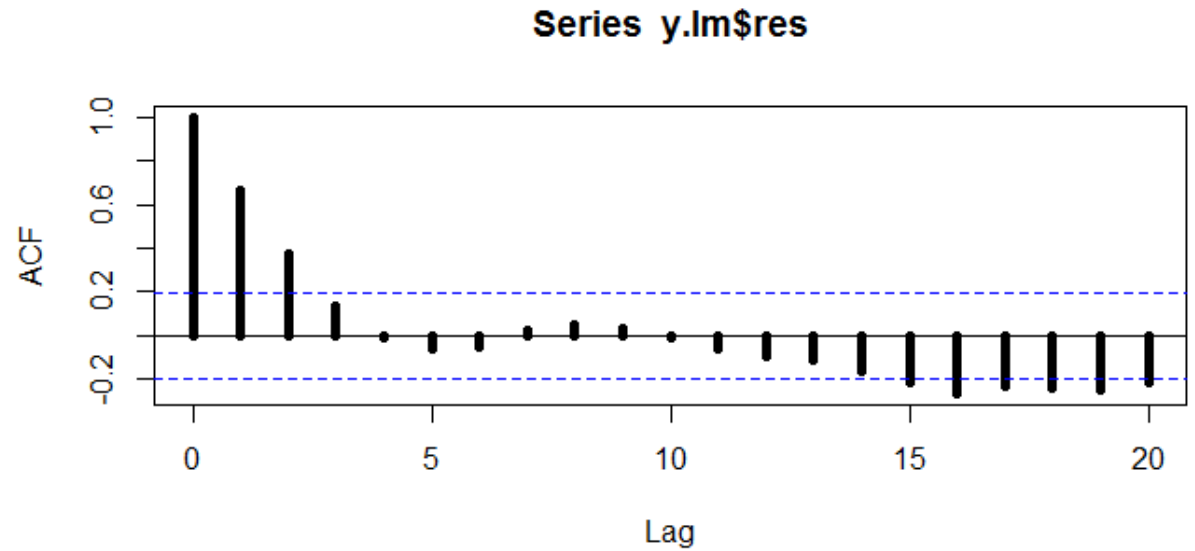
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	58.5512	4.8801	12.00	<2e-16	***
Temps	3.0633	0.0839	36.51	<2e-16	***

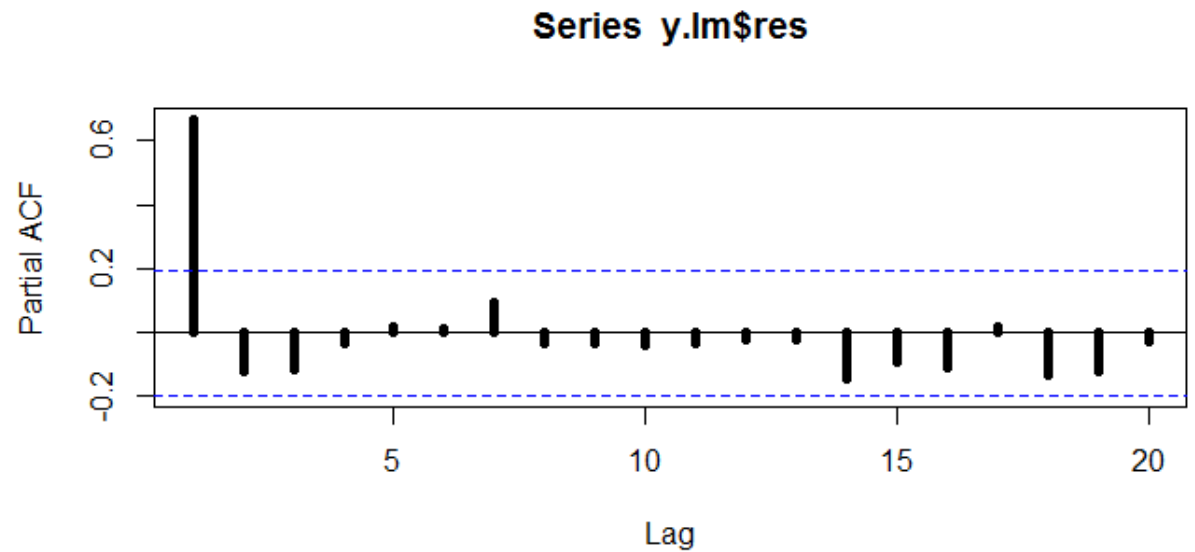
Sous-estimé

Test incorrect

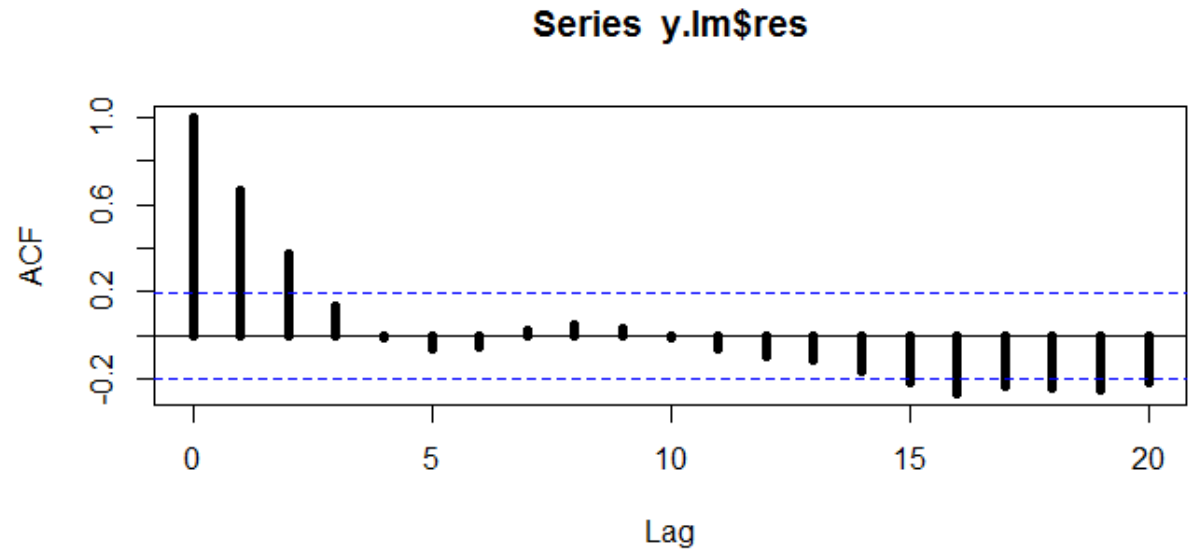
```
layout(1:2)
acf(y.lm$res)
pacf(y.lm$res)
```



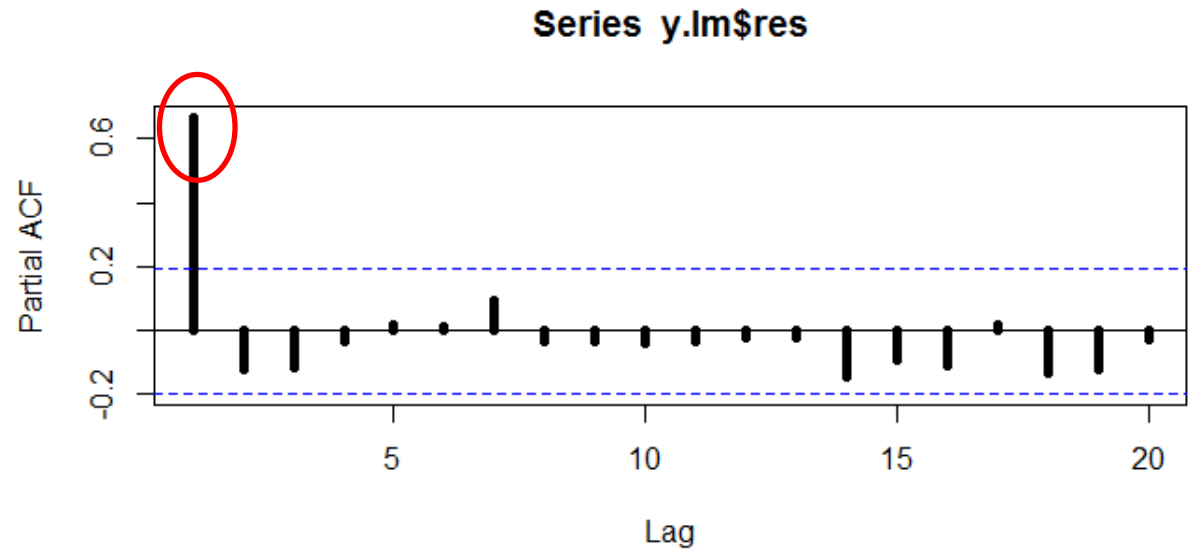
?Interprétez?



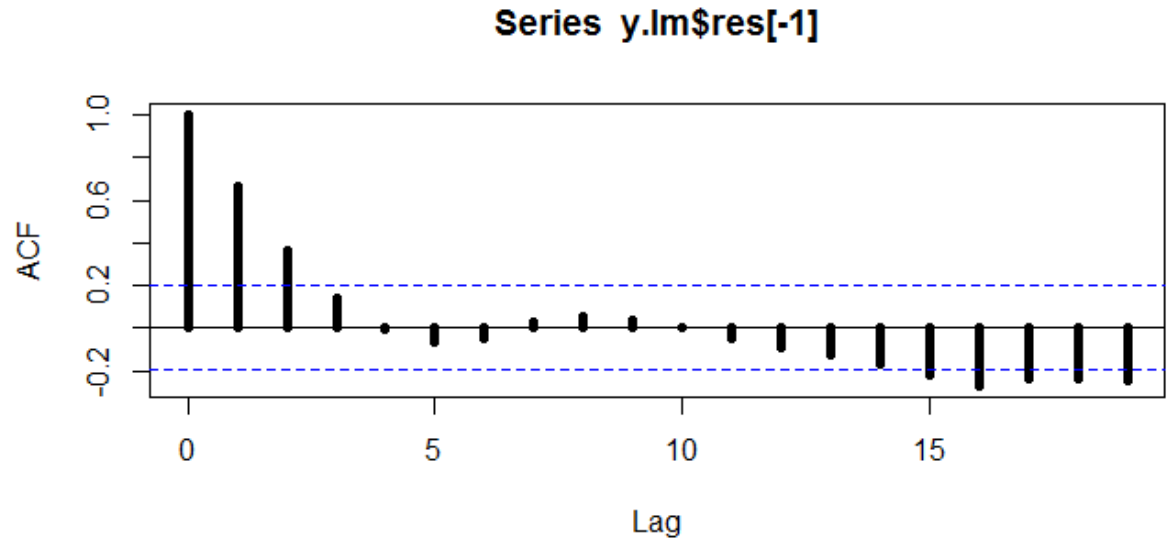
```
layout(1:2)
acf(y.lm$res)
pacf(y.lm$res)
```



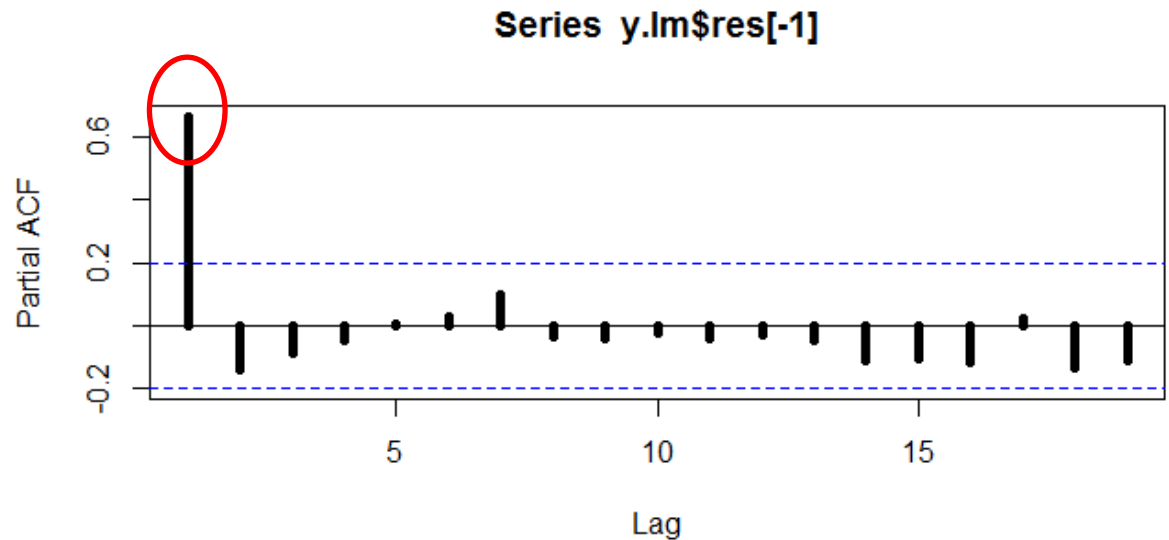
acf: décroissance exp
pacf: autocorr lag 1
=>AR(1) probable




```
layout(1:2)
acf(y.lm$res)
pacf(y.lm$res)
```



acf: décroissance exp
pacf: autocorr lag 1
=>AR(1) probable



Rappel: simulation

```
y<-50+3*Temps+ar1
```

Reprise Ex. 10: Température Globale

```
Global<-scan("C:/Mon_dossier/global.dat")
Global.ts<-ts(Global,start=c(1856,1),end=c(2005,12),freq=12)
Global.70<-window(Global.ts, start=c(1970, 1), end=c(2005,12))
Global.lm<-lm(Global.70~time(Global.70))
summary(Global.lm)
```

Coefficients:

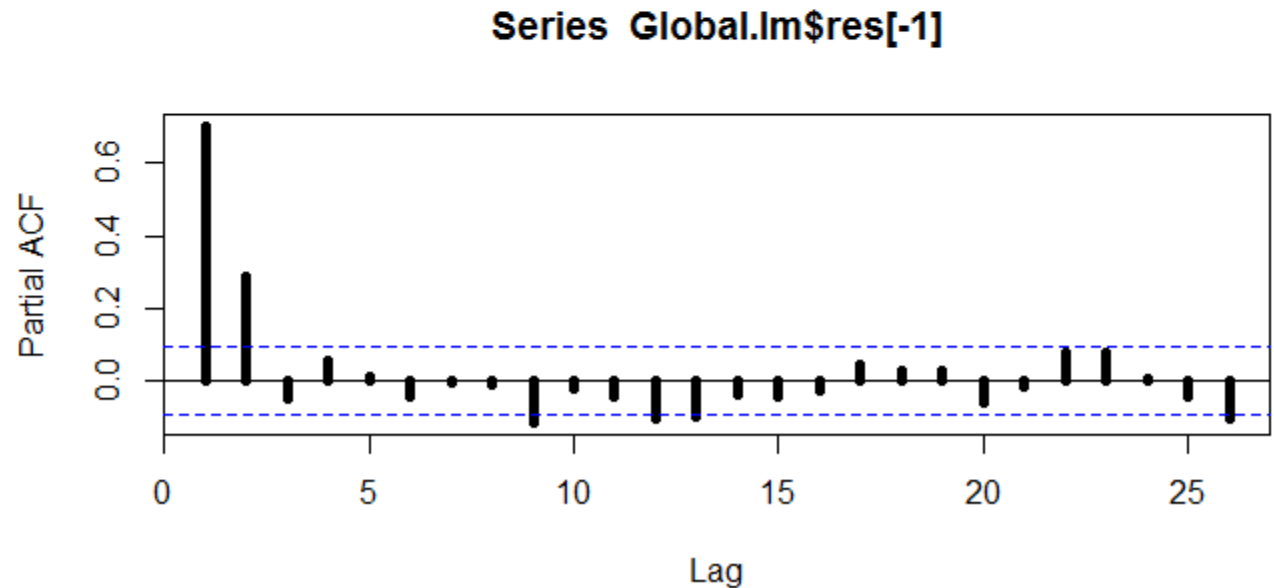
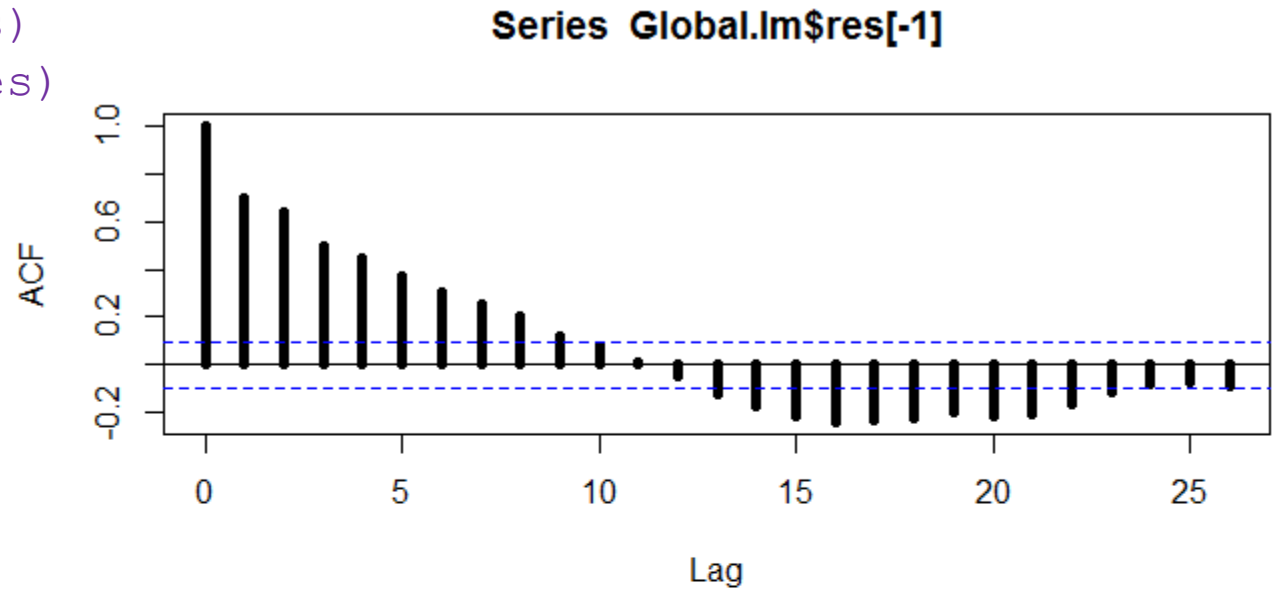
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-34.920409	1.164899	-29.98	<2e-16	***
time(Global.70)	0.017654	0.000586	30.13	<2e-16	***

```
confint(Global.lm)
```

	2.5 %	97.5 %
(Intercept)	-37.21001248	-32.63080554
time(Global.70)	0.01650228	0.01880572

```
layout(1:2)
acf(Global.lm$res)
pacf(Global.lm$res)
```

?Interprétez?



