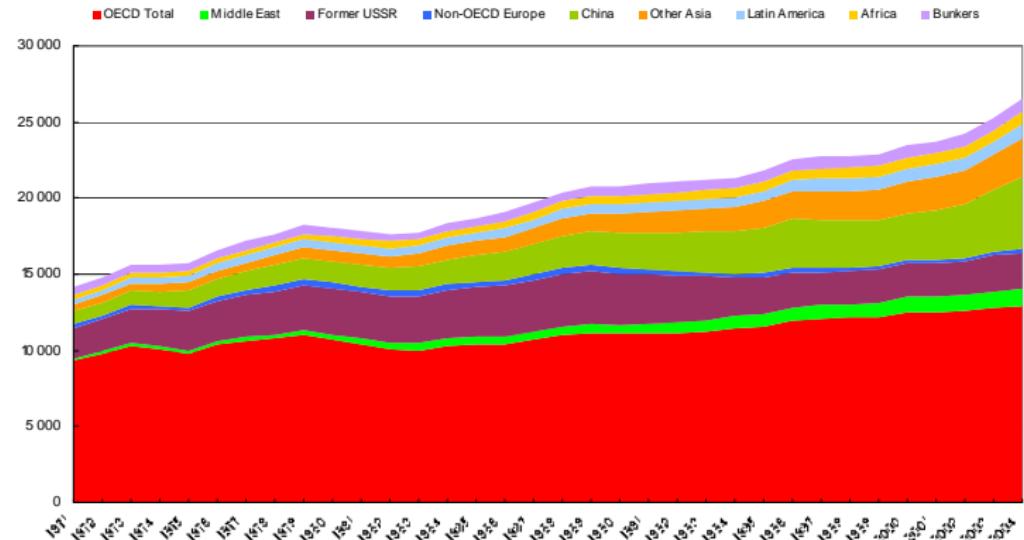


# Introduction aux séries temporelles

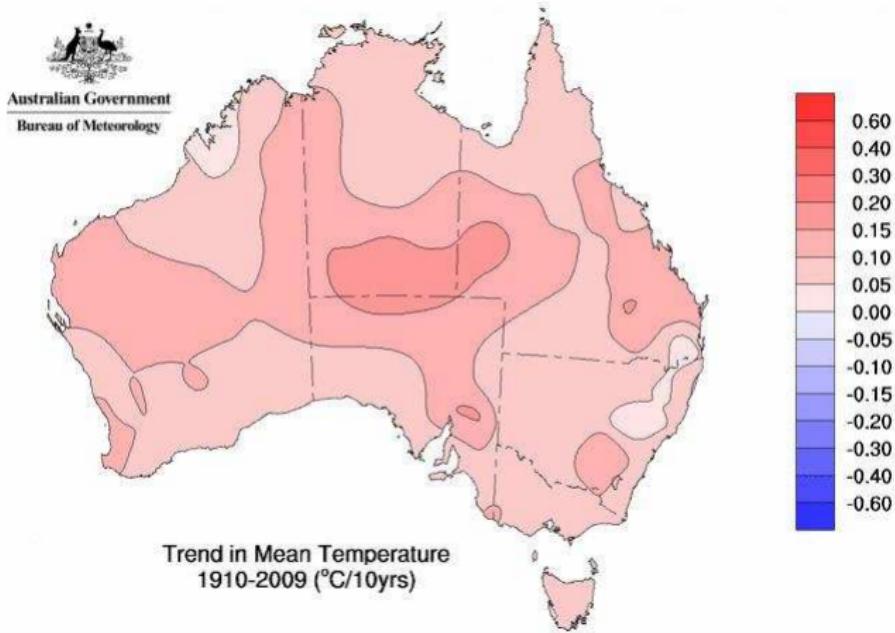
ISMA-2012

# Motivation

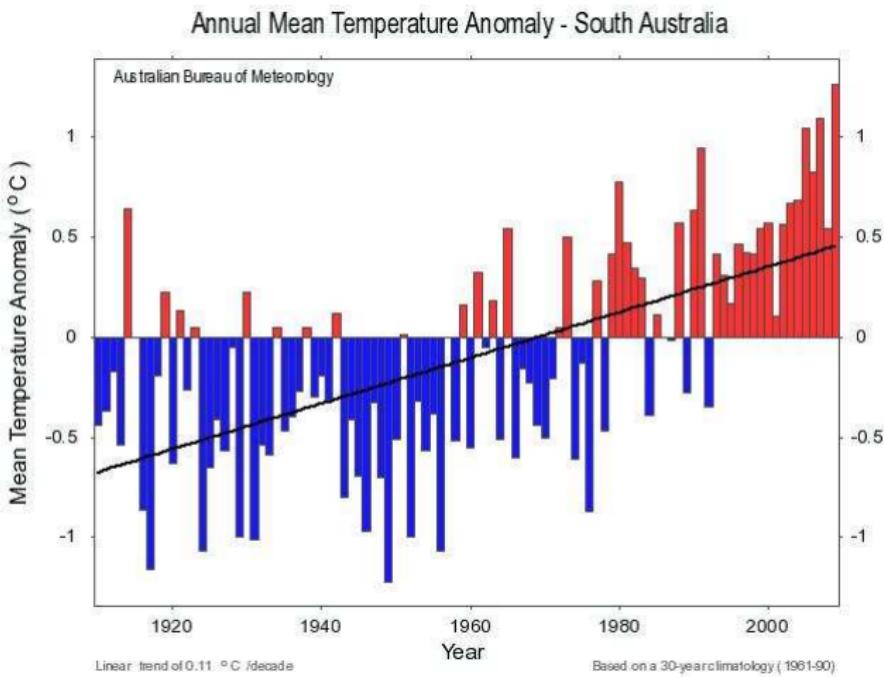


World CO<sub>2</sub> emissions from energy use, by region: Million tonnes

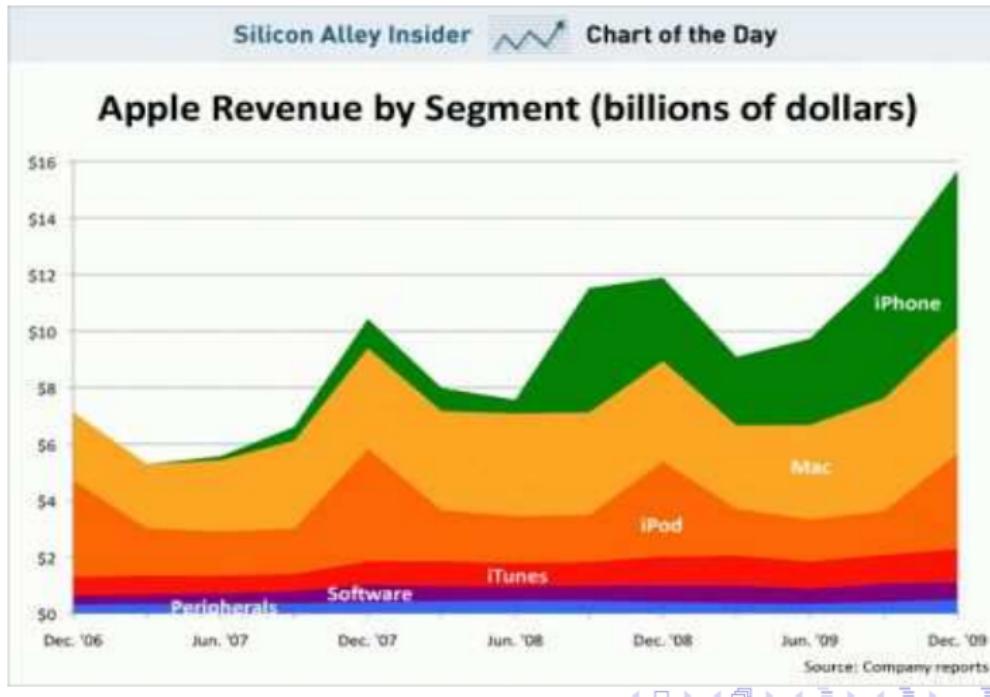
# Motivation



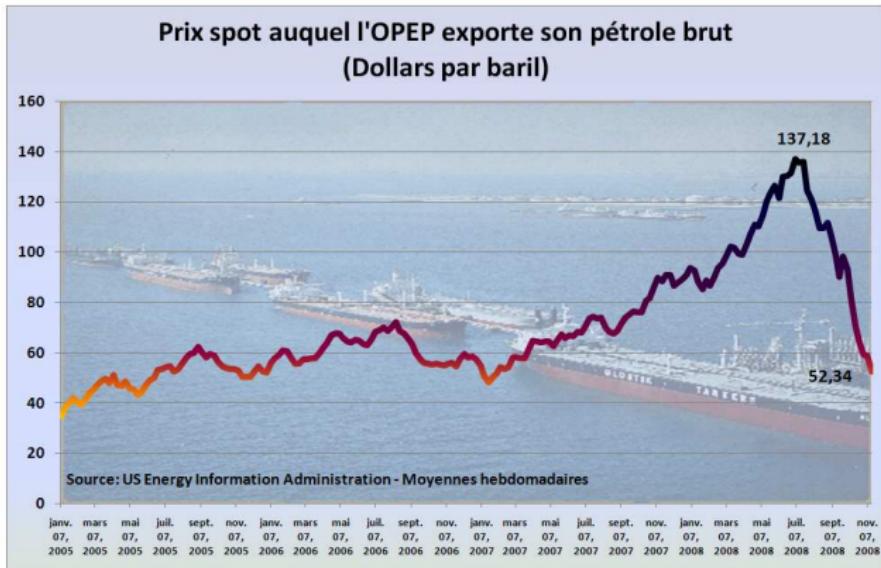
# Motivation



# Motivation



# Motivation



- Apprentissage: Comprendre l'évolution de la série temporelle
- Prévision: Proposer une approximation réaliste des valeurs futures de la série temporelles