Generalised Least Squares: GLS

Résidus corrélés => pb d'estimation avec `lm()`

Max de vraisemblance étant donnée une autocorrelation => `gls()`

Estimation du modèle linéaire : série simulée

```
chooseCRANmirror(ind=6)
install.packages("nlme")
library(nlme)
y.gls<-gls(y~Temps,cor=corAR1(0.8))
summary(y.gls)
```

Generalized least squares fit by REML

Model: y ~ Temps

Data: NULL

	AIC	BIC	logLik
	862.8866	873.2265	-427.4433

Correlation Structure: AR(1)

Formula: ~1

Parameter estimate(s):

Phi
0.7161368

Coefficients:

	Value	Std.Error	t-value	p-value
(Intercept)	58.23302	11.924568	4.883449	0
Temps	3.04225	0.202445	15.027538	0

Correlation:

(Intr)
Temps -0.857

Standardized residuals:

	Min	Q1	Med	Q3	Max
	-1.6171329	-0.6195428	0.0353972	0.5836326	2.3184155

Residual standard error: 25.58595

Degrees of freedom: 100 total; 98 residual

Reprise Ex. 10 : Température globale

```
print(acf(Global.lm$res))          0.706
Global.gls<-gls(Global.70~time(Global.70),cor=corAR1(0.8))
summary(Global.gls)
```

Coefficients:

	Value	Std.Error	t-value	p-value
(Intercept)	-34.15115	2.8850347	-11.83734	0
time(Global.70)	0.01727	0.0014512	11.89822	0

```
confint(Global.gls)
```

	2.5 %	97.5 %
(Intercept)	-39.80571681	-28.49658850
time(Global.70)	0.01442274	0.02011148

⇒Tendance modélisée par le modèle linéaire

⇒AR(1) modélisé

Interprétation: il existe bien une augmentation significative de la température pendant cette période

VIII.2 Modèle linéaire avec saisonnalité

Saisonnalité : modélisée par une série de Fourier

Simulation:

$$y_t = 0,1 + 0,005t + 0,001t^2 + \sin(2\pi t / 12) + 0,2 \sin(4\pi t / 12) + 0,1 \sin(8\pi t / 12) + 0,1 \cos(8\pi t / 12) + \varepsilon_t$$

```
Temps<-1:(10*12)
```

```
e<-rnorm(10*12,sd=0.5)
```

```
Trend<-0.1+0.005*Temps+0.001*Temps^2
```

```
Season<-sin(2*pi*Temps/12)+0.2*sin(4*pi*Temps/12)+0.1*sin(8*pi*Temps/12)+0.1*cos(8*pi*Temps/12)
```

```
y<-Trend+Season+e
```

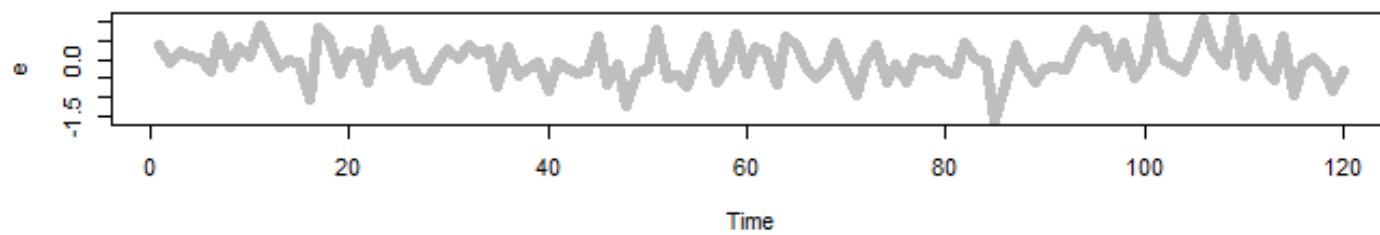
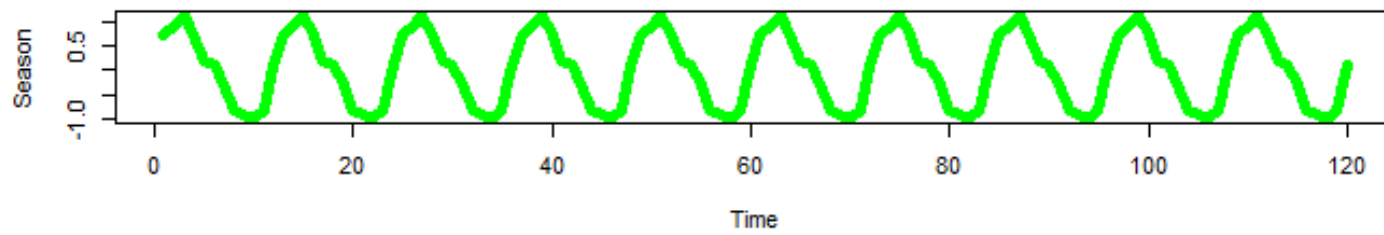
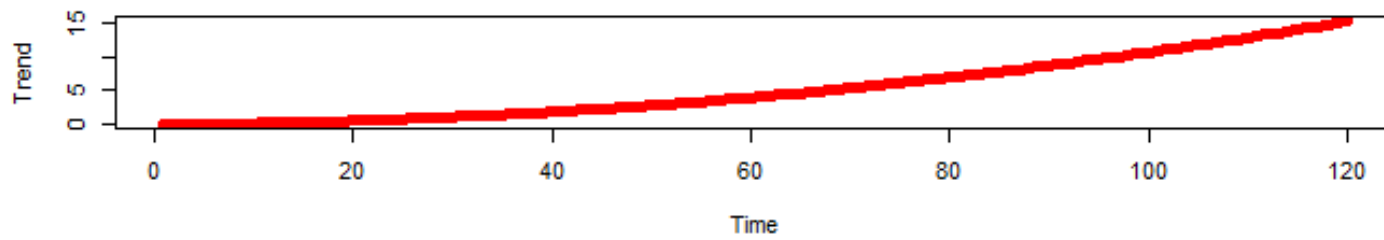
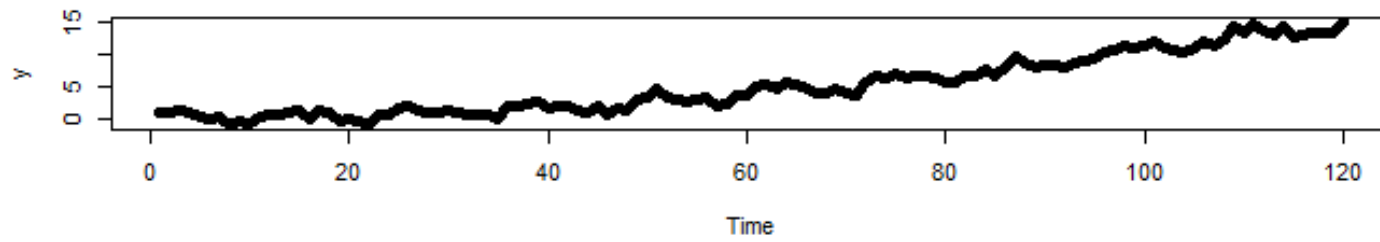
```
layout(1:4)
```

```
ts.plot(y,col="black")
```

```
ts.plot(Trend,col="red")
```

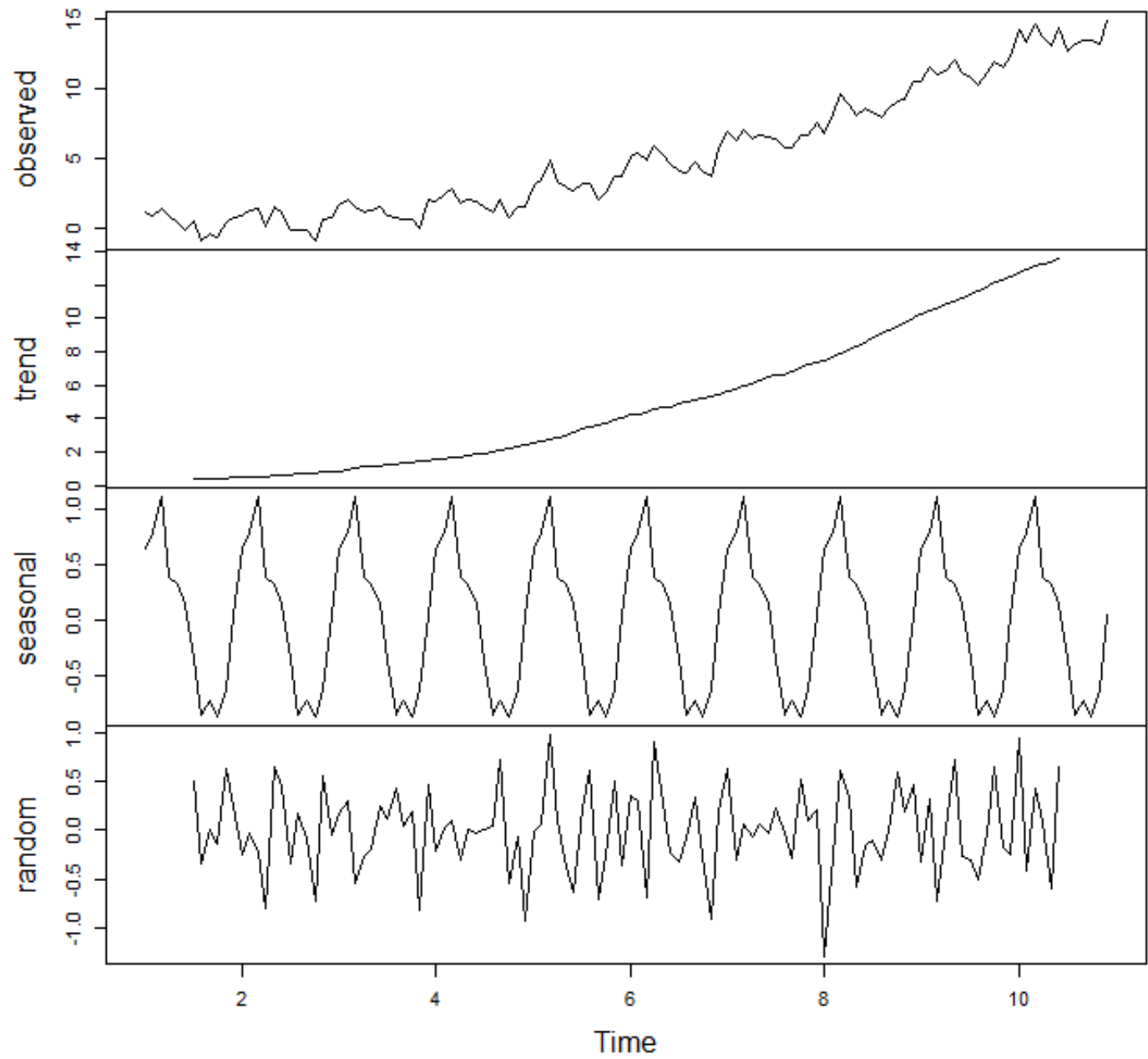
```
ts.plot(Season,col="green")
```

```
ts.plot(e,col="grey")
```



Decomposition of additive time series

```
plot(decompose(ts(y, freq=12)))
```



Estimation du modèle sur série simulée

Construction d'une matrice de sin et cos (pour rendre le script plus facile)

```
SIN<-COS<-matrix(nr=length(Temps),nc=6)
for (i in 1:6){
  COS[,i]<-cos(2*pi*i*Temps/12)
  SIN[,i]<-sin(2*pi*i*Temps/12)
}
```

Ecriture du modèle linéaire :

```
y.lm<-lm(y~Temps+I(Temps^2)
+COS[,1]+SIN[,1]+COS[,2]+SIN[,2]+COS[,3]+SIN[,3]
+COS[,4]+SIN[,4]+COS[,5]+SIN[,5]+COS[,6]+SIN[,6])
```

```
summary(y.lm)
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.745e-01	1.387e-01	1.979	0.0504	.
Temps	-4.143e-03	5.284e-03	-0.784	0.4347	
I (Temps^2)	1.069e-03	4.231e-05	25.258	<2e-16	***
COS[, 1]	2.943e-02	6.420e-02	0.458	0.6475	
SIN[, 1]	9.197e-01	6.437e-02	14.288	<2e-16	***
COS[, 2]	-1.232e-02	6.444e-02	-0.191	0.8487	
SIN[, 2]	1.559e-01	6.477e-02	2.407	0.0178	*
COS[, 3]	-6.407e-02	6.423e-02	-0.997	0.3208	
SIN[, 3]	4.488e-02	6.436e-02	0.697	0.4871	
COS[, 4]	1.473e-01	6.434e-02	2.289	0.0241	*
SIN[, 4]	4.144e-02	6.423e-02	0.645	0.5202	
COS[, 5]	5.602e-03	6.505e-02	0.086	0.9315	
SIN[, 5]	5.561e-02	6.512e-02	0.854	0.3950	
COS[, 6]	-6.179e-02	5.240e-02	-1.179	0.2410	
SIN[, 6]	-1.082e+12	3.113e+12	-0.348	0.7288	

step(y.lm)

Step: AIC=-168.04

y ~ I(Temps^2) + SIN[, 1] + SIN[, 2] + COS[, 4]

	Df	Sum of Sq	RSS	AIC
<none>			27.22	-168.04
- COS[, 4]	1	1.27	28.49	-164.56
- SIN[, 2]	1	1.52	28.74	-163.52
- SIN[, 1]	1	50.55	77.76	-44.06
- I(Temps^2)	1	2395.44	2422.66	368.62

Call:

lm(formula = y ~ I(Temps^2) + SIN[, 1] + SIN[, 2] + COS[, 4])

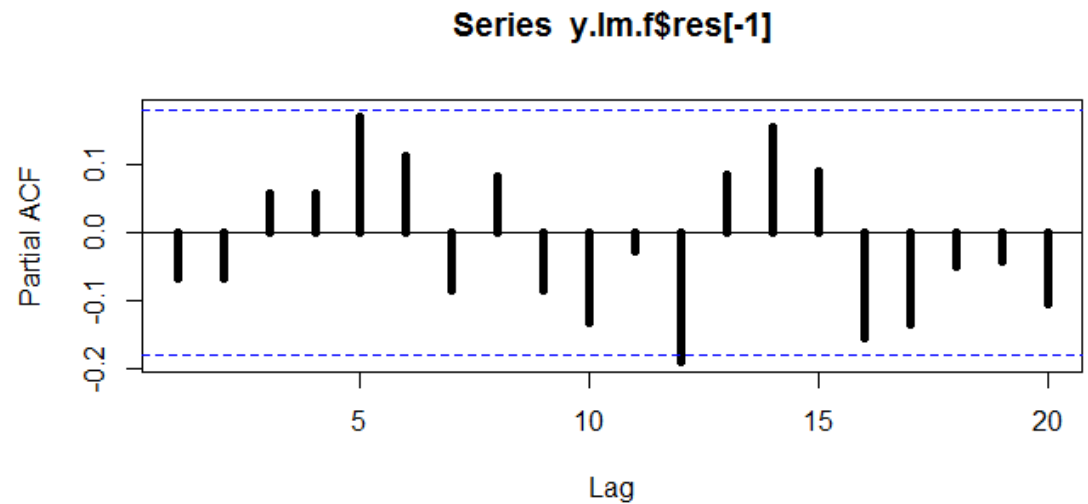
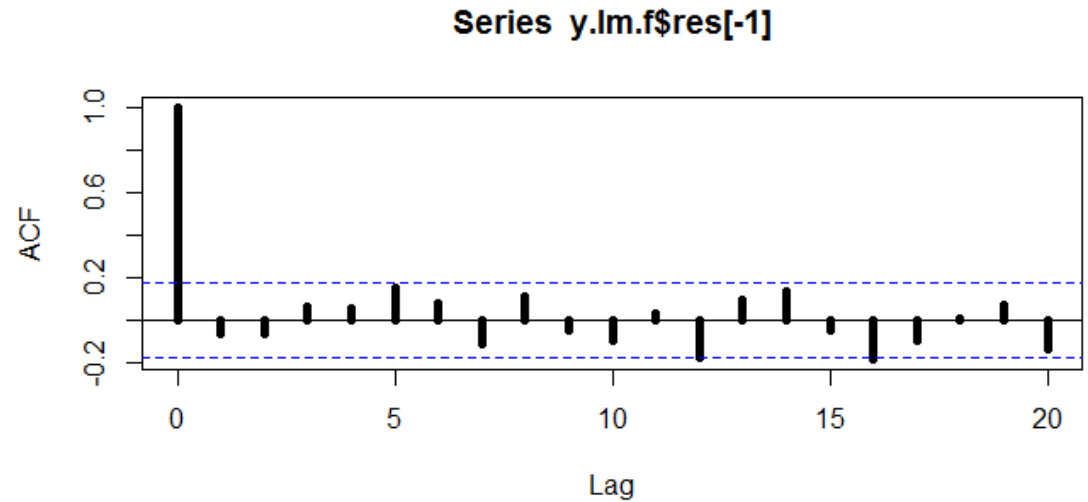
Coefficients:

(Intercept)	I(Temps^2)	SIN[, 1]	SIN[, 2]	COS[, 4]
0.182435	0.001036	0.920317	0.159228	0.145588

```
y.lm.f<-  
lm(y~I(Temps^2)+SIN[,1]+SIN[,2]+COS[,4])
```

```
layout(1:2)  
acf(y.lm.f$res)  
pacf(y.lm.f$res)
```

?Interprétez?



Estimation du modèle ex. 10: température globale 1970-2005

Construction d'une matrice de sin et cos (pour rendre le script plus facile)

```
SIN<-COS<-matrix(nr=length(Global.70),nc=6)
for (i in 1:6){
  COS[,i]<-cos(2*pi*i*time(Global.70))
  SIN[,i]<-sin(2*pi*i*time(Global.70))
}
```

Standardisation de la variable Temps pour éviter l'impact des grands nombres

```
Temps<-(time(Global.70)-mean(time(Global.70)))/sd(time(Global.70))
```

Ecriture du modèle

```
Global.lm<-lm(Global.70~Temps+I(Temps^2)
+COS[,1]+SIN[,1]+COS[,2]+SIN[,2]+COS[,3]+SIN[,3]
+COS[,4]+SIN[,4]+COS[,5]+SIN[,5]+COS[,6]+SIN[,6])
```

summary(Global.lm)

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.664e-01	9.117e-03	18.257	<2e-16	***
Temps	1.843e-01	6.087e-03	30.280	<2e-16	***
I (Temps^2)	8.673e-03	6.810e-03	1.274	0.2035	
COS[, 1]	6.402e-03	8.595e-03	0.745	0.4568	
SIN[, 1]	2.050e-02	8.596e-03	2.385	0.0175	*
COS[, 2]	1.065e-02	8.810e-03	1.209	0.2275	
SIN[, 2]	1.667e-02	8.633e-03	1.931	0.0541	.
COS[, 3]	5.541e-03	8.593e-03	0.645	0.5194	
SIN[, 3]	3.413e-03	8.594e-03	0.397	0.6914	
COS[, 4]	4.699e-03	8.593e-03	0.547	0.5848	
SIN[, 4]	1.445e-03	8.593e-03	0.168	0.8665	
COS[, 5]	2.724e-03	8.593e-03	0.317	0.7514	
SIN[, 5]	3.012e-03	8.593e-03	0.350	0.7262	
COS[, 6]	-2.858e-03	6.354e-03	-0.450	0.6530	
SIN[, 6]	-7.821e+08	1.258e+09	-0.622	0.5345	

step(Global.lm)

Step: AIC=-1789.14

Global.70 ~ Temps + SIN[, 1] + SIN[, 2]

	Df	Sum of Sq	RSS	AIC
<none>			6.7421	-1789.1
- SIN[, 2]	1	0.0564	6.7985	-1787.5
- SIN[, 1]	1	0.0900	6.8322	-1785.4
- Temps	1	14.5992	21.3414	-1293.4

Call:

lm(formula = Global.70 ~ Temps + SIN[, 1] + SIN[, 2])

Coefficients:

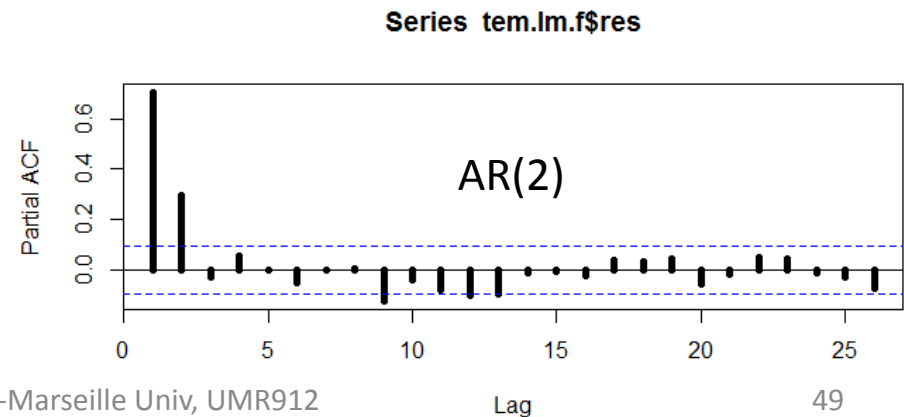
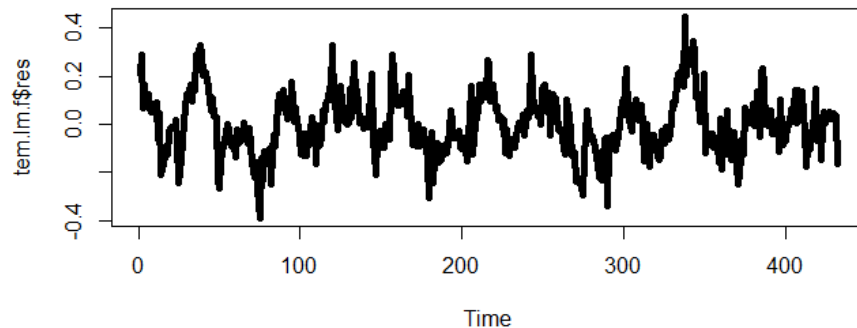
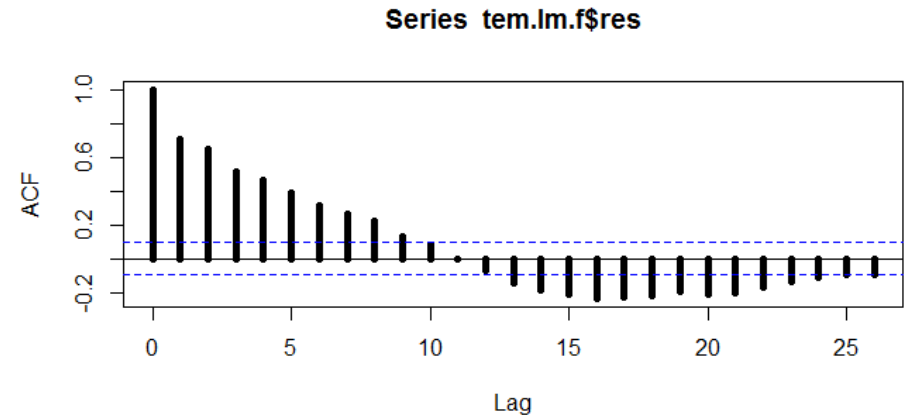
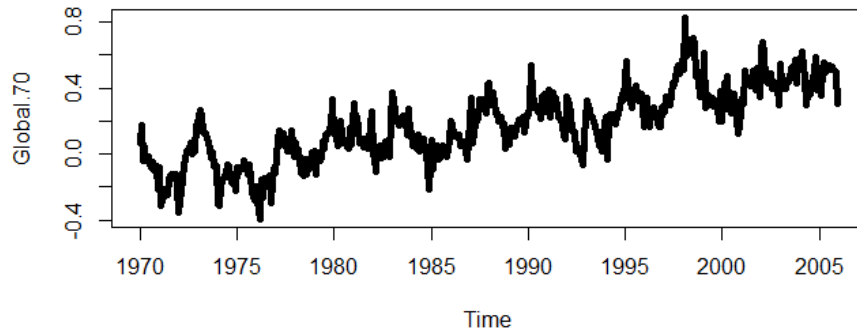
(Intercept)	Temps	SIN[, 1]	SIN[, 2]
0.17501	0.18410	0.02042	0.01615


```
Global.lm.f<-lm(Global.70~Temps+SIN[,1]+SIN[,2])
```

```
layout(1:2)  
ts.plot(Global.70)  
ts.plot(Global.lm.f$res)
```

```
layout(1:2)  
acf(Global.lm.f$res)  
pacf(Global.lm.f$res)
```

?Interprétez?



Vérification de l'autoregressif

```
res.ar<-ar(Global.lm.f$res, method="mle")
```

```
res.ar$order [1] 2
```

```
res.ar$ar
```

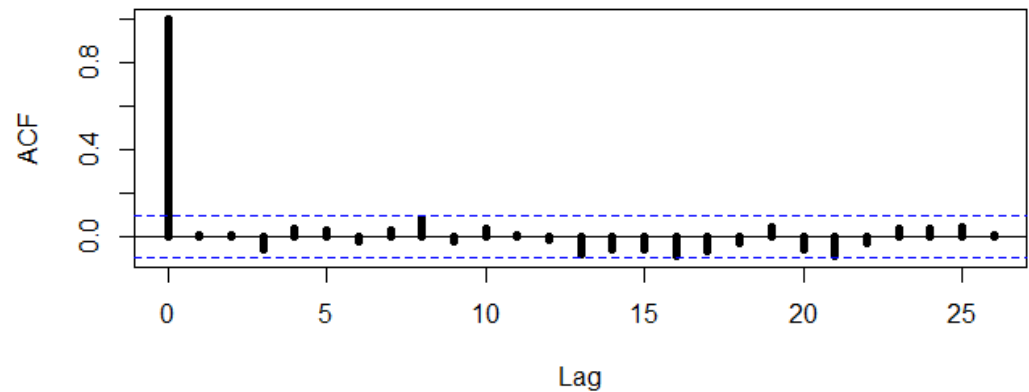
```
[1] 0.4938189 0.3071598
```

```
layout(1:2)
```

```
acf(res.ar$res[-(1:2)])
```

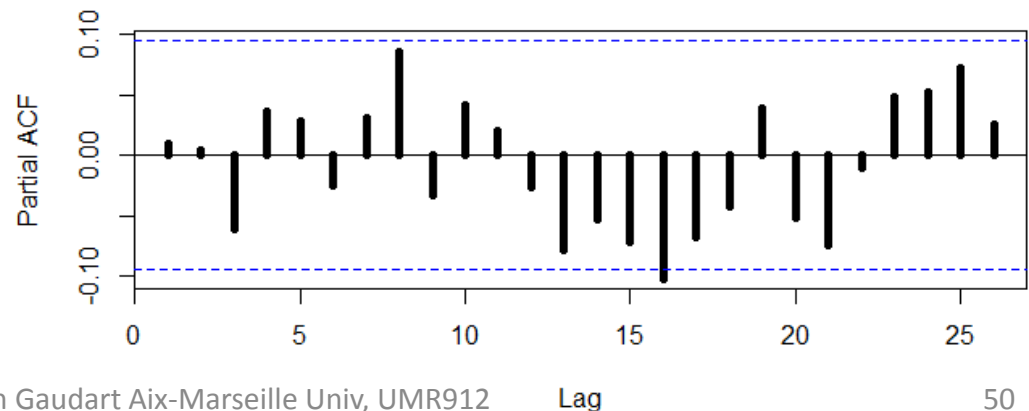
```
pacf(res.ar$res[-(1:2)])
```

Series res.ar\$res[-(1:2)]



?Interprétez?

Series res.ar\$res[-(1:2)]



```
Global.gls.f<-gls(Global.70~Temps+SIN[,1]
+SIN[,2],cor=corARMA(c(0.5,0.3),p=2,q=0))
```

```
summary(Global.gls.f)
```

Generalized least squares fit by REML

Model: Global.70 ~ Temps + SIN[, 1] + SIN[, 2]

Data: NULL

AIC	BIC	logLik
-871.5613	-843.1475	442.7807

Correlation Structure: ARMA(2,0)

Formula: ~1

Parameter estimate(s):

Phi1	Phi2
0.5022575	0.3111635

Coefficients:

	Value	Std.Error	t-value	p-value
(Intercept)	0.17665852	0.021450027	8.235818	0.0000
Temps	0.17612720	0.020894431	8.429385	0.0000
SIN[, 1]	0.01782806	0.008686392	2.052412	0.0407
SIN[, 2]	0.01486394	0.005015381	2.963670	0.0032

Estimation du modèle ex. 1: voyageurs PanAM

Construction d'une matrice de sin et cos (pour rendre le script plus facile)

```
SIN<-COS<-matrix(nr=length(AP),nc=6)
for (i in 1:6){
  COS[,i]<-cos(2*pi*i*time(AP))
  SIN[,i]<-sin(2*pi*i*time(AP))
}
```

Standardisation de la variable Temps pour éviter l'impact des grands nombres

```
TIME<-(time(AP)-mean(time(AP)))/sd(time(AP))
```

Ecriture du modèle

```
AP.lm<-lm(log(AP) ~ TIME +I(TIME^2)
+COS[,1]+SIN[,1]+COS[,2]+SIN[,2]+COS[,3]+SIN[,3]
+COS[,4]+SIN[,4]+COS[,5]+SIN[,5]+COS[,6]+SIN[,6])
```

Pas-à-Pas `step(AP.lm)`

Log: la variance croit avec le tps

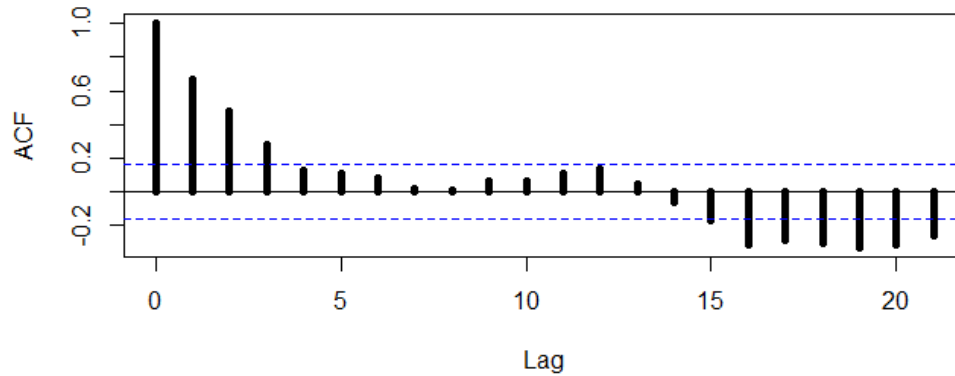
Modèle final

```
AP.lm2<-lm(log(AP) ~TIME+I(TIME^2)
+COS[,1]+SIN[,1]+COS[,2]+SIN[,2]+COS[,3]
+COS[,4]+SIN[,4]+SIN[,5])
```

```
layout(1:2)
acf(AP.lm2$res)
pacf(AP.lm2$res)
```

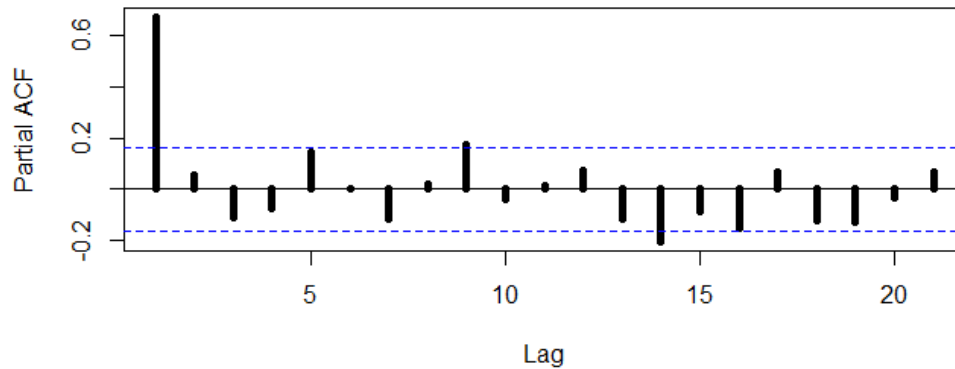
```
AP.ar<-ar(AP.lm2$res,method="mle")
AP.ar$order
[1] 1
AP.ar$ar
[1] 0.673178
```

Series AP.lm2\$res



?Interprétez?

Series AP.lm2\$res



Modèle gls

```
AP.gls2<-gls (log (AP) ~TIME+I (TIME^2)  
+COS [, 1]+SIN [, 1]+COS [, 2]+SIN [, 2]+COS [, 3]  
+COS [, 4]+SIN [, 4]+SIN [, 5], cor=corAR1 (0.67) )
```

Préparation des données pour une prédiction

```
new.t<-time (ts (start=1949, end=c (1970, 12) , freq=12) )
```

```
TIME<- (new.t-mean (time (AP) ) ) /sd (time (AP) )
```

```
SIN<-COS<-matrix (nr=length (new.t) , nc=6)
```

```
for (i in 1:6) {  
  COS [, i]<-cos (2*pi*i*new.t)  
  SIN [, i]<-sin (2*pi*i*new.t)  
}
```

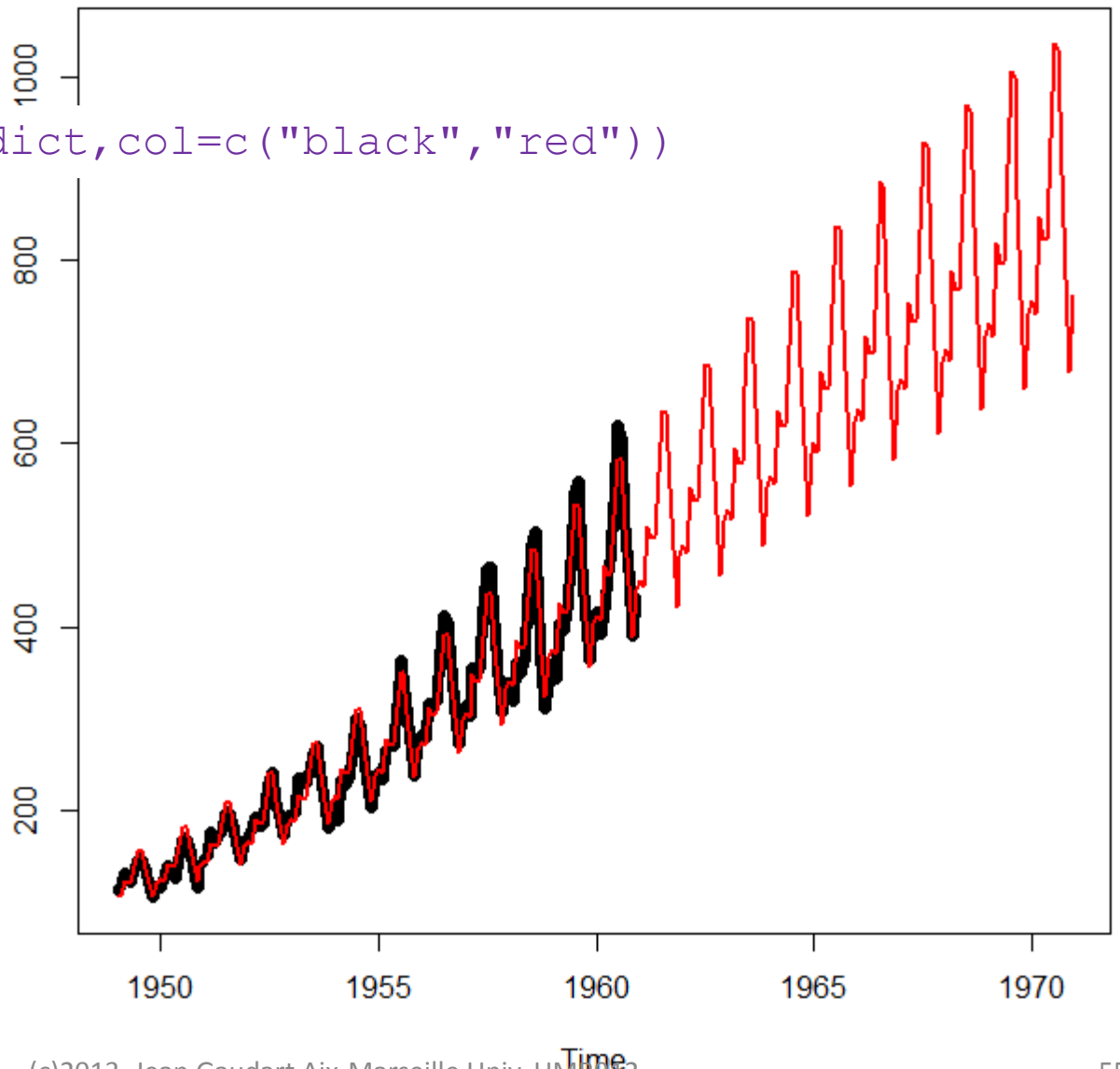
```
SIN<-SIN [, -6]
```