- **Apprentissage: Comprendre l'évolution de la série temporelle**
- **Prévision: Proposer une approximation réaliste des valeurs futures de la série temporelles**

- Modèle de régression multiple
- Processus autorégressif à moyenne mobile (ARMA)
- Processus non stationnaire ARIMA
- Processus non stationnaire SARIMA
- Processus VARMA
- Estimation de la densité de probabilité
- Régression non paramétrique

# 1. Modèle de régression multiple

## Modèle

$$y_t = b_1 x_{t,1} + \dots + b_p x_{t,p} + u_t, \quad 1 \leq t \leq T, \quad (1)$$

$u_t$  est un bruit blanc c'est à dire une suite de variables aléatoires réelles telles que:

- (H.1):  $E(u_t) = 0, \quad 1 \leq t \leq T,$

- (H.2):  $E(u_t u_s) = \begin{cases} \sigma^2 & \text{si } t = s \\ 0 & \text{si } t \neq s \end{cases}.$

# 1. Modèle de régression multiple

## Modèle

$$y_t = b_1 x_{t,1} + \dots + b_p x_{t,p} + u_t, \quad 1 \leq t \leq T, \quad (1)$$

$u_t$  est un bruit blanc c'est à dire une suite de variables aléatoires réelles telles que:

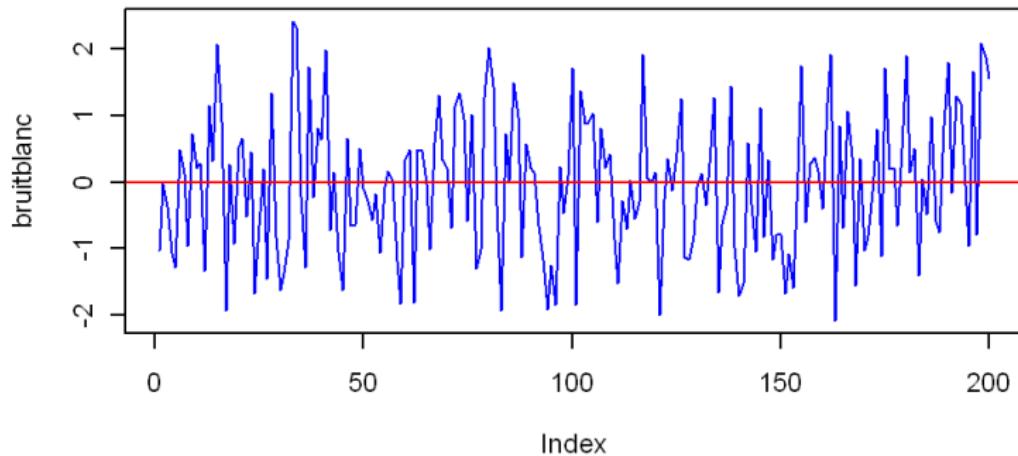
- (H.1):  $E(u_t) = 0, \quad 1 \leq t \leq T,$

- (H.2):  $E(u_t u_s) = \begin{cases} \sigma^2 & \text{si } t = s \\ 0 & \text{si } t \neq s \end{cases}.$

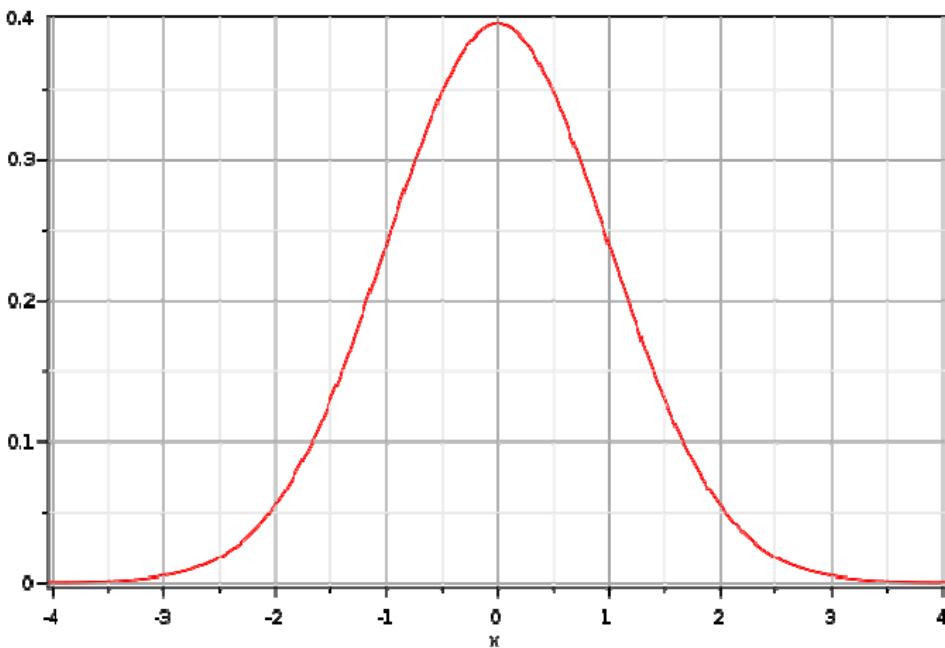
# 1. Modèle de régression multiple

## Commandes sur R

```
bruitblanc <- rnorm(200,0,1)  
plot(bruitblanc,type="l", col="blue")
```



# 1. Modèle de régression multiple



# 1. Modèle de régression multiple

Posons  $b = {}^t(b_1, \dots, b_p)$ ,  $y = {}^t(y_1, \dots, y_T)$ ,  $\mathbf{X}$  la matrice  $(T, p)$  dont la  $j^{ieme}$  colonne est  $x_j = {}^t(x_{1,j}, \dots, x_{T,j})$ , et  $u = {}^t(u_1, \dots, u_T)$ , le système prend alors la forme matricielle suivante:

$$y = \mathbf{X}b + u \quad (2)$$

$x_j, 1 \leq j \leq p$ : **les variables explicatives** (ou facteurs exogènes)

$y$ : **la variable expliquée** (ou facteur endogène),

$u$ : **la perturbation**,

$b$ : **le vecteur des paramètres inconnus.**

# 1. Modèle de régression multiple

	Aléatoire	Déterministe
Observable	$y$	$x_j, 1 \leq j \leq p$
Non observable	$u$	$b$

Estimateur des moindres carrés de  $b$ :

$$\hat{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (3)$$

# 1. Modèle de régression multiple

	Aléatoire	Déterministe
Observable	$y$	$x_j, 1 \leq j \leq p$
Non observable	$u$	$b$

**Estimateur des moindres carrés de  $b$ :**

$$\hat{b} = ({}^t \mathbf{X} \mathbf{X})^{-1} {}^t \mathbf{X} y. \quad (3)$$

# 1. Modèle de régression multiple

## Theorem

i)  $E(\hat{b}) = b$  (*estimateur sans biais*)

ii)  $Var(\hat{b}) = \sigma^2 ({}^t \mathbf{X} \mathbf{X})^{-1}$  (*matrice de covariance*).

- Hypothèse de normalité (conséquence)

En plus des hypothèses **(H.1)** et **(H.2)** on ajoute une hypothèse **(H.3)** de normalité du vecteur aléatoire  $u$  :

$$u \sim N(0, \sigma^2 \mathbf{I}_T).$$

Posons  $\hat{\sigma}^2 = \|\hat{u}\|^2 / (T - p)$ ,  $\hat{u} = y - \hat{y}$  (vecteur des résidus),  
 $\hat{y} = X\hat{b}$ .

# 1. Modèle de régression multiple

## Theorem

i)  $E(\hat{b}) = b$  (*estimateur sans biais*)

ii)  $Var(\hat{b}) = \sigma^2 ({}^t \mathbf{X} \mathbf{X})^{-1}$  (*matrice de covariance*).