```
new.dat<-data.frame (TIME=as.vector (TIME) , SIN=SIN, COS=COS)
```

```
lAP.p<-predict (AP.gls2,new.dat)
lAP.p.ts<-ts (lAP.p,start=1949,freq=12)
AP.predict<-exp (lAP.p.ts)
```

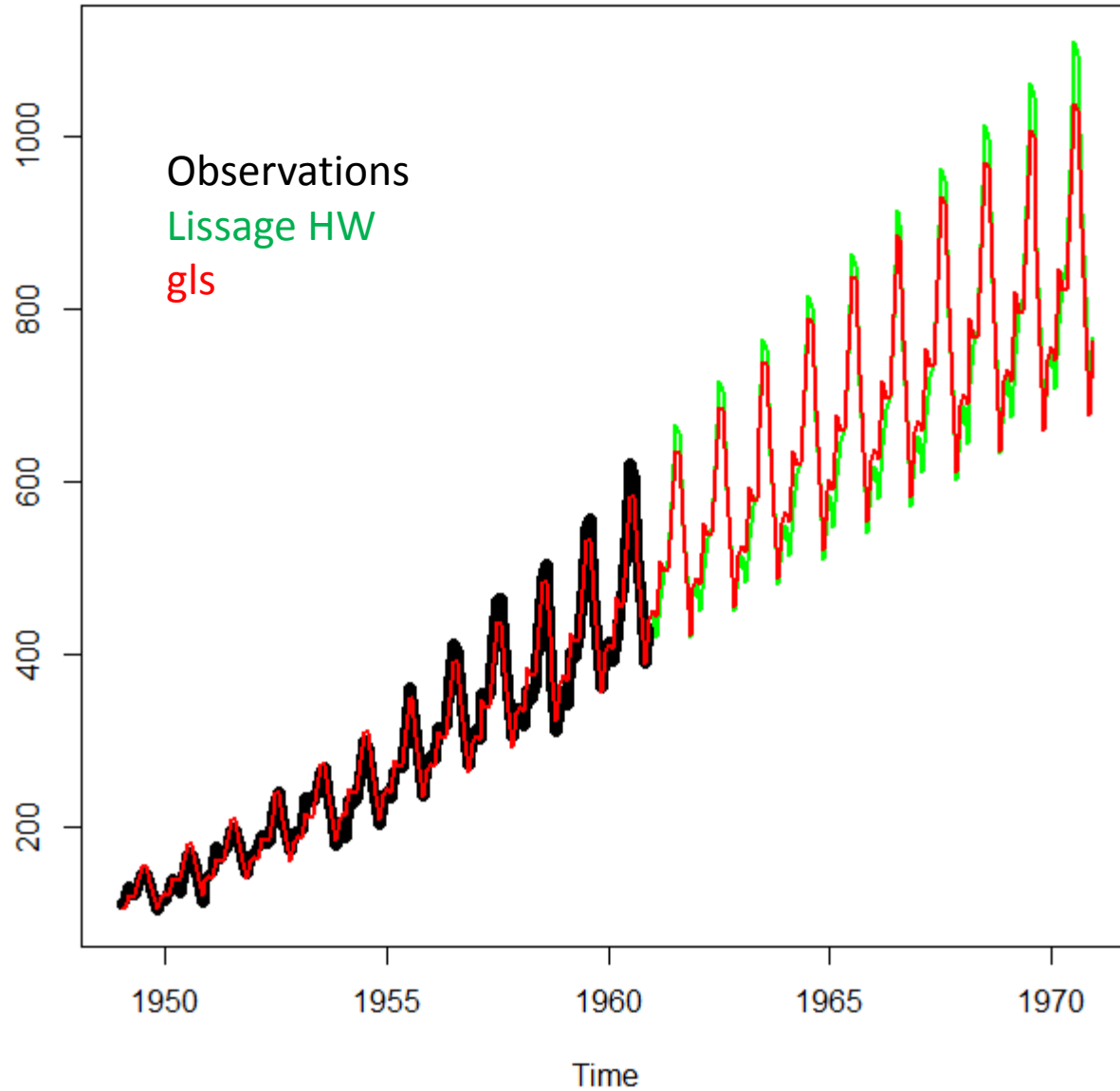
```
ts.plot (AP,AP.predict,col=c ("black","red"))
```

?Interprétez?



```
ts.plot(AP, AP.liss.p, AP.predict,  
+col=c("black", "green", "red"), lwd=c(6, 2, 2))
```

?Interprétez?



IX. Modèles stationnaires : ARMA

Objectif: améliorer la modélisation de l'autocorrélation de modèles stationnaires

Attention: si on tente de modéliser une série temporelle par un modèle stationnaire, on doit assumer que les observations sont une réalisation d'une série stationnaire.

Vérifier tendance et saison et ajuster , par ex avec un modèle linéaire => résidus

IX.1 Moyenne Mobile: MA

$$Y_t = \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q} + \varepsilon_t$$

Ré-écrit avec l'opérateur retard:

Polynôme d'ordre q

$$Y_t = \left(1 + \beta_1 B + \beta_2 B^2 + \dots + \beta_q B^q\right) \varepsilon_t = \phi_q(B) \varepsilon_t$$

MA => Somme finie de bruits blancs stationnaires => processus stationnaire

$$E(Y_t) = 0$$

$$\text{VAR}(Y_t) = \sigma_\varepsilon^2 (1 + \beta_1^2 + \dots + \beta_q^2)$$

$$\rho(k) \begin{cases} 1 & \text{si } k = 0 \\ \frac{\sum_{i=0}^{q-k} \beta_i \beta_{i+k}}{\sum_{i=0}^q \beta_i^2} & \text{si } k = 1, \dots, q \\ 0 & \text{si } k > q \end{cases}$$

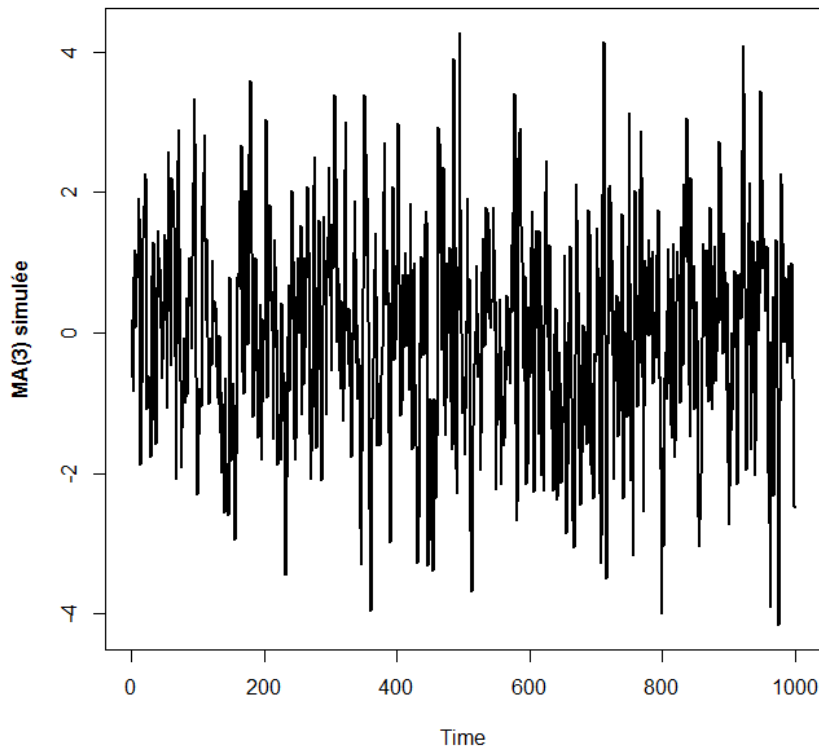
Avec $\beta_0 = 1$

En général $\beta_0 \leq 1$, pour convergence

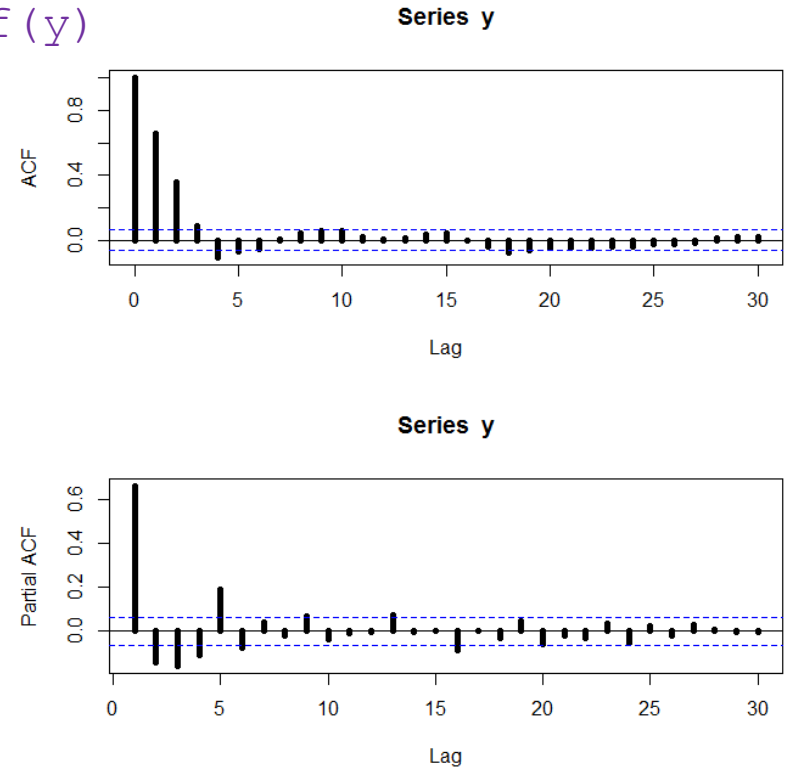
Simulation d'un MA(3)

```
b<-c(0.8,0.6,0.4)
y<-e<-rnorm(1000)
for (t in 4:1000){
  for (j in 1:3){y[t]<-y[t]+b[j]*e[t-j]}
}
```

```
ts.plot(y)
```



```
layout(1:2)
acf(y)
pacf(y)
```



Estimation du modèle simulé

```
y.ma<-arima(y,order=c(0,0,3),include.mean=FALSE)
```

```
y.ma
```

← Si la moyenne est nulle

Call:

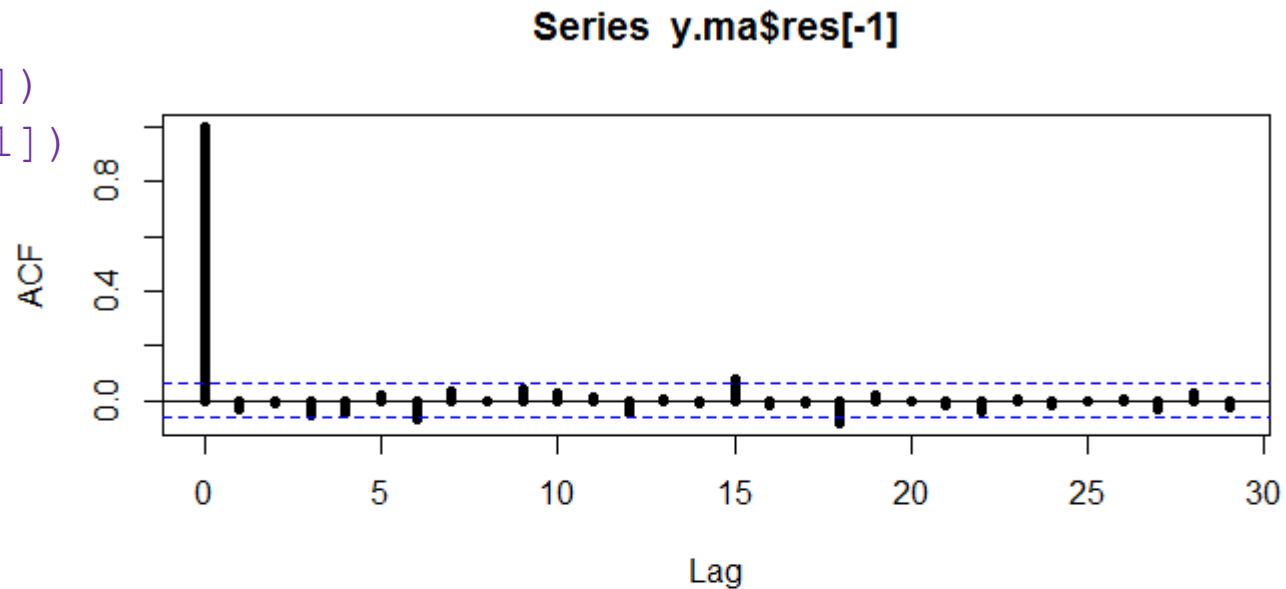
```
arima(x = y, order = c(0, 0, 3), include.mean = FALSE)
```

Coefficients:

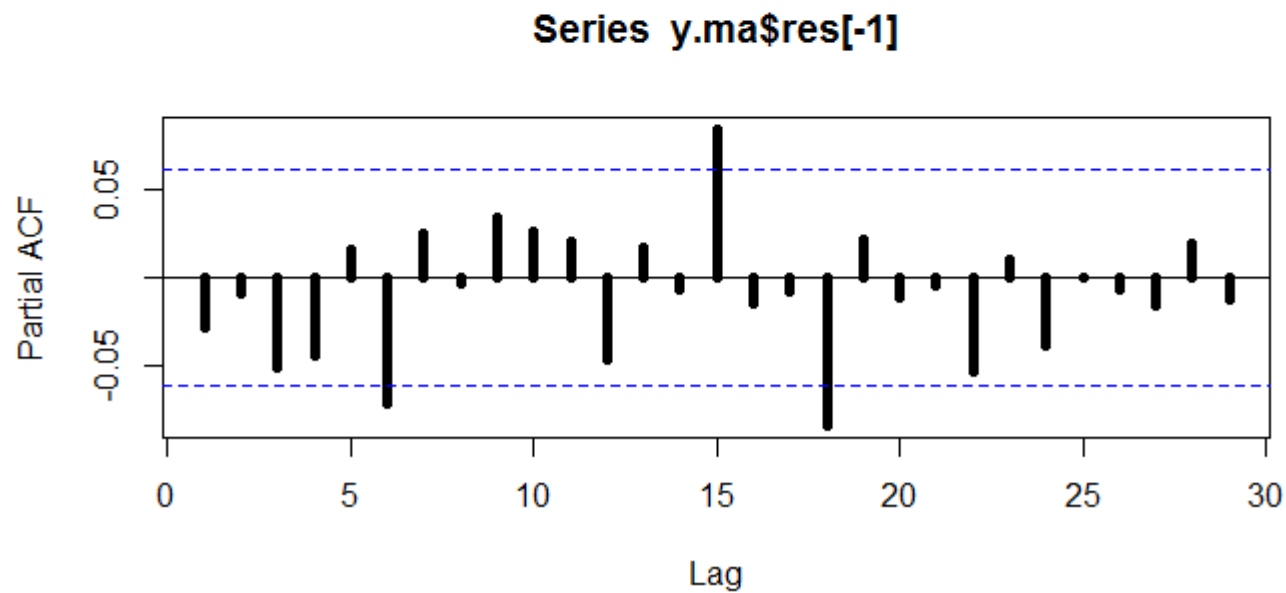
	ma1	ma2	ma3
	0.7899	0.5666	0.396
s.e.	0.0306	0.0351	0.032

```
sigma^2 estimated as 1.068: log likelihood = -  
1452.48, aic = 2912.96
```

```
layout(1:2)
acf(y.ma$res[-1])
pacf(y.ma$res[-1])
```



?Interprétez?



Modélisation du taux de change £GBP - \$NZ

```
Z<- read.table("C:/Mon_dossier/pounds_nz.dat",header=TRUE)
Z.ts<-ts(Z, start=1991,freq=4)
```

```
Z.ma<-arima(Z.ts,order=c(0,0,1))
Z.ma
```

```
Call:
arima(x = Z.ts, order = c(0, 0, 1))
```

Coefficients:

	ma1	intercept
	1.000	2.8329
s.e.	0.072	0.0646

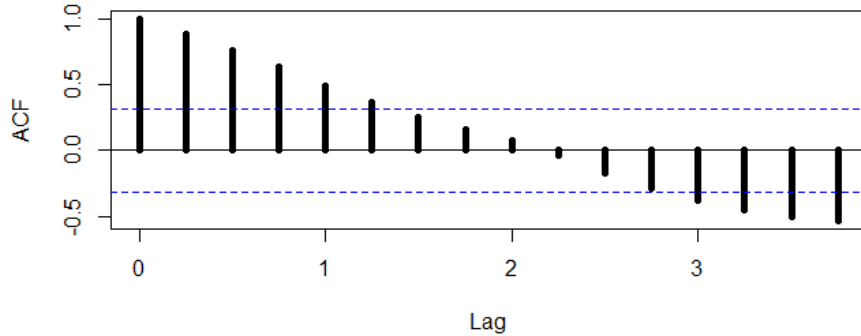
```
sigma^2 estimated as 0.04172:log likelihood= 4.76,aic= -3.53
```