- Hypothèse de normalité (conséquence)

En plus des hypothèses **(H.1)** et **(H.2)** on ajoute une hypothèse **(H.3)** de normalité du vecteur aléatoire  $u$  :

$$u \sim N(0, \sigma^2 \mathbf{I}_T).$$

Posons  $\hat{\sigma}^2 = \|\hat{u}\|^2 / (T - p)$ ,  $\hat{u} = y - \hat{y}$  (vecteur des résidus),  
 $\hat{y} = X\hat{b}$ .

# 1. Modèle de régression multiple

## Theorem

$$i) \hat{b} \sim N \left( b, \sigma^2 ({}^t \mathbf{X} \mathbf{X})^{-1} \right),$$

$$ii) \frac{(T-p)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(T-p),$$

iii)  $\hat{b}$  et  $\hat{\sigma}^2$  sont des statistiques indépendantes.

## Corollary

$$\frac{\hat{b}_j - b_j}{\sqrt{\hat{\sigma}^2 a_{jj}}} \sim t(T-p), \quad 1 \leq j \leq p;$$

$a_{jj}$  est le  $j^{i\text{eme}}$  élément diagonal de  $({}^t \mathbf{X} \mathbf{X})^{-1}$ .

# 1. Modèle de régression multiple

## Theorem

i)  $\hat{b} \sim N \left( b, \sigma^2 ({}^t \mathbf{X} \mathbf{X})^{-1} \right),$

ii)  $\frac{(T-p)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(T-p),$

iii)  $\hat{b}$  et  $\hat{\sigma}^2$  sont des statistiques indépendantes.

## Corollary

$$\frac{\hat{b}_j - b_j}{\sqrt{\hat{\sigma}^2 a_{jj}}} \sim t(T-p), \quad 1 \leq j \leq p;$$

$a_{jj}$  est le  $j^{i\text{eme}}$  élément diagonal de  $({}^t \mathbf{X} \mathbf{X})^{-1}$ .

# 1. Modèle de régression multiple

- Construction des tests d'hypothèses au seuil  $\alpha$ :

$$H_0 : b_j = 0 \quad \xrightarrow{\text{région critique}} \quad \frac{|\hat{b}_j|}{\sqrt{\hat{\sigma}^2 a_{jj}}} \geq t_{1-\alpha/2}(T-p) \quad (4)$$

# 1. Modèle de régression multiple

$$H_0 : b_j \leq 0 \quad \xrightarrow{\text{région critique}} \quad \frac{\hat{b}_j}{\sqrt{\hat{\sigma}^2 a_{jj}}} \geq t_{1-\alpha}(T-p),$$

où  $t_{1-\alpha}(T-p)$  désigne le quantile d'ordre  $1 - \alpha$  de la loi de Student à  $(T-p)$  degrés de liberté.

- Construction de région de confiance (au niveau  $\alpha$ ) pour  $b_j$ , ( $1 \leq j \leq p$ ):

$$\hat{b}_j \pm \sqrt{\hat{\sigma}^2 a_{jj}} t_{1-\alpha/2}(T-p)$$

# 1. Modèle de régression multiple

$$H_0 : b_j \leq 0 \quad \xrightarrow{\text{région critique}} \quad \frac{\hat{b}_j}{\sqrt{\hat{\sigma}^2 a_{jj}}} \geq t_{1-\alpha}(T-p),$$

où  $t_{1-\alpha}(T-p)$  désigne le quantile d'ordre  $1 - \alpha$  de la loi de Student à  $(T-p)$  degrés de liberté.

- **Construction de région de confiance (au niveau  $\alpha$ ) pour  $b_j$ , ( $1 \leq j \leq p$ ):**

$$\hat{b}_j \pm \sqrt{\hat{\sigma}^2 a_{jj}} t_{1-\alpha/2}(T-p)$$

# 1. Modèle de régression multiple

## ● Prévision

Comment, à l'instant  $T$ , prévoir la réalisation de  $y_{T+1}$  disposant de  $x_{T+1} = {}^t(x_{T+1,1}, \dots, x_{T+1,p})$ , c'est à dire d'une réalisation supplémentaire de chacune des  $p$  variables explicatives. Les hypothèses sont **(H.1)**, **(H.2)** et

$$\left\{ \begin{array}{l} y_{T+1} = {}^t x_{T+1} b + u_{T+1}, \\ E(u_{T+1}) = 0, \\ E(u_{T+1} u_t') = 0, \quad 1 \leq t \leq T. \end{array} \right.$$

# 1. Modèle de régression multiple

On a deux incertitudes:

- i) Celle due à  $u_{T+1}$  (perturbation aléatoire),
- ii) Celle due au fait que  $b$  est inconnu.

## Prévision optimale

$$\hat{y}_{T+1} = {}^t x_{T+1} \hat{b}.$$

L'intervalle de confiance au niveau  $(1 - \alpha)$  est donné par

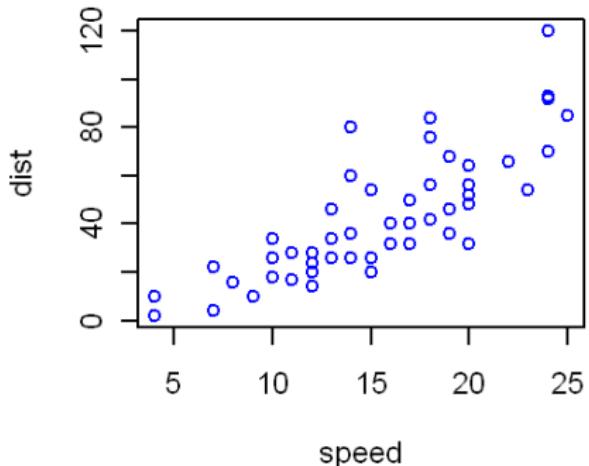
$$\hat{y}_{T+1} \pm \hat{\sigma} t_{1-\alpha/2} (T - p) \sqrt{1 + {}^t x_{T+1} ({}^t \mathbf{X} \mathbf{X})^{-1} x_{T+1}}.$$

# 1. Modèle de régression multiple

- Exemples de modélisation. Exemple 1: Modèle linéaire simple.

Commandes sur R

`data(cars), plot(cars)`

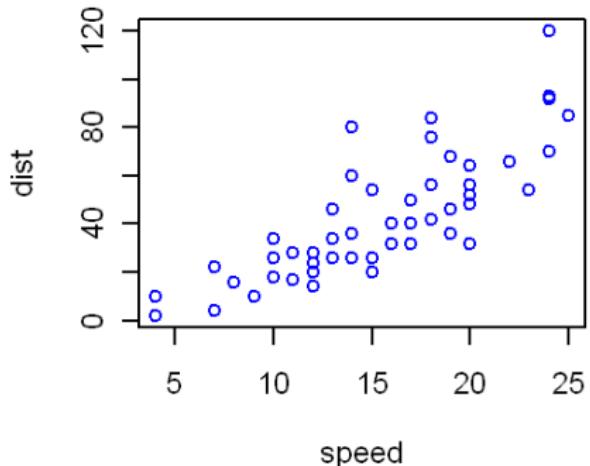


# 1. Modèle de régression multiple

- Exemples de modélisation. Exemple 1: Modèle linéaire simple.

Commandes sur R

`data(cars), plot(cars)`



# 1. Modèle de régression multiple

- On cherche la relation qui lie la variable dist (distance de freinage) et la variable speed (vitesse).
- On propose le modèle  $dist_t = b_1 + b_2 speed_t + u_t$ ,

## Commandes sur R

```
fit <- lm(cars$dist ~ cars$speed), summary(fit)
```

*Call: lm(formula = cars\$dist ~ cars\$speed)*

*Residuals:*

<i>Min</i>	<i>1Q</i>	<i>Median</i>	<i>3Q</i>	<i>Max</i>
-29.069	-9.525	-2.272	9.215	43.201

# 1. Modèle de régression multiple

- On cherche la relation qui lie la variable dist (distance de freinage) et la variable speed (vitesse).
- On propose le modèle  $dist_t = b_1 + b_2 speed_t + u_t$ ,

## Commandes sur R

```
fit <- lm(cars$dist ~ cars$speed), summary(fit)
```

*Call: lm(formula = cars\$dist ~ cars\$speed)*

*Residuals:*

<i>Min</i>	<i>1Q</i>	<i>Median</i>	<i>3Q</i>	<i>Max</i>
-29.069	-9.525	-2.272	9.215	43.201

# 1. Modèle de régression multiple

*Coefficients:*

<i>Estimate</i>	<i>Std. Error</i>	<i>t value</i>	<i>Pr(&gt;  t )</i>
(Intercept) -17.5791	6.7584	-2.601	0.0123 *
cars\$speed 3.9324	0.4155	9.464	1.49e-12 ***

— Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

*Residual standard error: 15.38 on 48 degrees of freedom*

*Multiple R-Squared: 0.6511, Adjusted R-squared: 0.6438*

*F-statistic: 89.57 on 1 and 48 DF, p-value: 1.490e-12*

# 1. Modèle de régression multiple

- Interprétation des résultats:

*Residuals:* En premier lieu R affiche les statistiques descriptives des résidus de la régression.

*Coefficients:* Les estimations  $\hat{b}_1 = -17.5791$ ,  $\hat{b}_2 = 3.9324$ .

*Std. Error* désigne l'écart type estimé.

*t. value* est la statistique de Student.

*Pr(> |t|)* est la p-value associé à la statistique de Student t value, une valeur plus petite que 0.01 nous conduit au rejet de  $H_0$  c'est à dire que le paramètre est significatif. Pour notre modèle le paramètre  $b_1$  n'est pas significatif alors que  $b_2$  l'est.

# 1. Modèle de régression multiple

- Interprétation des résultats:

*Residuals:* En premier lieu R affiche les statistiques descriptives des résidus de la régression.

*Coefficients:* Les estimations  $\hat{b}_1 = -17.5791$ ,  $\hat{b}_2 = 3.9324$ .

*Std. Error* désigne l'écart type estimé.

*t. value* est la statistique de Student.

*Pr(> |t|)* est la p-value associé à la statistique de Student t value, une valeur plus petite que 0.01 nous conduit au rejet de  $H_0$  c'est à dire que le paramètre est significatif. Pour notre modèle le paramètre  $b_1$  n'est pas significatif alors que  $b_2$  l'est.

# 1. Modèle de régression multiple

- Interprétation des résultats:

*Residuals:* En premier lieu R affiche les statistiques descriptives des résidus de la régression.

*Coefficients:* Les estimations  $\hat{b}_1 = -17.5791$ ,  $\hat{b}_2 = 3.9324$ .

*Std. Error* désigne l'écart type estimé.

*t. value* est la statistique de Student.

*Pr(> |t|)* est la p-value associé à la statistique de Student t value, une valeur plus petite que 0.01 nous conduit au rejet de  $H_0$  c'est à dire que le paramètre est significatif. Pour notre modèle le paramètre  $b_1$  n'est pas significatif alors que  $b_2$  l'est.

# 1. Modèle de régression multiple

- Interprétation des résultats:

*Residuals:* En premier lieu R affiche les statistiques descriptives des résidus de la régression.

*Coefficients:* Les estimations  $\hat{b}_1 = -17.5791$ ,  $\hat{b}_2 = 3.9324$ .

*Std. Error* désigne l'écart type estimé.

*t. value* est la statistique de Student.

*Pr(> |t|)* est la p-value associé à la statistique de Student t value, une valeur plus petite que 0.01 nous conduit au rejet de  $H_0$  c'est à dire que le paramètre est significatif. Pour notre modèle le paramètre  $b_1$  n'est pas significatif alors que  $b_2$  l'est.

# 1. Modèle de régression multiple

- Interprétation des résultats:

*Residuals:* En premier lieu R affiche les statistiques descriptives des résidus de la régression.

*Coefficients:* Les estimations  $\hat{b}_1 = -17.5791$ ,  $\hat{b}_2 = 3.9324$ .

*Std. Error* désigne l'écart type estimé.

*t. value* est la statistique de Student.

*Pr(> |t|)* est la p-value associé à la statistique de Student t value, une valeur plus petite que 0.01 nous conduit au rejet de  $H_0$  c'est à dire que le paramètre est significatif. Pour notre modèle le paramètre  $b_1$  n'est pas significatif alors que  $b_2$  l'est.

# 1. Modèle de régression multiple

$$\text{Residual standard error} = \sqrt{\frac{||\hat{u}||^2}{T-p}}.$$

*Multiple R-Squared:*

## Definition

Lorsqu'il y'a une constante dans le modèle de régression multiple, on appelle coefficient de détermination le scalaire

$$R^2 = \frac{||\hat{y} - \bar{y}\delta_T||^2}{||y - \bar{y}\delta_T||^2}$$

où  $\bar{y} = \frac{1}{T} \sum_{n=1}^T y_n$  et  $\delta_T = {}^t(1, \dots, 1)$ , vecteur  $(T, 1)$ .

Plus  $R^2$  est proche de 1, plus l'ajustement est meilleur.

# 1. Modèle de régression multiple

*Adjusted R-squared:*  $R_a^2 = 1 - (1 - R^2) \frac{T-1}{T-p-1}$ , où  $p$  est le nombre de paramètres sans compter la constante.  $R_a^2$  ne croit que si la nouvelle variable explicative ajoutée améliore l'ajustement, elle peut être négative, et  $R_a^2 \leq R^2$ .

*F-statistic* désigne la statistique de Fisher qui correspond à l'hypothèse nulle  $H_0 : b_1 = \dots = b_p = 0$ ; elle est donnée par  $F = \frac{\|\tilde{u}\|^2 - \|\hat{u}\|^2}{\|\hat{u}\|^2}$  avec  $\tilde{u}$ : résidus sous  $H_0$  et  $\hat{u}$ : résidus sous  $H_1$ . On a sous  $H_0$  la statistique  $F$  suit une loi de Fisher  $F(p, T - p)$ .

# 1. Modèle de régression multiple

*Adjusted R-squared:*  $R_a^2 = 1 - (1 - R^2) \frac{T-1}{T-p-1}$ , où  $p$  est le nombre de paramètres sans compter la constante.  $R_a^2$  ne croit que si la nouvelle variable explicative ajoutée améliore l'ajustement, elle peut être négative, et  $R_a^2 \leq R^2$ .

*F-statistic* désigne la statistique de Fisher qui correspond à l'hypothèse nulle  $H_0 : b_1 = \dots = b_p = 0$ ; elle est donnée par  $F = \frac{\|\tilde{u}\|^2 - \|\hat{u}\|^2}{\|\hat{u}\|^2}$  avec  $\tilde{u}$  : résidus sous  $H_0$  et  $\hat{u}$  : résidus sous  $H_1$ . On a sous  $H_0$  la statistique  $F$  suit une loi de Fisher  $F(p, T - p)$ .

# 1. Modèle de régression multiple

## Commandes sur R

```
fit1 <- lm(cars$dist ~ cars$speed-1), summary(fit1)
```

Call:  $lm(formula = cars\$dist \sim cars\$speed - 1)$

Residuals:

Min	1Q	Median	3Q	Max
-26.183	-12.637	-5.455	4.590	50.181

Coefficients:

Estimate Std. Error t value Pr(> |t|)

$\text{cars\$speed}$  2.9091 0.1414 20.58 < 2e-16 \*\*\*

# 1. Modèle de régression multiple

## Commandes sur R

```
fit1 <- lm(cars$dist ~ cars$speed-1), summary(fit1)
```

Call:  $lm(formula = cars\$dist \sim cars\$speed - 1)$

Residuals:

<i>Min</i>	<i>1Q</i>	<i>Median</i>	<i>3Q</i>	<i>Max</i>
-26.183	-12.637	-5.455	4.590	50.181

Coefficients:

*Estimate Std. Error t value Pr(> |t|)*

*cars\$speed* 2.9091 0.1414 20.58 < 2e-16 \*\*\*

# 1. Modèle de régression multiple

*Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1*

*Residual standard error: 16.26 on 49 degrees of freedom*

*Multiple R-Squared: 0.8963, Adjusted R-squared: 0.8942*

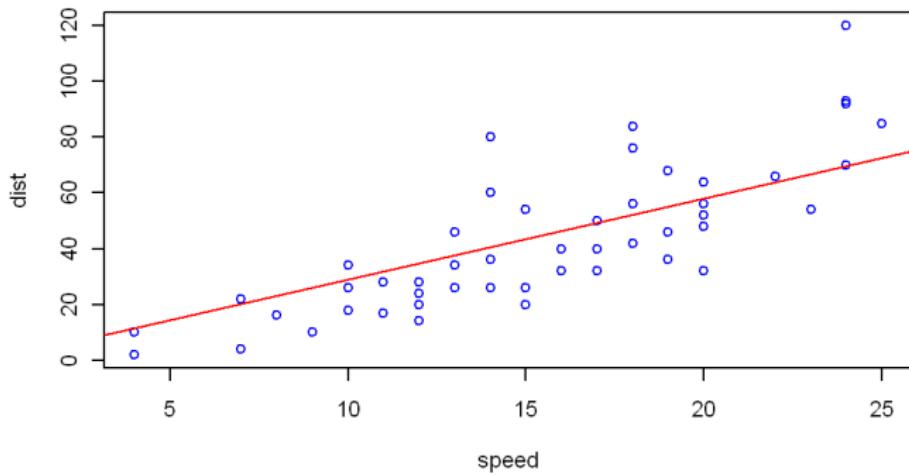
*F-statistic: 423.5 on 1 and 49 DF, p-value: < 2.2e - 16.*

Le  $R^2$  s'approche de 1, donc l'ajustement est meilleur.