```
layout(1:2)
acf(Z.ts)
pacf(Z.ts)
```

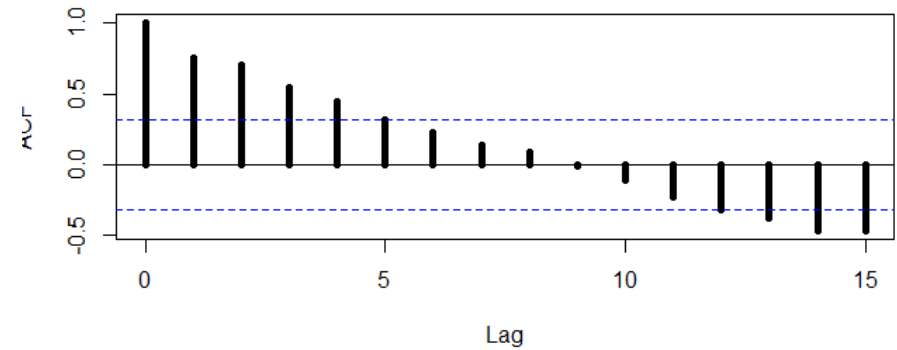
```
layout(1:2)
acf(Z.ma$res[-1])
pacf(Z.ma$res[-1])
```

?Interprétez?

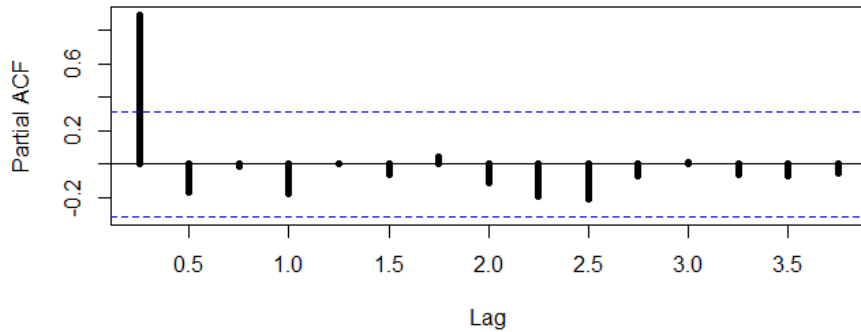
xrate



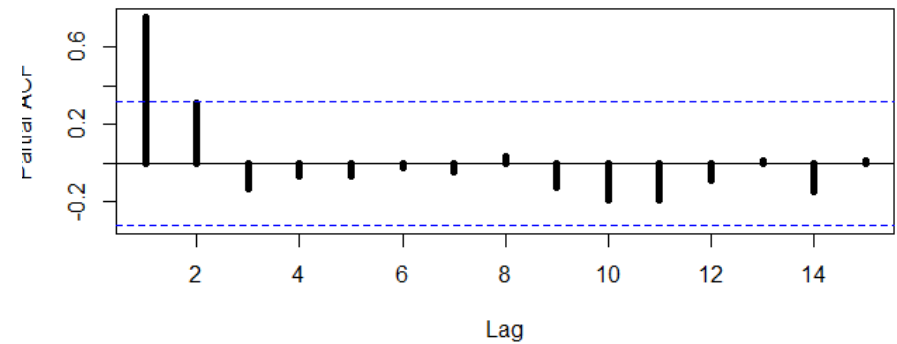
Series Z.ma\$res[-1]



Series Z.ts



Series Z.ma\$res[-1]

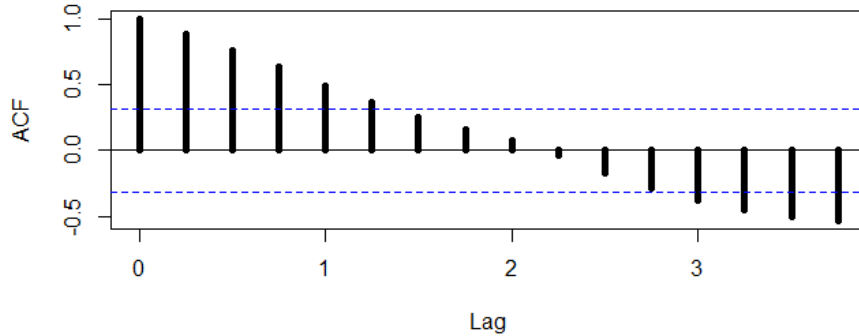


```
layout(1:2)
acf(Z.ts)
pacf(Z.ts)
```

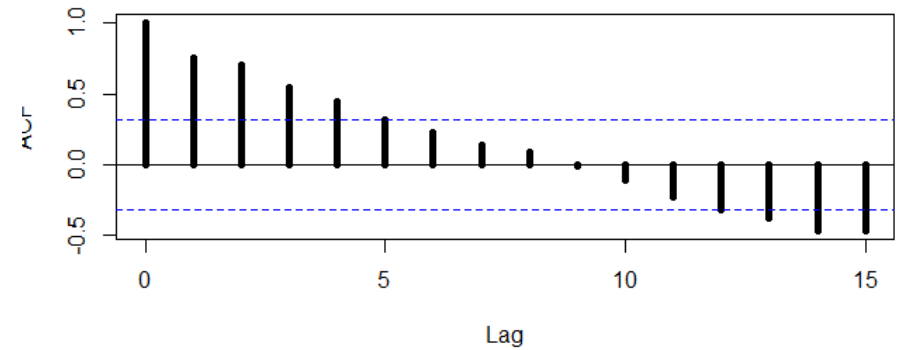
```
layout(1:2)
acf(Z.ma$res[-1])
pacf(Z.ma$res[-1])
```

?Interprétez?

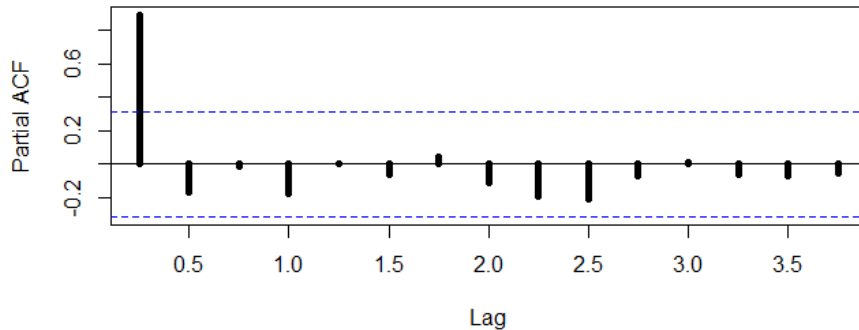
xrate



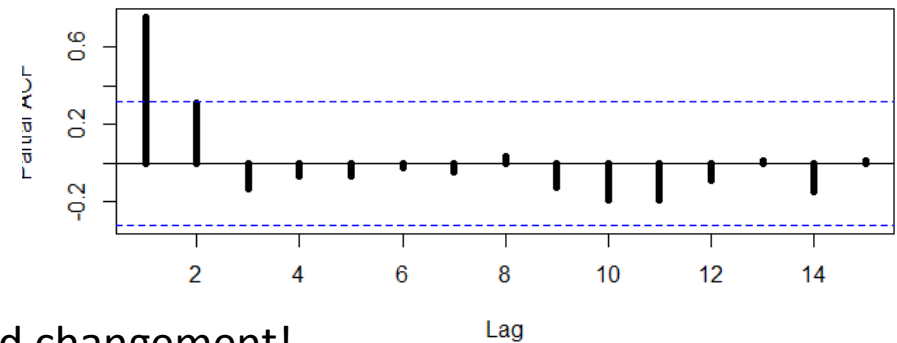
Series Z.ma\$res[-1]



Series Z.ts



Series Z.ma\$res[-1]



Pas de grand changement!
Mais Z n'était pas stationnaire!

IX.2 Processus ARMA(p,q)

$\forall t,$

$$Y_t = \underbrace{\alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p}}_{\text{AR}(p)} + \underbrace{\varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}}_{\text{MA}(q)}$$

Se ré-écrit:

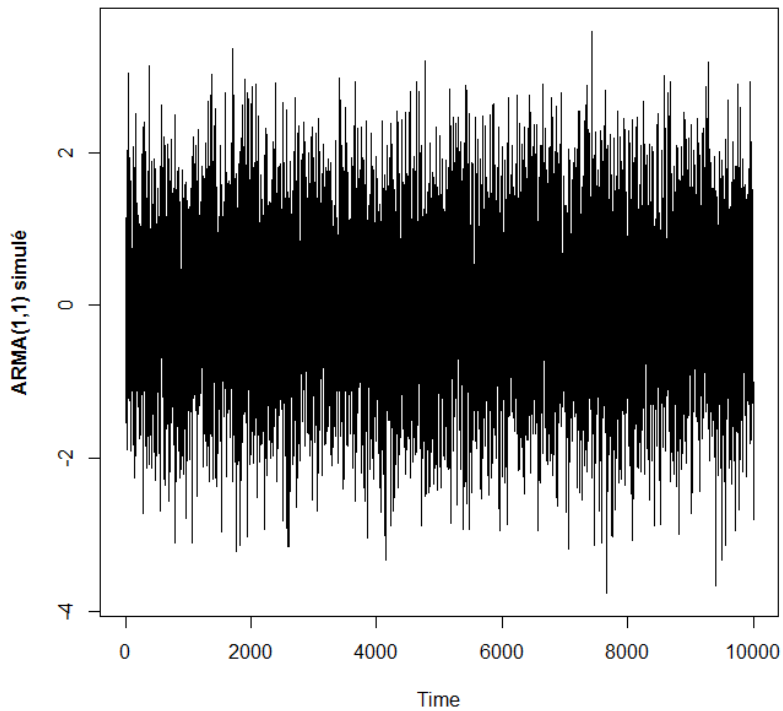
$$\theta_p(\mathbf{B})y_t = \phi_q(\mathbf{B})\varepsilon_t$$

polynômes

Simulation avec la fonction `arima.sim()`

```
y<-arima.sim(n=10000,list(ar=-0.6, ma=0.5))
```

```
ts.plot(y)
```

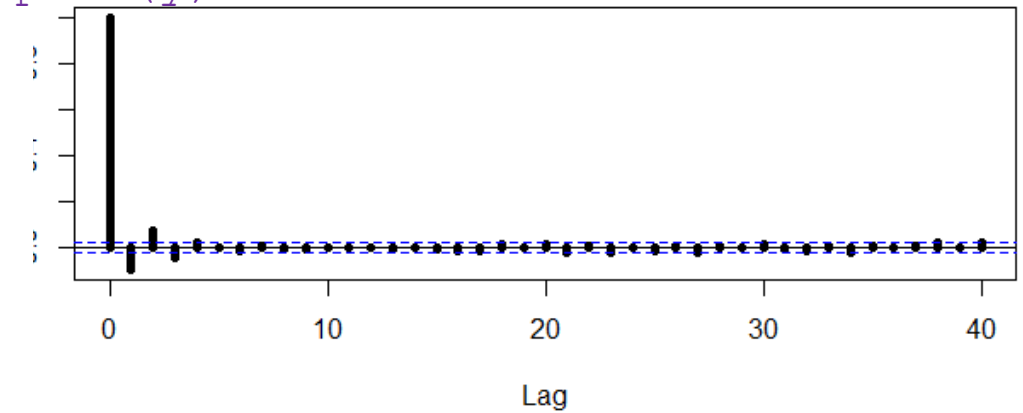


```
layout(1:2)
```

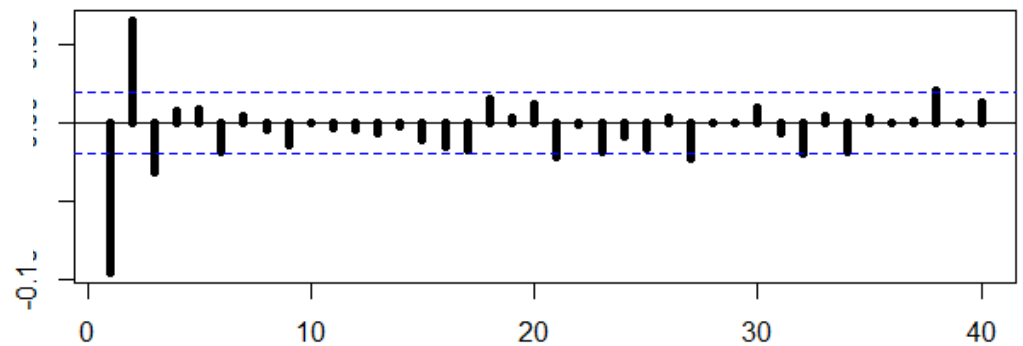
```
acf(y)
```

```
pacf(y)
```

Series y



Series y



Modélisation de ARMA(1,1) simulé

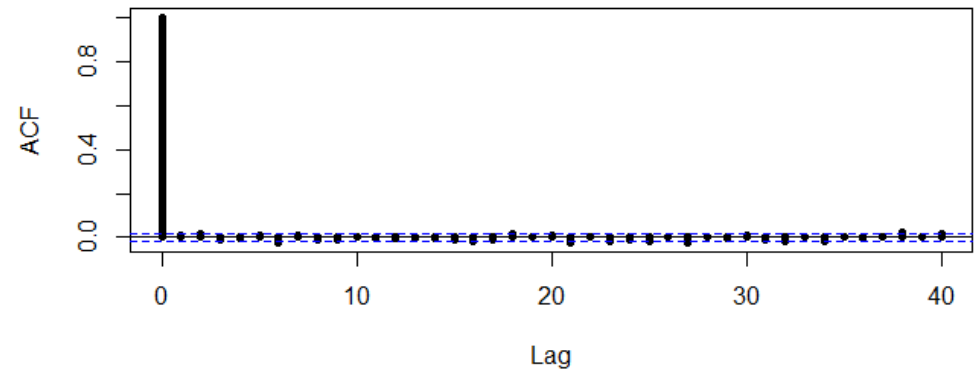
```
y.arma<-arima(y,order=c(1,0,1))  
y.arma
```

Coefficients:

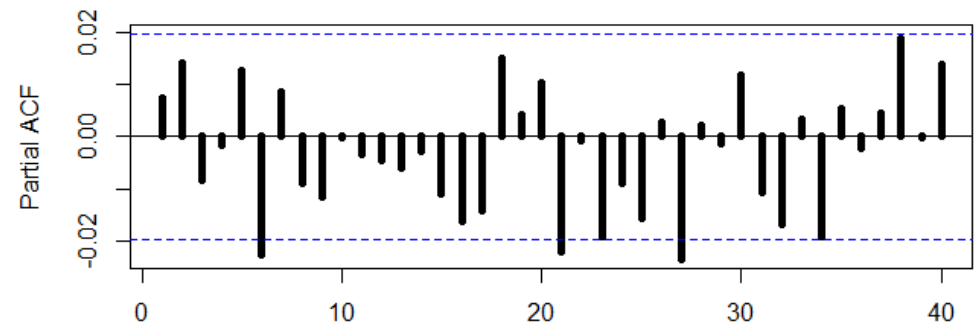
	ar1	ma1	intercept
	-0.5970	0.5027	-0.0066
s.e.	0.0494	0.0530	0.0095

```
layout(1:2)  
acf(y.arma$res[-1])  
pacf(y.arma$res[-1])
```

Series y.arma\$res

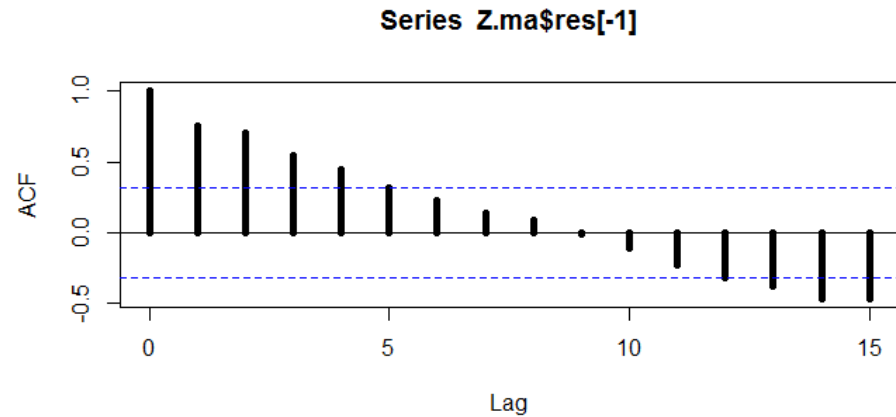


Series y.arma\$res

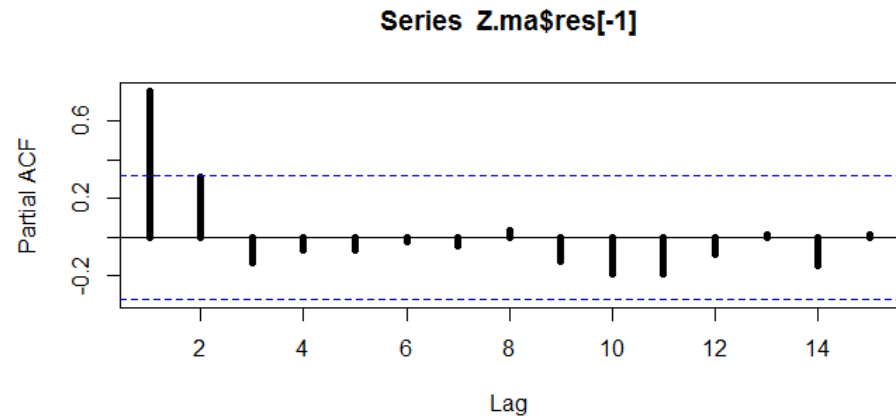


Modélisation de ARMA(p,q) du taux de change (ex. 5)

```
MA(1)  Z.ma<-arima(Z.ts,order=c(0,0,1))
        AIC(Z.ma)           [1] -3.526895
        layout(1:2)
        acf(Z.ma$res[-1])
        pacf(Z.ma$res[-1])
```

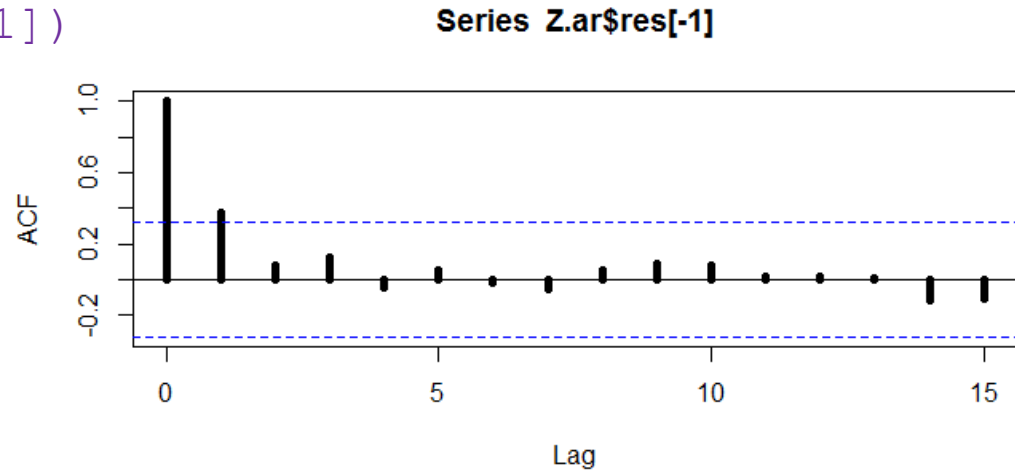


?Interprétez?

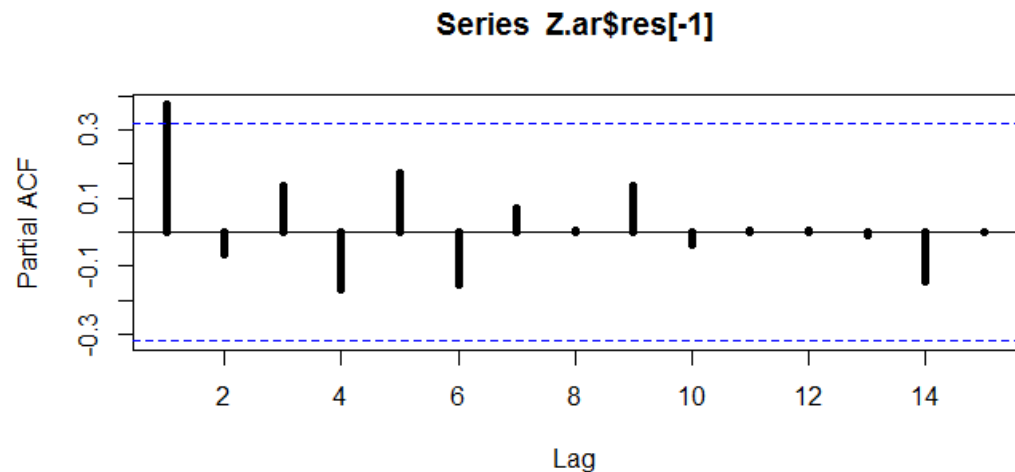


AR(1)