# 1. Modèle de régression multiple

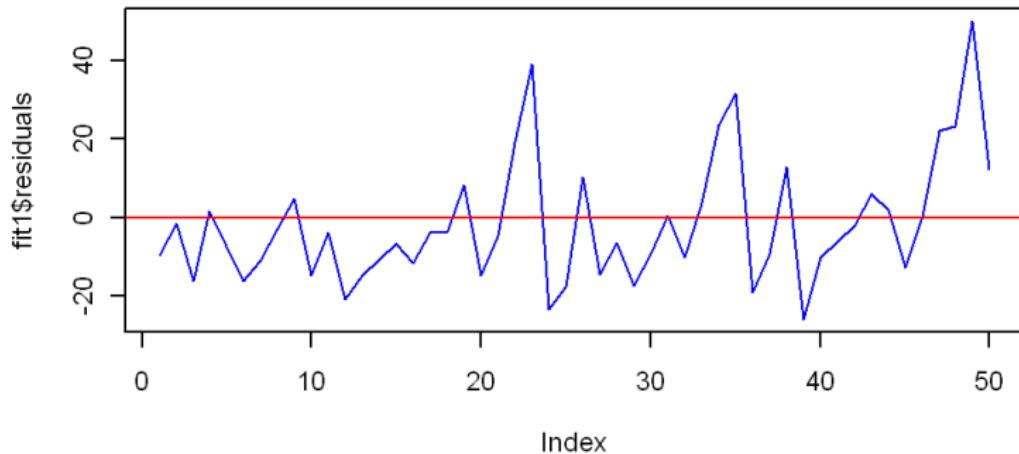
## Commandes sur R

```
plot(cars,col="blue"), z <- lm(dist ~ speed, data = cars),  
abline(z,col="red")
```



# 1. Modèle de régression multiple

## Analyse des résidus de la régression



# 1. Modèle de régression multiple

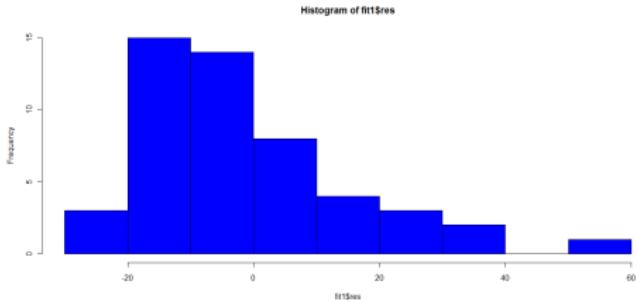
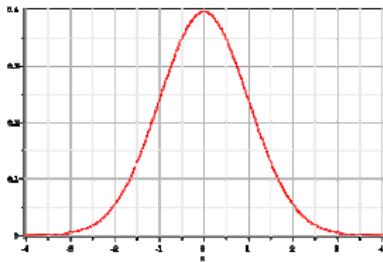
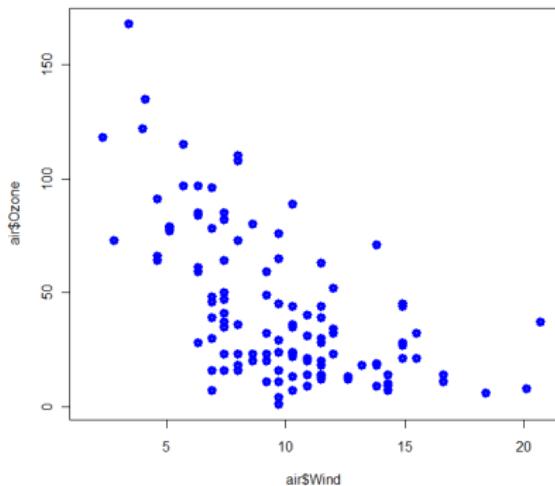
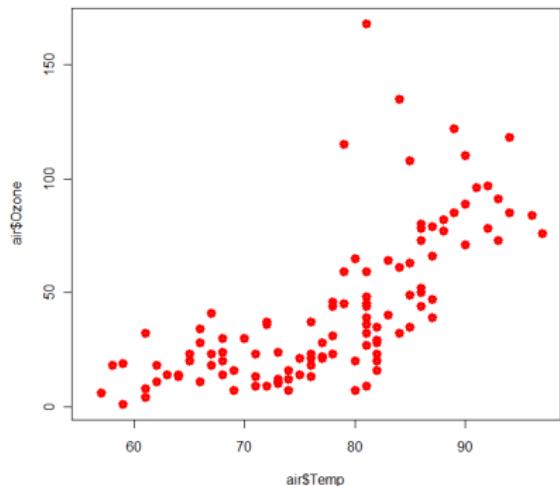


Figure:

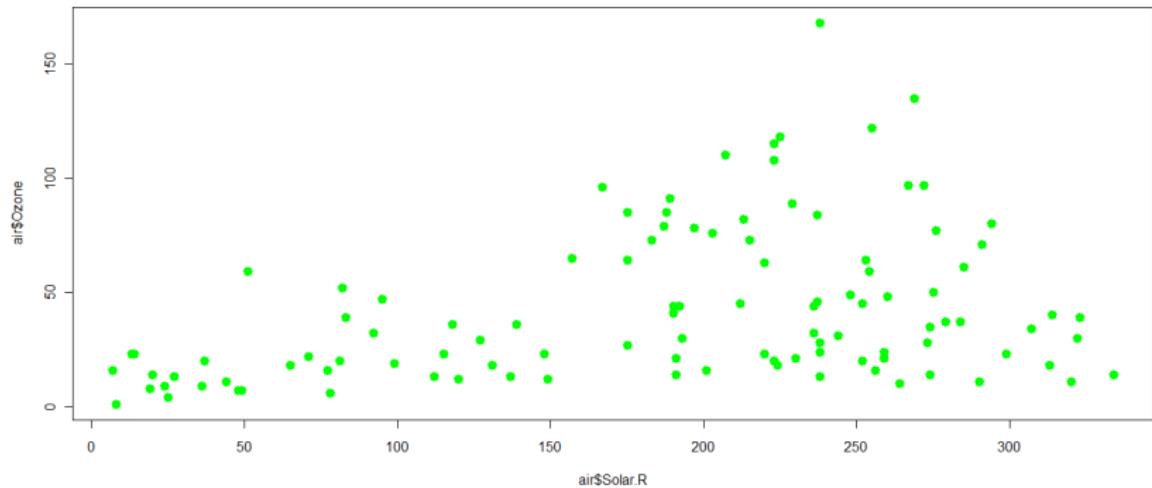


# 1. Modèle de régression multiple

**Exemple 2.** Régression multiple: On considère l'évolution de l'ozone en fonction de la température, du vent et de la radiation solaire.



# 1. Modèle de régression multiple



## Commandes sur R

```
fit3 <- lm(air$Ozone ~ air$Temp + air$wind + air$Solar.R),  
summary(fit3)
```



# 1. Modèle de régression multiple

Call: lm(formula = air\$Ozone ~ air\$Temp + air\$Wind +  
air\$Solar.R)

Residuals:

Min	1Q	Median	3Q	Max
-40.485	-14.219	-3.551	10.097	95.619

Coefficients:

	Estimate	Std. Error	t value	Pr(>  t  )
(Intercept)	-64.34208	23.05472	-2.791	0.00623 **
air\$Temp	1.65209	0.25353	6.516	2.42e-09 ***
air\$Wind	-3.33359	0.65441	-5.094	1.52e-06 ***
air\$Solar.R	0.05982	0.02319	2.580	0.01124 *

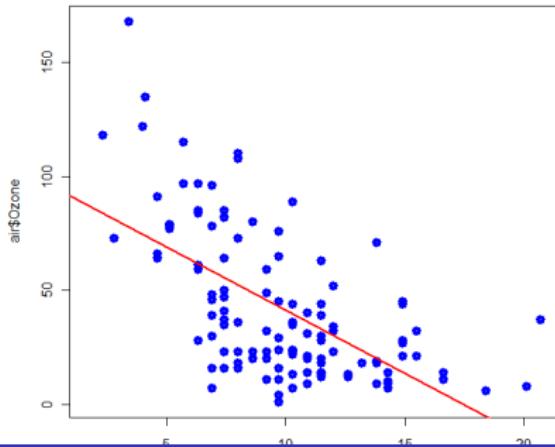
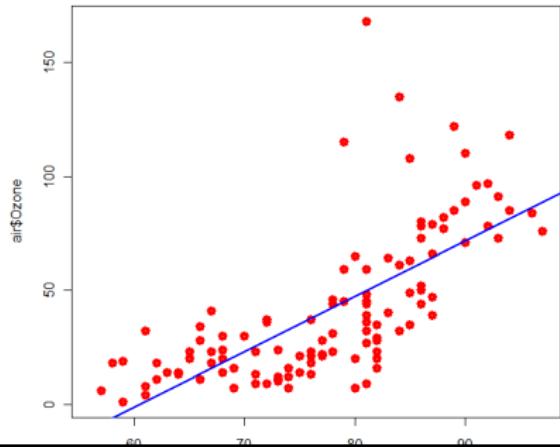
— Signif. codes: 0 \*\*\* 0.001 \*\* 0.01 \* 0.05 . 0.1 1

# 1. Modèle de régression multiple

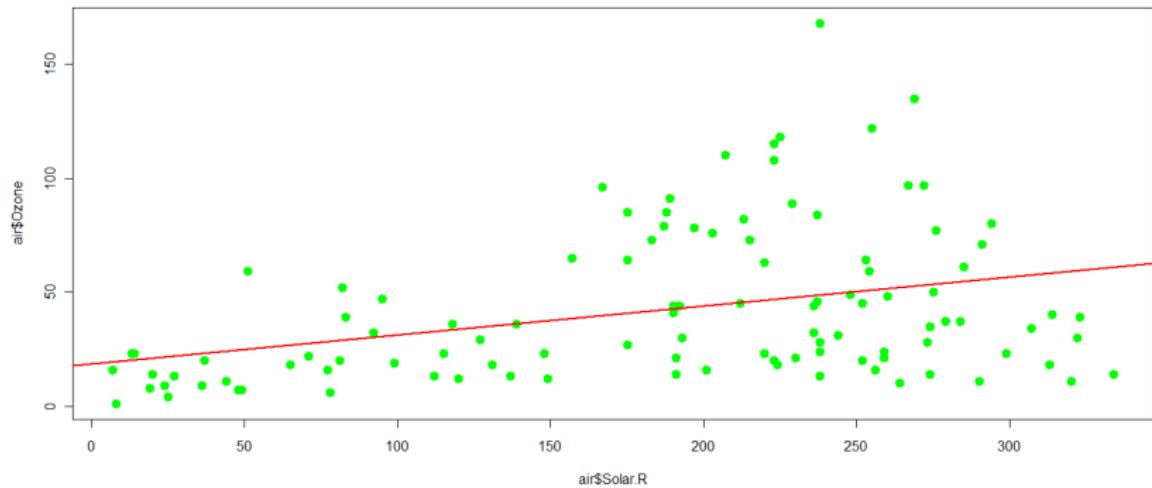
Residual standard error: 21.18 on 107 degrees of freedom (42 observations deleted due to missingness)

Multiple R-squared: 0.6059, Adjusted R-squared: 0.5948

F-statistic: 54.83 on 3 and 107 DF, p-value: < 2.2e - 16

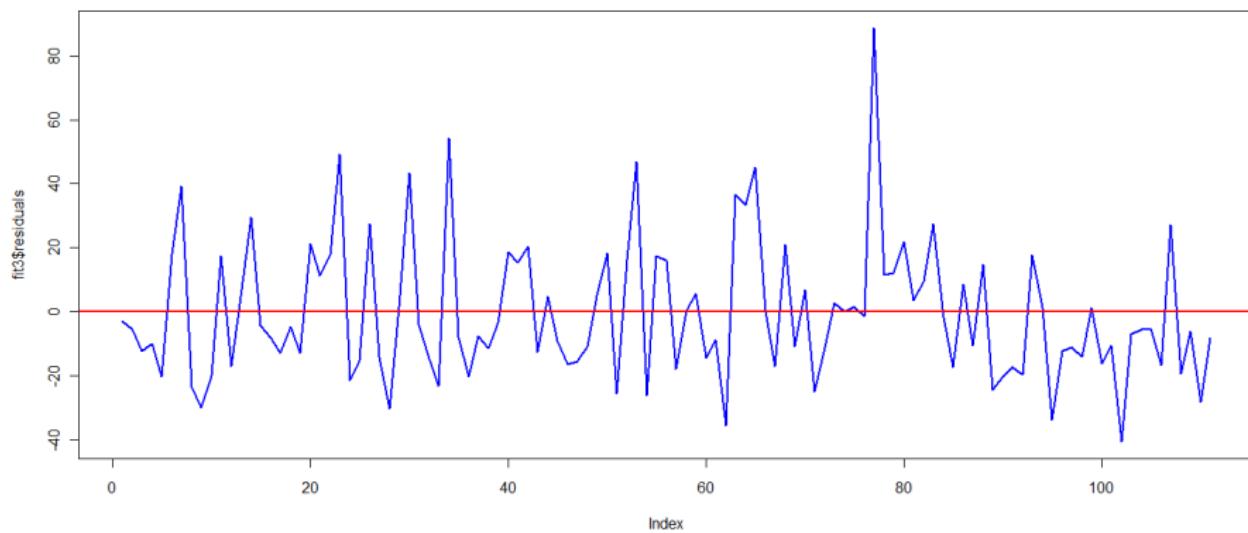


# 1. Modèle de régression multiple

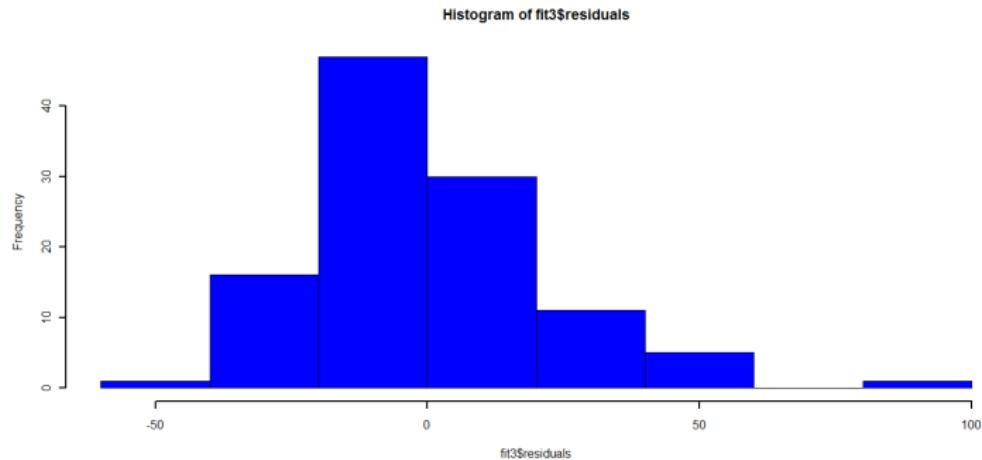


# 1. Modèle de régression multiple

Les résidus de la régression sont donnés par



# 1. Modèle de régression multiple



**Conclusion:** L'ajustement est assez bon, l'hypothèse de normalité des données est presque satisfaite.

## 2. Processus autorégressif-moyenne mobile (ARMA)

- Processus du second ordre

### Definition

On appelle **processus aléatoire du second ordre** à temps discret, toute suite réelle  $X_t : \Omega \longrightarrow \mathbb{R}$  définie sur un espace de probabilité  $(\Omega, \mathcal{F}, P)$  à valeurs dans  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , indexée par  $t \in \mathbb{N}$  ou  $\mathbb{Z}$  telle que:

$$E(|X_t|^2) < \infty \text{ pour tout } t.$$

## 2. Processus autorégressif-moyenne mobile (ARMA)

i) Moyenne:  $\mu : \mathbb{Z} \longrightarrow \mathbb{R}$ ,  $\mu(t) = E(X_t)$ ,  $\forall t \in \mathbb{Z}$ .

ii) Auto-covariance:  $K : \mathbb{Z} \times \mathbb{Z} \longrightarrow \mathbb{R}$ ,

$$K(t, s) = E((X_t - \mu(t))(X_s - \mu(s))), \forall t, s \in \mathbb{Z}.$$

Ces fonctions s'expriment encore à l'aide du produit scalaire:

$$\langle X, Y \rangle = E(XY)$$

défini sur l'espace de probabilité  $L^2(\Omega, \mathcal{F}, P)$

$$\mu(t) = \langle X_t, 1 \rangle, \quad K(t, s) = \langle X_t, X_s \rangle - \mu(t)\mu(s).$$

Le processus  $X_t$  est dit centré si  $\mu = 0$ .