```
Z.ar<-arima(Z.ts,order=c(1,0,0))  
AIC(Z.ar) [1] -37.40417  
layout(1:2)  
acf(Z.ar$res[-1])  
pacf(Z.ar$res[-1])
```

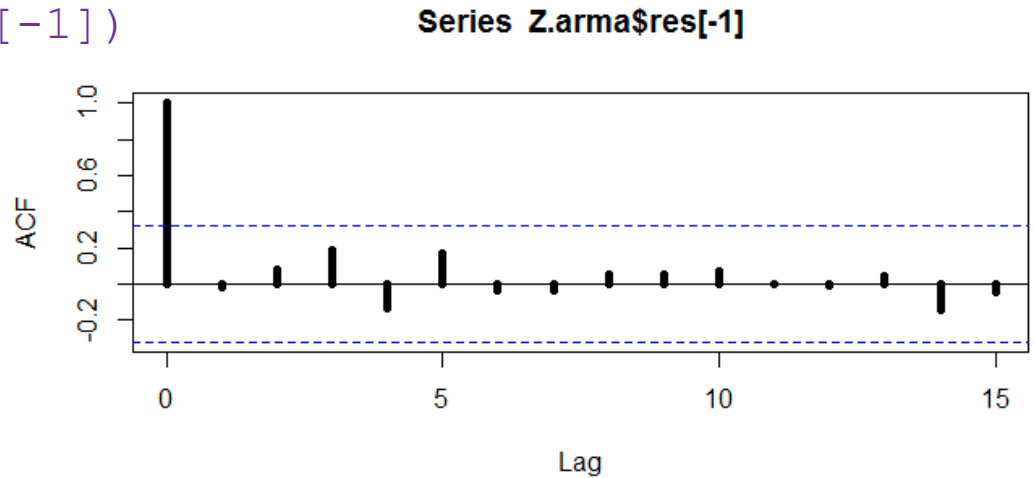


?Interprétez?

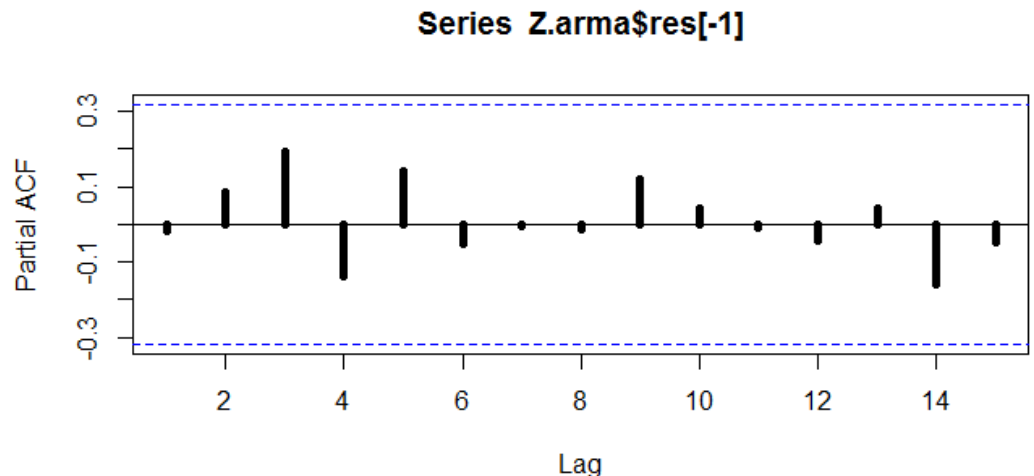


ARMA(1,0,1)

```
Z.arma<-arima(Z.ts,order=c(1,0,1))  
AIC(Z.arma) [1] -42.27357  
layout(1:2)  
acf(Z.arma$res[-1])  
pacf(Z.arma$res[-1])
```



?Interprétez?



Modélisation de la consommation électrique (ex. 4)

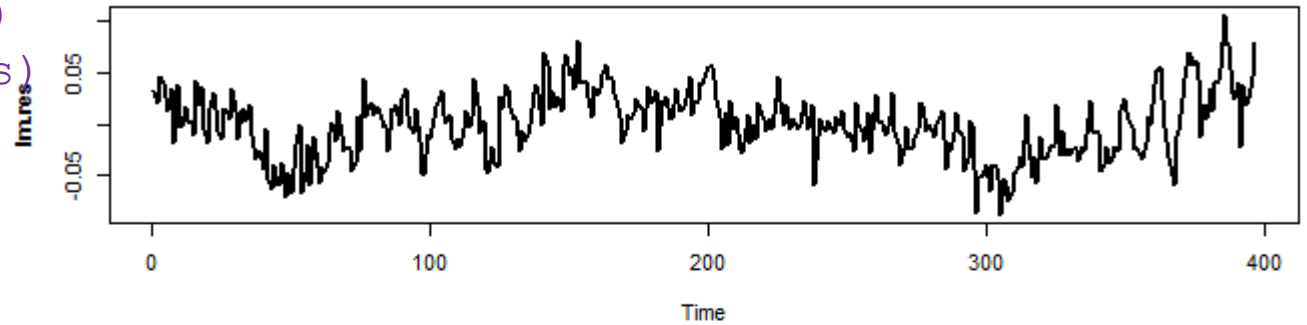
```
Elect.ts<-ts(CBE[,3],start=1958,freq=12)
Time<- 1:length(Elect.ts)
mth<-cycle(Elect.ts)
```

```
Elect.lm<-lm(log(Elect.ts)~Time +I(Time^2)+factor(mth))
```

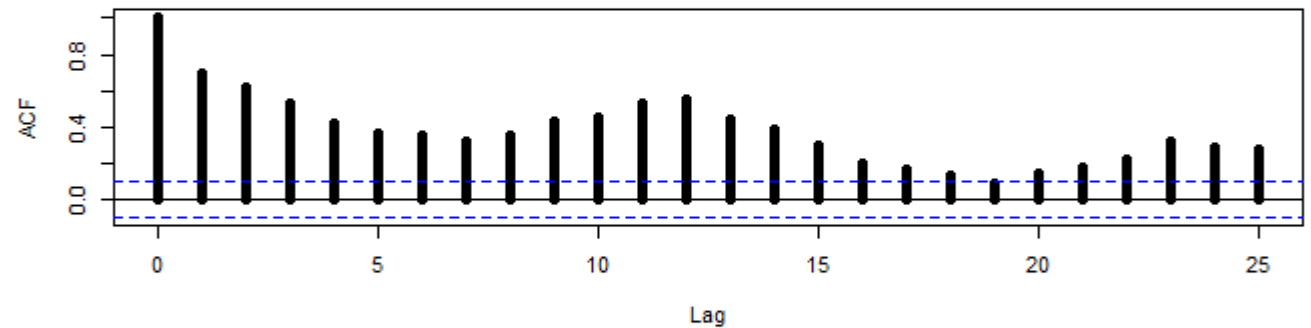
```
summary(Elect.lm)
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	7.271e+00	7.130e-03	1019.786	< 2e-16	***
Time	7.960e-03	5.603e-05	142.070	< 2e-16	***
I(Time^2)	-6.883e-06	1.367e-07	-50.359	< 2e-16	***
factor(mth) 2	-1.991e-02	7.826e-03	-2.544	0.01134	*
factor(mth) 3	6.598e-02	7.826e-03	8.430	7.14e-16	***
factor(mth) 4	3.288e-02	7.826e-03	4.201	3.31e-05	***
factor(mth) 5	1.462e-01	7.826e-03	18.687	< 2e-16	***
factor(mth) 6	1.777e-01	7.826e-03	22.711	< 2e-16	***
factor(mth) 7	2.375e-01	7.826e-03	30.344	< 2e-16	***
factor(mth) 8	1.994e-01	7.827e-03	25.477	< 2e-16	***
factor(mth) 9	1.074e-01	7.827e-03	13.719	< 2e-16	***
factor(mth) 10	9.044e-02	7.827e-03	11.555	< 2e-16	***
factor(mth) 11	4.278e-02	7.827e-03	5.465	8.35e-08	***
factor(mth) 12	2.350e-02	7.828e-03	3.002	0.00286	**

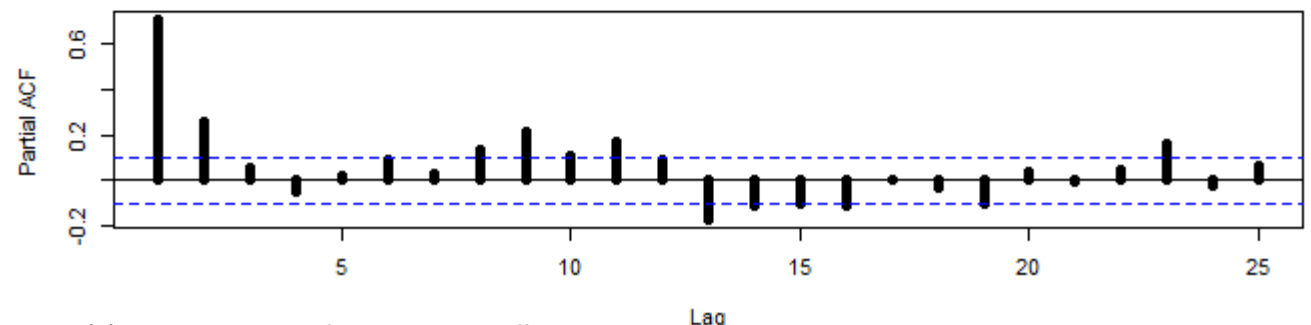
```
layout(1:3)
ts.plot(Elect.lm$res)
acf(Elect.lm$res)
pacf(Elect.lm$res)
```



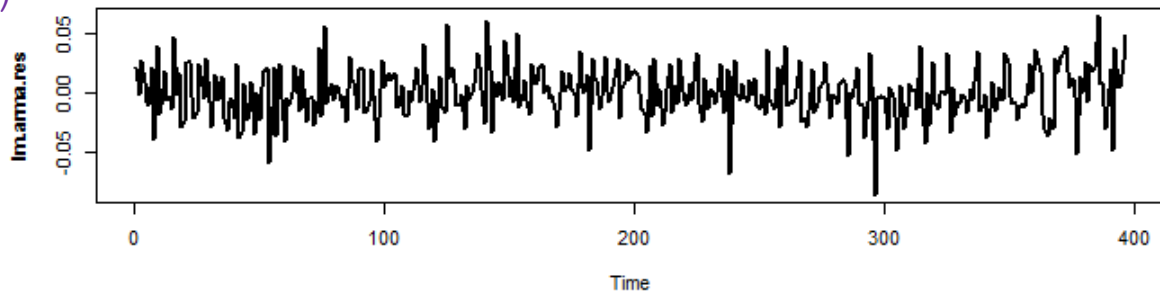
Series Elect.lm\$res



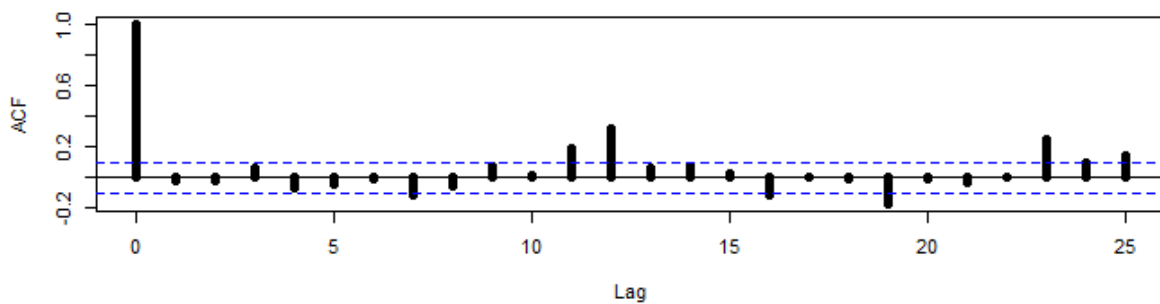
Series Elect.lm\$res



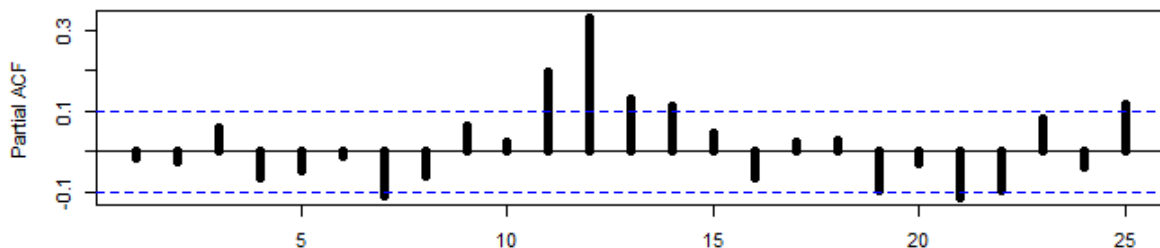

```
Elect.arma<-arima(Elect.lm$res,order=c(2,0,0))  
layout(1:3)  
ts.plot(Elect.arma$res)  
acf(Elect.arma$res)  
pacf(Elect.arma$res)
```



Series Elect.arma\$res



Series Elect.arma\$res



X. Modèles non-stationnaires : ARIMA

En général, séries non stationnaires: tendance, effet saison,...

ex: une marche aléatoire n'est pas stationnaire

=> Stationnarisation (ex marche aléatoire : différenciation d'ordre 1)

Intégration de différenciation dans un modèle ARMA:

AR Intergrated MA = ARIMA(p,d,q)

$$\theta_p(\mathbf{B})(1-\mathbf{B})^d y_t = \phi_q(\mathbf{B})\varepsilon_t$$

Modéliser la saison dans un modèle ARIMA:

= processus SARIMA(p,d,q)(P,D,Q)_s

Box-Jenkins 1970

$$\Theta_P(\mathbf{B}^s)\theta_p(\mathbf{B})(1-\mathbf{B}^s)^D(1-\mathbf{B})^d y_t = \Phi_Q(\mathbf{B}^s)\phi_q(\mathbf{B})\varepsilon_t$$

```
arima(object.ts, order=c(p, d, q), seas=list(c(ps, ds, qs), freq))
```

Remarque: modélisation d'une tendance déterministe / stochastique

Différencier une série $\{y_t\} \Rightarrow$ ajuste sur la tendance \forall déterministe ou stochastique

• Marche Aléatoire:

$$y_t = y_{t-1} + \varepsilon_t$$

$$\Leftrightarrow \nabla^{(1)} y_t = y_t - y_{t-1} = \varepsilon_t \quad \text{Bruit Blanc} \Rightarrow \text{stationnaire}$$

• Tendance linéaire:

$$y_t = a + bt + \varepsilon_t$$

$$\Leftrightarrow \nabla^{(1)} y_t = y_t - y_{t-1}$$

$$= a + bt - (a + b(t-1)) + (\varepsilon_t - \varepsilon_{t-1})$$

$$= b + \varepsilon_t - \varepsilon_{t-1}$$

MA(1) stationnaire

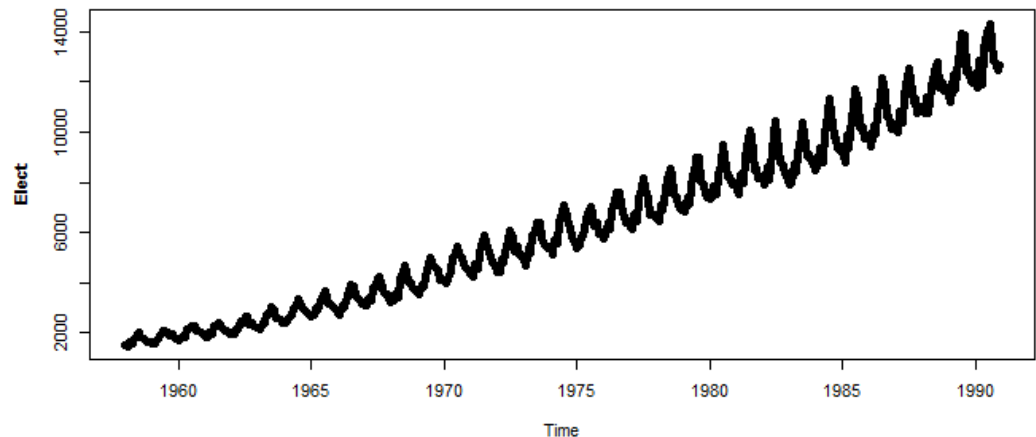
On peut utiliser `arima()` pour ajuster une tendance linéaire déterministe, ou la modéliser par un modèle linéaire préalable (dans ce cas `order(xx, 0, xx)` et `include.mean=FALSE`)

Différenciation: Reprise Exemple 4: consommation en Australie d'électricité

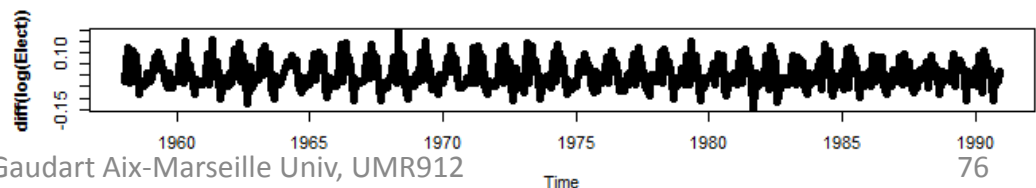
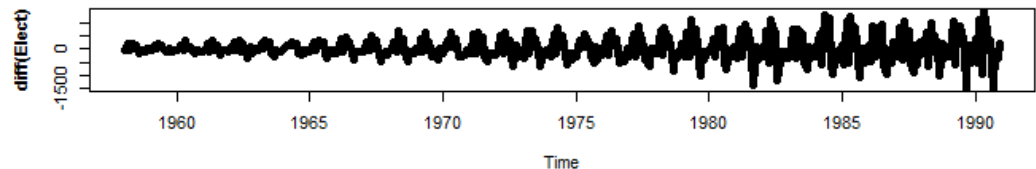
Fonction `diff(object.ts, d=1, lag=1)`

```
CBE<- read.table("C:/Mon_Dossier/cbe.dat",header=TRUE)
Elect.ts<-ts(CBE[,3],start=1958, freq=12)
```

```
layout(c(1,1,2,3))
plot(Elect.ts)
plot(diff(Elect.ts))
plot(diff(log(Elect.ts)))
```

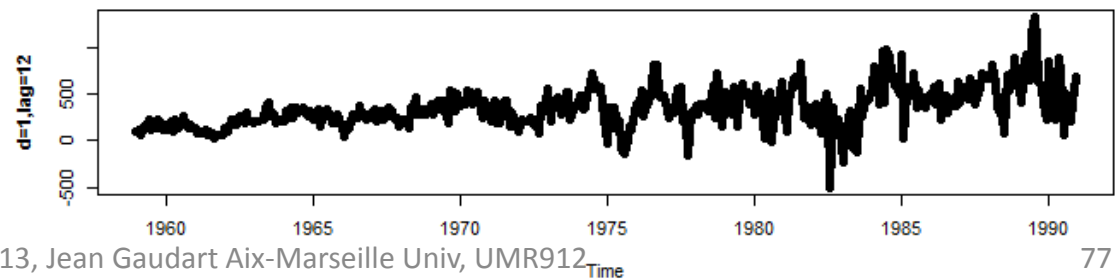
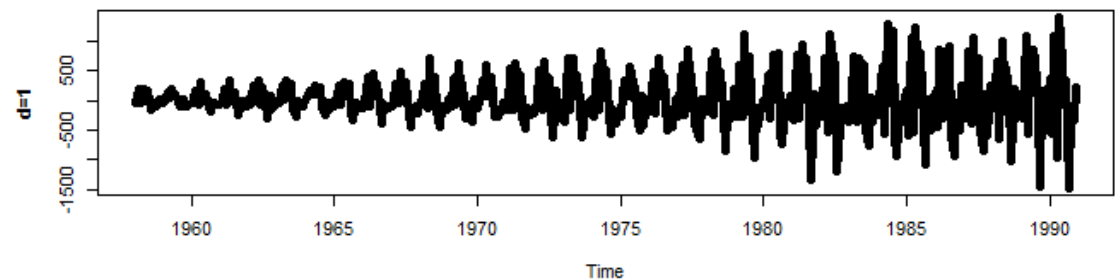
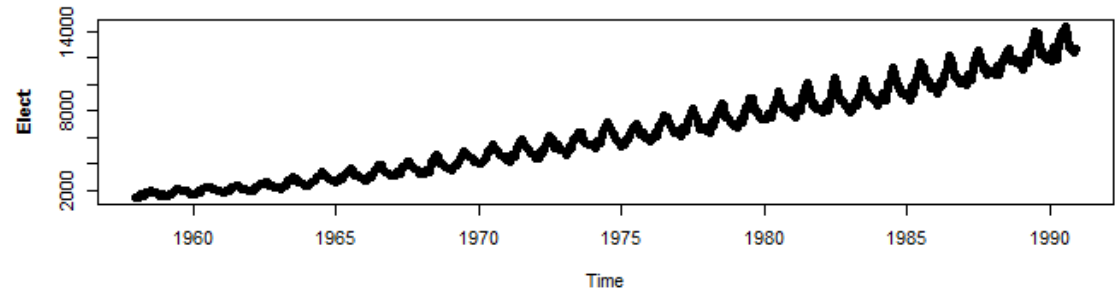


?Interprétez?



Différencier à $d=1$ et $\text{lag}=12$ ajuste sur une tendance linéaire et une saisonnalité additive

```
layout(c(1,1,2,3))  
plot(Elect.ts)  
plot(diff(Elect.ts))  
plot(diff(Elect.ts,d=1,lag=12))
```



Simulation d'un Modèle ARIMA(1,1,1)

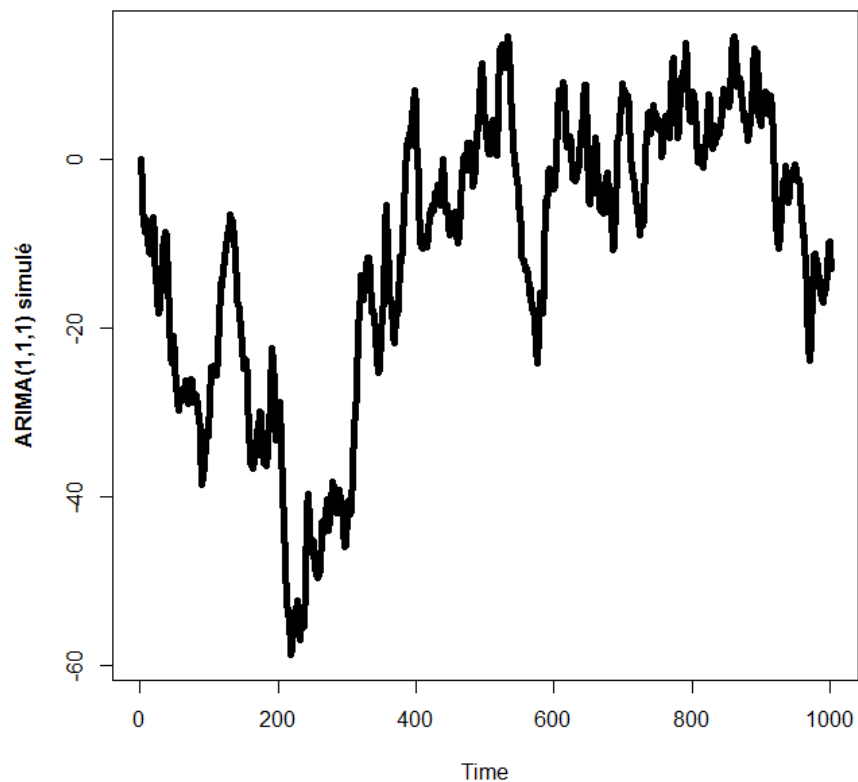
```
y<-arima.sim(n=1000,list(order = c(1,1,1),ar=0.5,ma=0.3))
```

```
ts.plot(y)
```

```
layout(1:2)
```

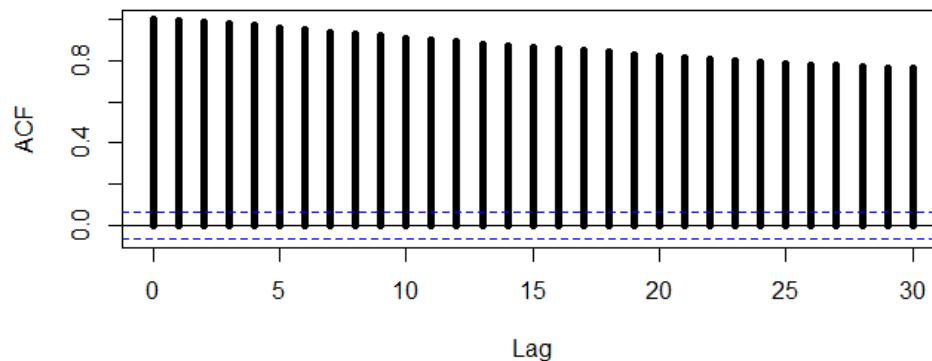
```
acf(y)
```

```
pacf(y)
```

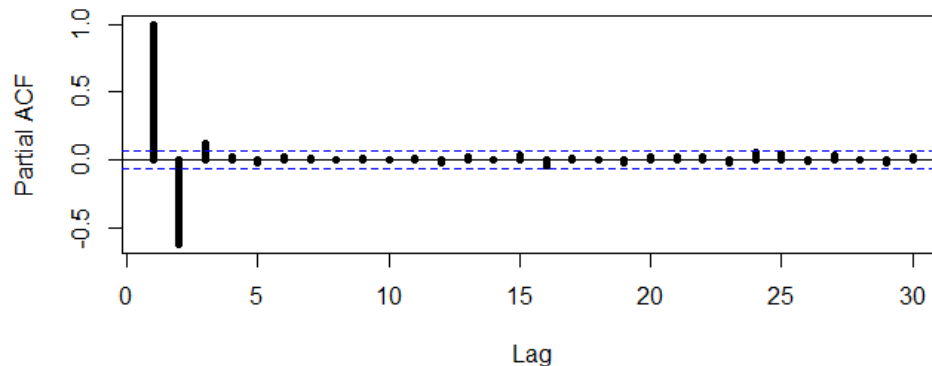


?Interprétez?

Series ts(y)



Series ts(y)



Estimation d'un Modèle ARIMA(1,1,1) sur la série simulée

```
y.arima<-arima(y,order = c(1,1,1))
```

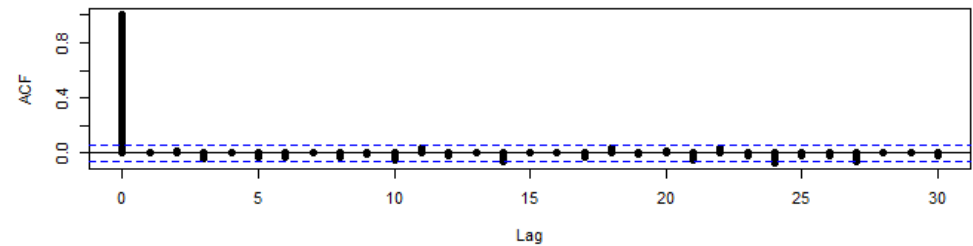
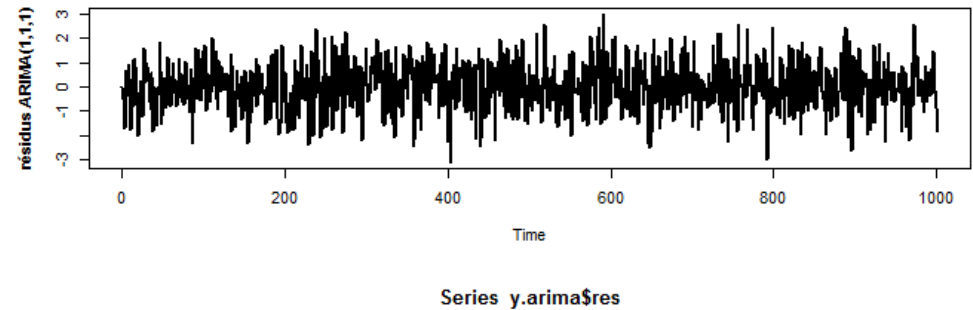
```
y.arima
```

Coefficients:

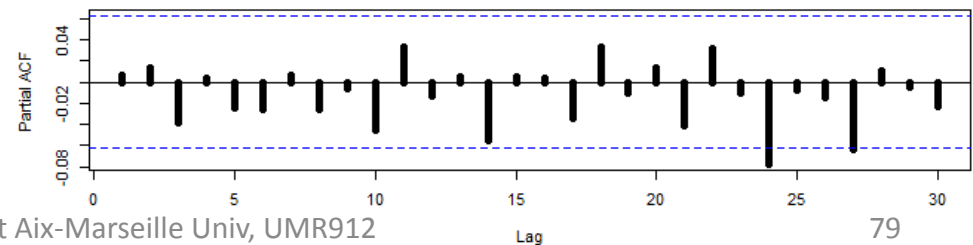
	ar1	ma1
	0.5269	0.2657
s.e.	0.0375	0.0412

sigma² estimated as 0.9766:
log likelihood = -1407.45,
aic = 2820.9

```
layout(1:3)  
ts.plot(y.arima$res)  
acf(y.arima$res)  
pacf(y.arima$res)
```



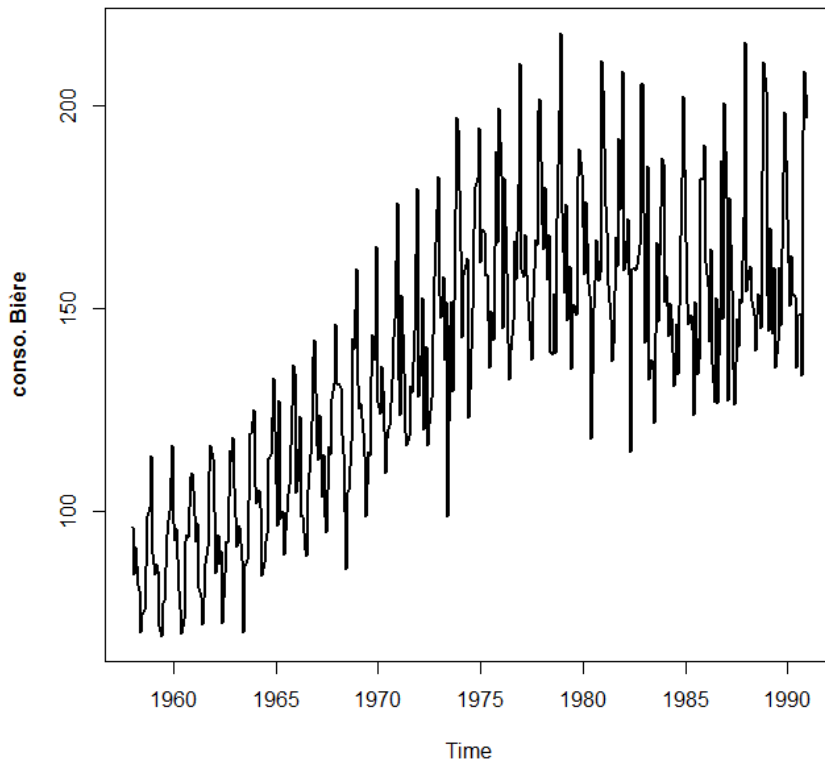
?Interprétez?



Estimation d'un Modèle ARIMA, reprise exemple 4 consommation de Bière

```
B.ts<-ts(CBE[,2],start=1958, freq=12)
```

```
ts.plot(B.ts)
```



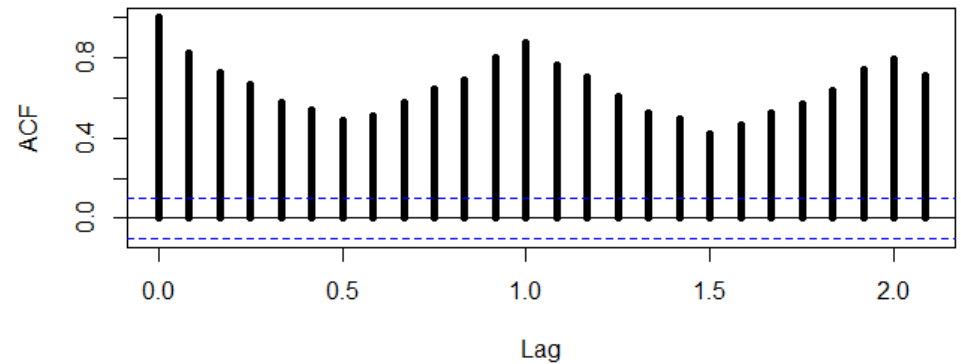
Tendance => Diff

Mais diff(1) introduit un MA(1)

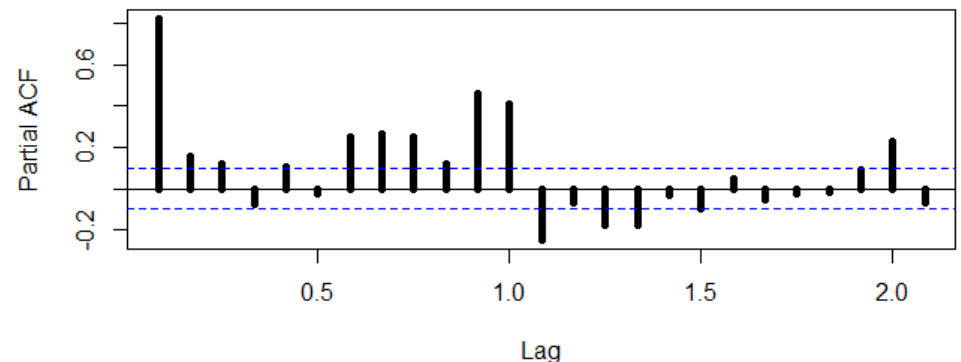
```
layout(1:2)
```

```
acf(B.ts)
```

```
pacf(B.ts) Series B.ts
```



Series B.ts

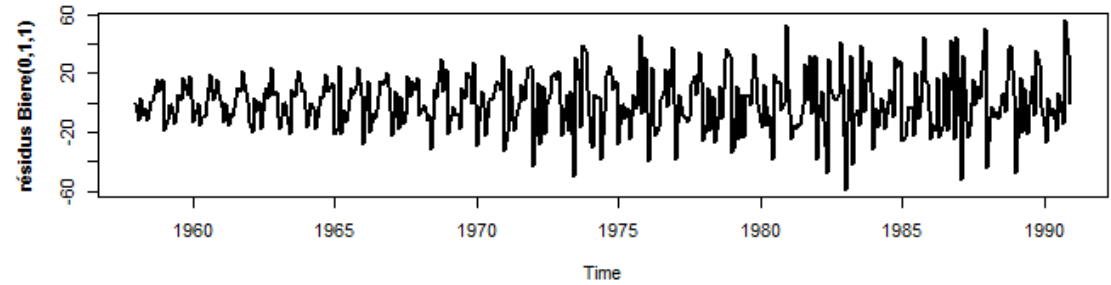



```
B.ima<-arima(B.ts,order=c(0,1,1))
```

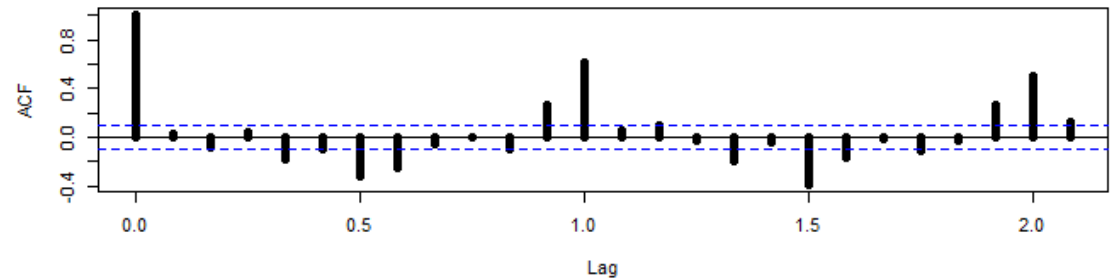
```
B.ima
```

```
Coefficients:  
      ma1  
      -0.3334  
s.e.      0.0558
```