## 2.Processus autorégressif-moyenne mobile (ARMA)

### Theorem

*La fonction d'auto-covariance est symétrique de type positif, i.e.*

$$K(t, s) = K(s, t); \quad \sum_t \sum_s a(t) K(t, s) a(s) \geq 0$$

*pour toute fonction  $a : \mathbb{Z} \rightarrow \mathbb{R}$  à support fini (quels que soient les entiers  $(t_j)_{1 \leq j \leq k}$ , la matrice  $[K(t_i, t_j)]_{1 \leq i, j \leq k}$  est symétrique positive).*

## 2.Processus autorégressif-moyenne mobile (ARMA)

- Processus stationnaire au second ordre

### Definition

On dit que le processus  $X$  est **strictement (ou fortement) stationnaire** si pour toute suite finie d'instants  $t_1, \dots, t_k$  éléments de  $\mathbb{Z}$  et tout entier  $r \in \mathbb{Z}$ , les lois jointes de  $(X_{t_1}, \dots, X_{t_k})$  et de  $(X_{t_1+r}, \dots, X_{t_k+r})$  sont les mêmes (lois jointes invariantes par translation dans le temps).

## 2.Processus autorégressif-moyenne mobile (ARMA)

### Definition

Un processus est **stationnaire au second ordre** (on dit aussi **faiblement stationnaire**) si moyenne et auto-covariance sont invariantes par translation dans le temps, i.e.:

$$\left\{ \begin{array}{l} E(X_t) = \mu, \forall t \in \mathbb{Z}, \\ \text{et} \\ E(X_t - \mu)(X_s - \mu) = E(X_{t+r} - \mu)(X_{s+r} - \mu), \quad \forall (t, s, r) \in \mathbb{Z}^3. \end{array} \right.$$

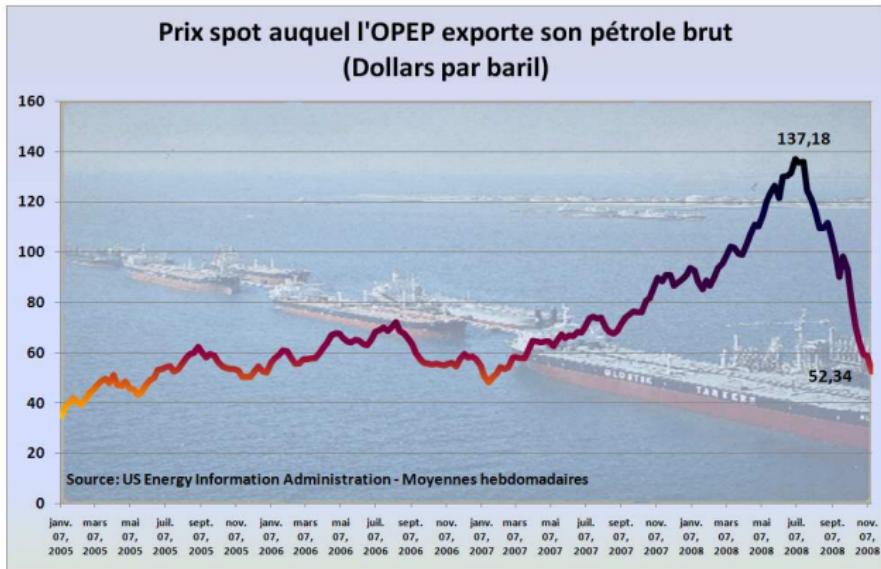
où  $\mu$  est une constante et  $K(t, s)$  ne dépend plus que de  $t - s$ , donc il existe une fonction  $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$ , telle que:  
 $K(t, s) = \gamma(t - s)$ .

## 2. Processus autorégressif-moyenne mobile (ARMA)

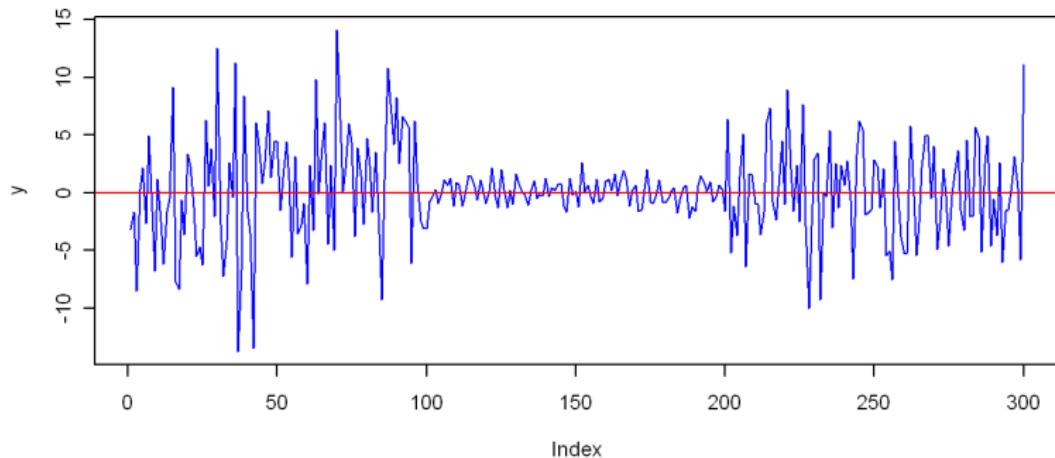
Pour un processus faiblement stationnaire, on a

$$\left\{ \begin{array}{l} \text{var}(X_t) = \gamma(0), \forall t \in \mathbb{Z}, \\ |\gamma(k)| \leq \gamma(0), \forall k \in \mathbb{Z}, \text{ ( par l'inégalité de Shwartz)} \\ \gamma(k) = \gamma(-k), \forall k \in \mathbb{Z}, \\ \rho(j) = \frac{\gamma(j)}{\gamma(0)}, \text{ **fonction d'auto-corrélation ou mémoire de } X_t \\ r(j) = \frac{\text{cov}\left(X_1 - \mathbf{P}_{[x_2, \dots, x_j]}(X_1), X_{j+1} - \mathbf{P}_{[x_2, \dots, x_j]}(X_{j+1})\right)}{\text{var}\left(X_1 - \mathbf{P}_{[x_2, \dots, x_j]}(X_1)\right)} \\ \text{fonction d'auto-corrélation partielle .} \end{array} \right.**$$

## 2. Processus autorégressif-moyenne mobile (ARMA)



## 2. Processus autorégressif-moyenne mobile (ARMA)



## Analyse spectrale d'un processus stationnaire au second ordre

### Theorem

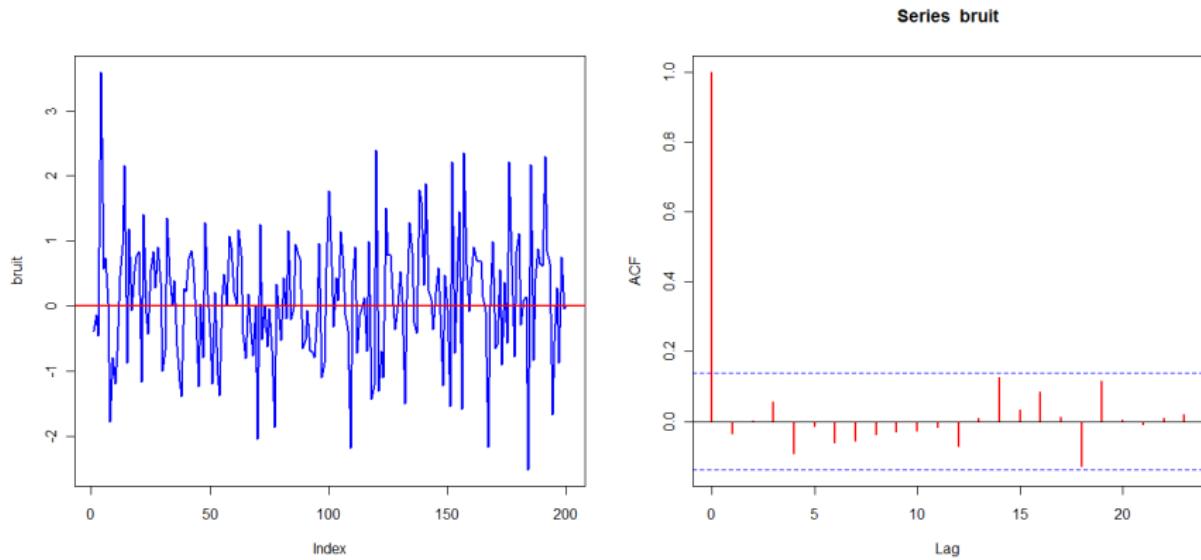
(Herglotz-Fourier) Soit  $X = (X_t)_{t \in \mathbb{Z}}$  un processus stationnaire au second ordre, de fonction d'auto-covariance  $\gamma(k)$  et de densité spectrale  $f(\lambda)$ . Alors

$$\gamma(k) = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda \quad \text{pour tout } k \in \mathbb{Z}$$