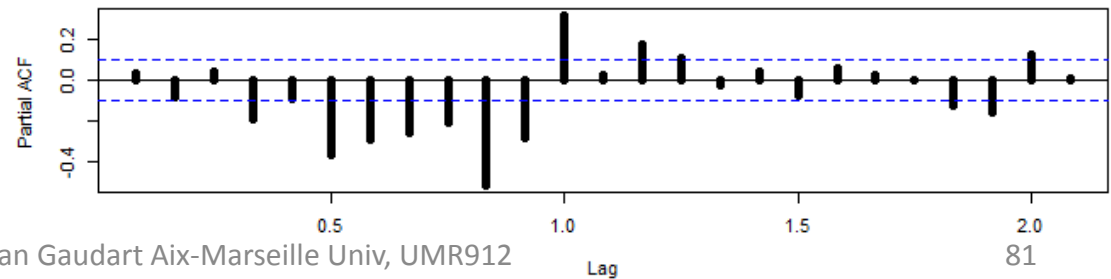
```
layout(1:3)  
ts.plot(B.ima$res)  
acf(B.ima$res)  
pacf(B.ima$res)
```



Series B.ima\$res



Series B.ima\$res



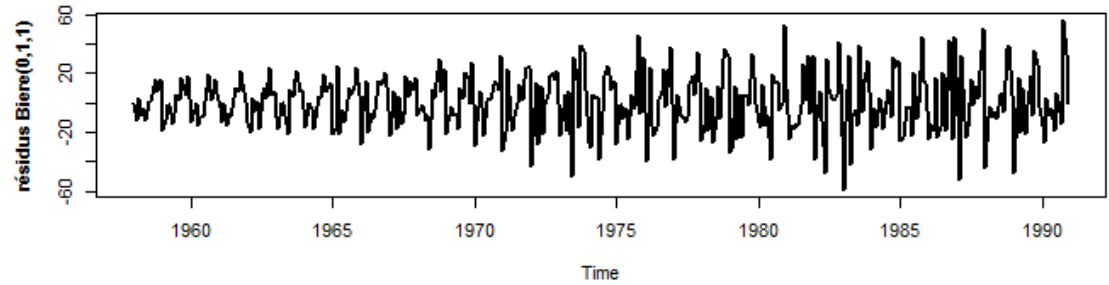
?Interprétez?

```
B.ima<-arima(B.ts,order=c(0,1,1))
```

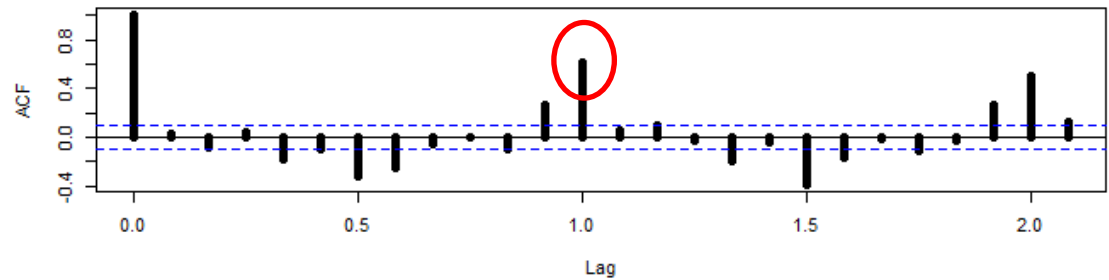
```
B.ima
```

```
Coefficients:  
      ma1  
      -0.3334  
s.e.      0.0558
```

```
layout(1:3)  
ts.plot(B.ima$res)  
acf(B.ima$res)  
pacf(B.ima$res)
```

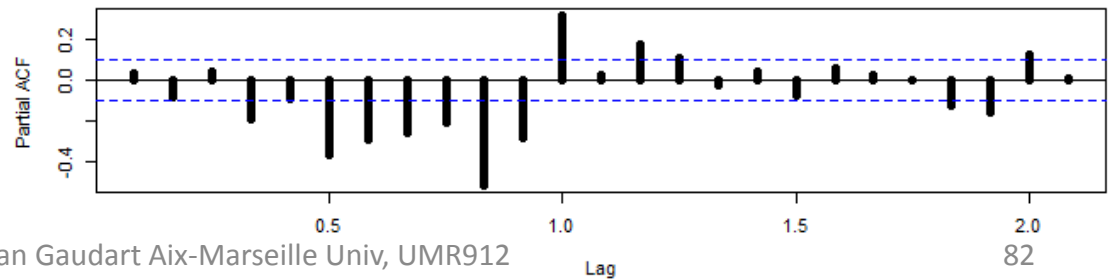


Series B.ima\$res



?Interprétez?

Series B.ima\$res



Modèle SARIMA(p,d,q)(P,D,Q)_s

Modélisation souvent difficile car nombreux paramètres à définir
Comment choisir? Un des critères : $AIC = -2 * L + (\ln(n) + 1)p$

NB: ne pas créer plus d'Autocorr. qu'initialement

Ex. 4 électricité

```
AIC(arima(log(Elect.ts), order=c(1, 1, 0),  
+seas=list(order=c(1, 0, 0), 12)))
```

```
[1] -1764.741
```

```
AIC(arima(log(Elect.ts), order=c(0, 1, 1),  
+seas=list(order=c(0, 0, 1), 12)))
```

```
[1] -1361.586
```

programme Cowpertwait de sélection de modèle (get.best.arima.R) (à télécharger)

```
get.best.arima<-function(x.ts,maxord=c(1,1,1,1,1,1))
{
  best.aic<-1e8
  n <- length(x.ts)
  for(p in 0:maxord[1]) for(d in 0:maxord[2])
  for(q in 0:maxord[3]) for(P in 0:maxord[4])
  for(D in 0:maxord[5]) for (Q in 0:maxord[6])
  {
    fit <- arima(x.ts,order=c(p,d,q),
                 seas=list(order=c(P,D,Q), frequency(x.ts)),
                 method="CSS")
    fit.aic <- -2*fit$loglik+(log(n)+1)*length(fit$coef)
    if (fit.aic < best.aic)
    {
      best.aic <- fit.aic
      best.fit <- fit
      best.model <- c(p,d,q,P,D,Q)
    }
  }
  list(best.aic, best.fit, best.model)
}
```

Application à la consommation électrique

```
best.arima.elect<-  
get.best.arima(log(Elect.ts),maxord=c(2,2,2,2,2,2))
```

```
best.arima.elect
```

```
[[1]]  
[1] -1919.025  
[[2]]  
Call:  
arima(x = x.ts, order = c(p, d, q), seasonal = list(order =  
c(P, D, Q), frequency(x.ts)),  
      method = "CSS")  
Coefficients:  
          ma1      sar1      sar2      sma1      sma2  
    -0.6566  0.7315  0.2557  -0.3324  -0.3051  
s.e.    0.0420  0.1270  0.1259   0.1279   0.0958  
sigma^2 estimated as 0.0004161:  part log likelihood = 976.97  
[[3]]  
[1] 0 1 1 2 0 2
```

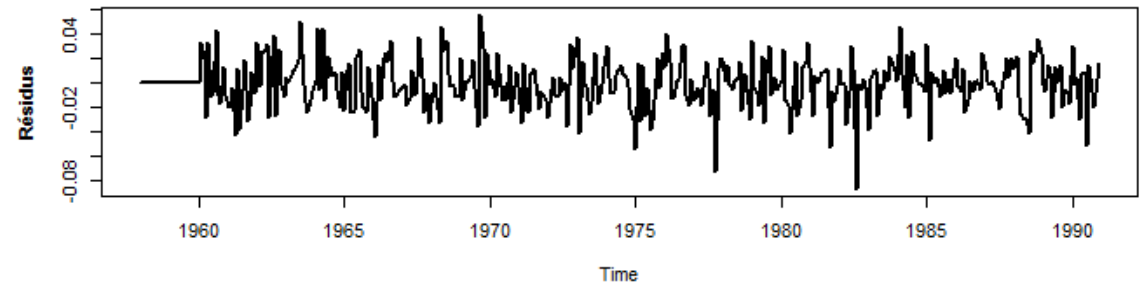
```
best.fit.elect<-best.arima.elect[[2]]
```

```
layout(1:3)
```

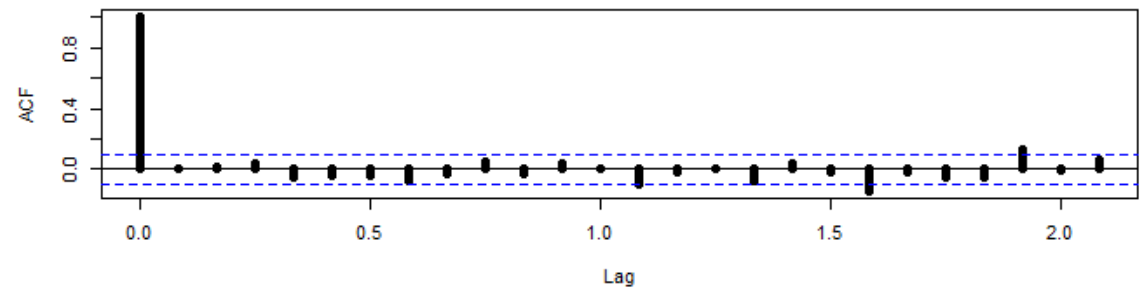
```
ts.plot(best.fit.elect$res)
```

```
acf(best.fit.elect$res)
```

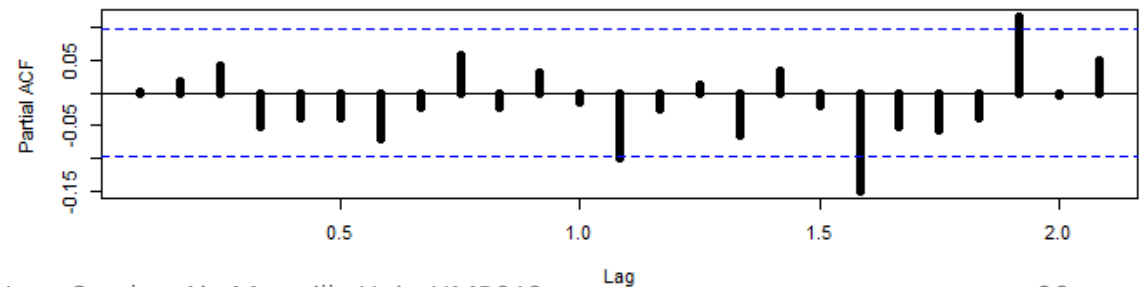
```
pacf(best.fit.elect$res)
```



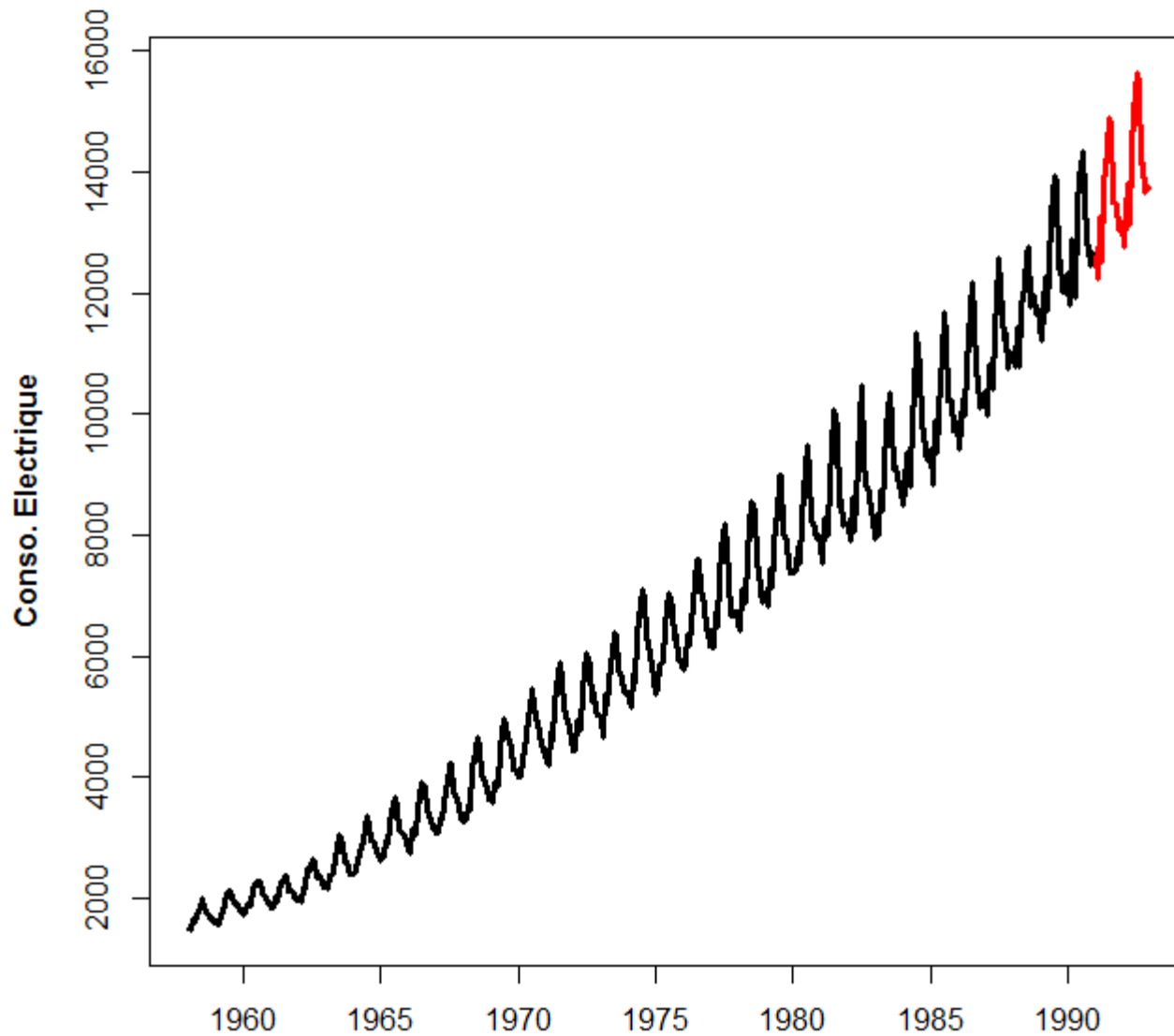
Series best.fit.elect\$res



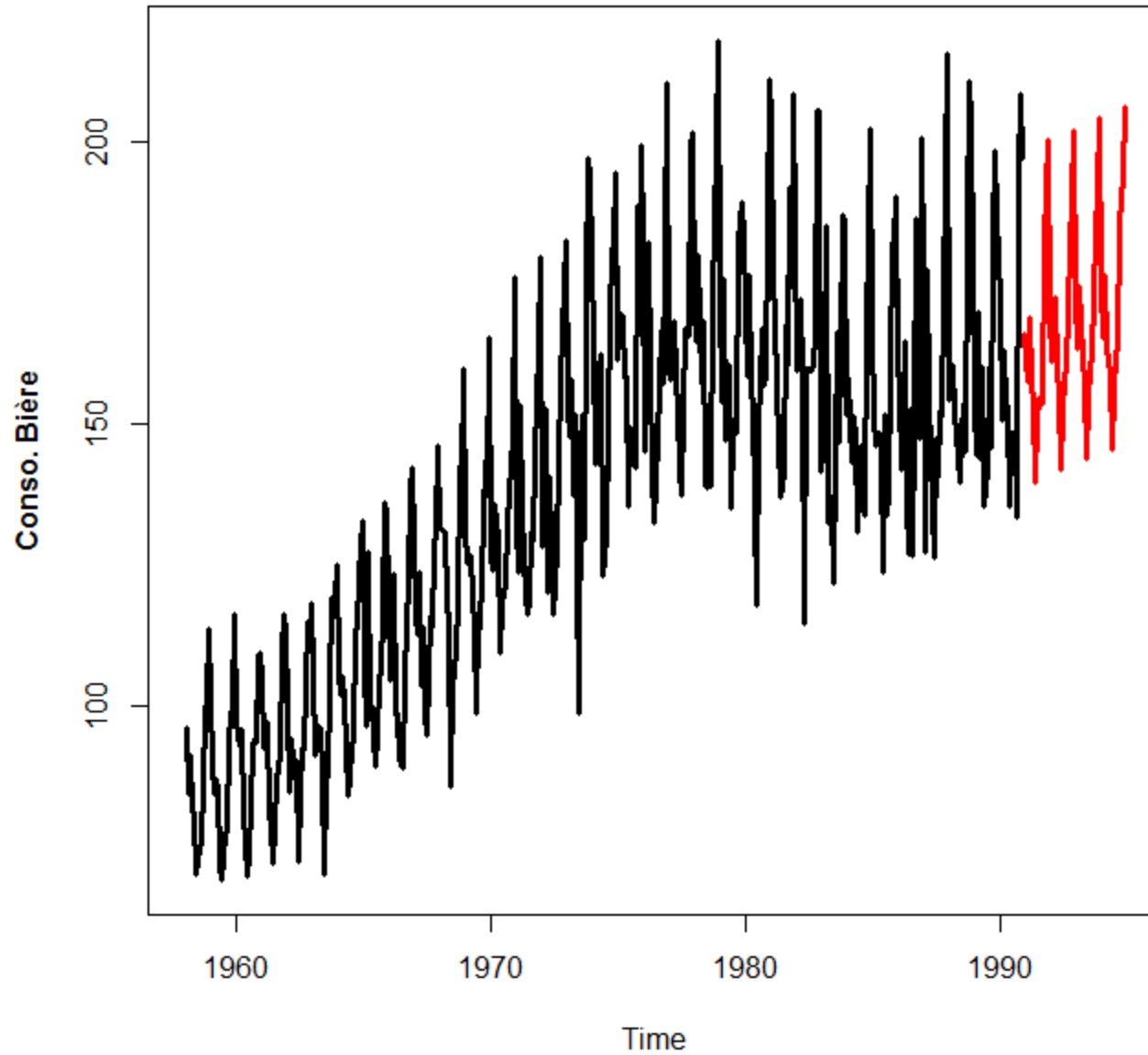
Series best.fit.elect\$res



```
best.predic<-predict(best.fit.elect,24)$pred
ts.plot(Elect.ts, exp(best.predic), col=c("black","red"))
```



Faire l'exercice pour la consommation de bière



Références:

- PSP Cowpertwait, AV Metcalfe. Introductory Time Series with R. Springer: Use R!
- RH Shumway, DS Stoffer. Time Series Analysis and Its Applications. Springer: Texts in Statistics