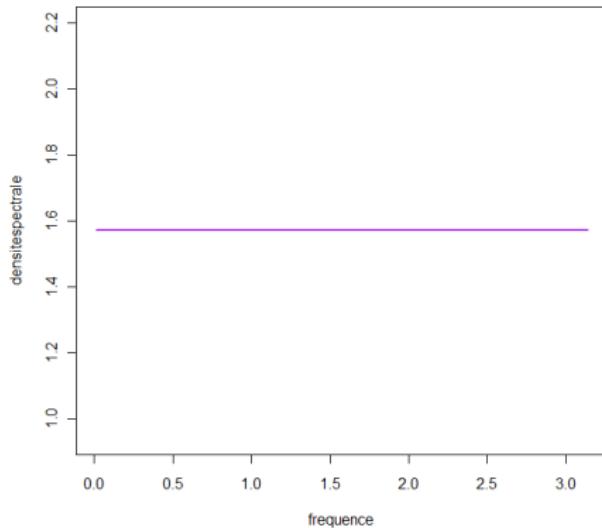
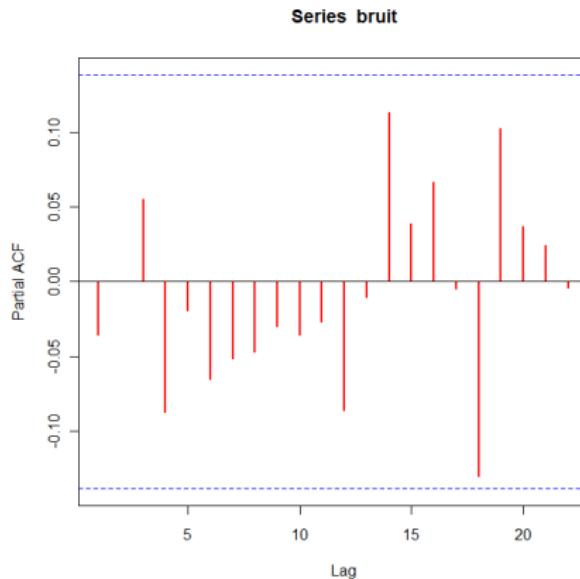
$$f(\lambda) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} e^{-ik\lambda} \gamma(k)$$

## 2. Processus autorégressif-moyenne mobile (ARMA)

Exemples. 1) Bruit blanc :  $\varepsilon = (\varepsilon_t)_{t \in \mathbb{Z}}$  un processus centré tel que  $E(\varepsilon_t \varepsilon_s) = \sigma^2 \delta_t^s$ . Alors  $f(\lambda) = \sigma^2 / 2\pi$ .



## 2. Processus autorégressif-moyenne mobile (ARMA)



## 2.Processus autorégressif-moyenne mobile (ARMA)

ii) Moyenne mobile: Soit  $(\varepsilon_t)_{t \in \mathbb{Z}}$  un bruit blanc, un processus  $MA(q)$  est défini par

$$X_t = \sum_{k=0}^q c_k \varepsilon_{t-k},, c_0 = 1.$$

$$\begin{aligned}\gamma(k) &= E(X_{t+k} X_t) \\ &= \sum_{l,m} c_l c_m E(\varepsilon_{t+k-l} \varepsilon_{t-m}), \quad |k| \leq q, \\ &= \sigma^2 \sum_{l=0}^{q-k} c_{k+l} c_l, \quad |k| \leq q,\end{aligned}$$

## 2. Processus autorégressif-moyenne mobile (ARMA)

$$\gamma(k) = 0, \quad \text{si } |k| > q;$$

donc un tel processus admet une densité spectrale (continue)

$$\begin{aligned} f(\lambda) &= \frac{1}{2\pi} \sum_k \gamma(k) e^{-ik\lambda}, \\ &= \frac{\sigma^2}{2\pi} \sum_{|k| \leq q} \sum_{l=0}^{q-k} c_{k+l} c_l e^{-ik\lambda} \end{aligned}$$

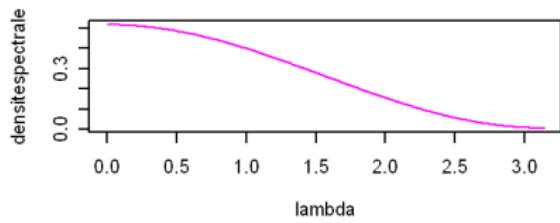
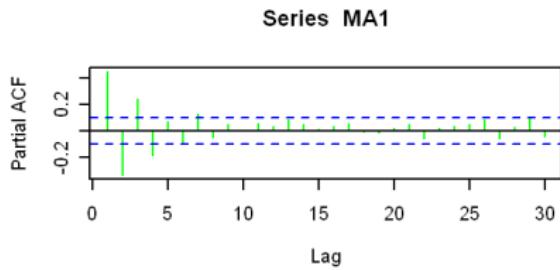
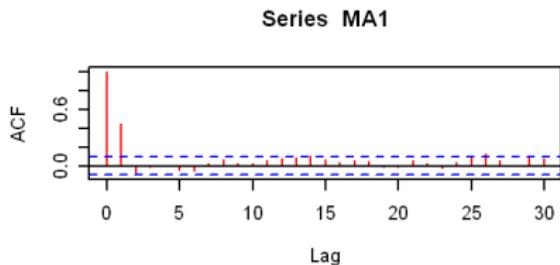
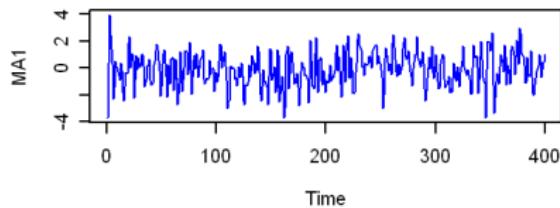
posons  $k + l = m$ ,

$$\begin{aligned} f(\lambda) &= \frac{\sigma^2}{2\pi} \sum_{m=0}^q \sum_{l=0}^q c_m c_l e^{-i(m-l)\lambda} \\ &= \frac{\sigma^2}{2\pi} \left| \sum_{l=0}^q c_l e^{-il\lambda} \right|^2 \end{aligned}$$

## 2. Processus autorégressif-moyenne mobile (ARMA)

### Commandes sur R

```
MA1 <- arima.sim(list(order= c(0,0,1), ma = 0.7),200).
```



## 2. Processus autorégressif-moyenne mobile (ARMA)

### • Processus ARMA

#### Definition

On dit que  $X$  est un processus ARMA( $p, q$ ), s'il existe un bruit blanc  $\varepsilon$  défini sur le même  $(\Omega, \mathcal{F}, P)$  et des nombres réels  $a_0 (= 1), a_1, \dots, a_p, c_0 (= 1), c_1, \dots, c_q$  tels que

$$\sum_{k=0}^p a_k X_{t-k} = \sum_{l=0}^q c_l \varepsilon_{t-l}. \quad (5)$$

En particulier:  $X$  est dit autorégressif d'ordre  $p$ , AR( $p$ ), si

$$\sum_{k=0}^p a_k X_{t-k} = \varepsilon_t \quad (6)$$

## 2. Processus autorégressif-moyenne mobile (ARMA)

### • Processus AR( $p$ )

- Ces processus jouent un rôle important dans les applications.
- $X_n$  est une extrapolation linéaire des  $p$  valeurs précédentes  $X_{n-1}, \dots, X_{n-p}$  à un bruit blanc près.

Si  $A$  désigne le polynôme unitaire ( $A(0) = 1$ ) de degré  $p$

$$A(x) = 1 + \sum_{k=1}^p a_k x^k, \quad (7)$$

l'équation (6) peut alors s'écrire, à l'aide de l'opérateur retard  $B$ , ( $BX_t = X_{t-1}$ ):

$$A(B)X_t = \varepsilon_t, \quad \forall t \in \mathbb{Z}. \quad (8)$$



## 2.Processus autorégressif-moyenne mobile (ARMA)

### Theorem

*Un processus autorégressif d'ordre  $p$ , possède une densité spectrale définie par*

$$f(\lambda) = \frac{\sigma^2}{2\pi |A(e^{-i\lambda})|^2} \quad (9)$$

*( $\sigma^2$  est la variance de  $\varepsilon$ ), et le polynôme  $A$  ne peut donc posséder de zéros de module 1.*

*Réciproquement si un processus centré stationnaire au second ordre  $X$  possède une densité spectrale de la forme (9) pour un nombre réel  $\sigma > 0$  et un polynôme unitaire  $A$  de degré  $p$  (nécessairement sans zéro de module 1) ce processus est autorégressif d'ordre  $p$  les  $(a_j)_{1 \leq j \leq p}$  étant donnés par (7).*

## 2.Processus autorégressif-moyenne mobile (ARMA)

### Theorem

*La fonction d'auto-corrélation partielle d'un processus AR( $p$ ) vérifie  $r(j) = 0 \forall j > p$ .*

### Theorem

*La densité spectrale d'un processus AR( $p$ ) s'écrit d'une seule manière sous la forme*

$$f(\lambda) = \frac{\sigma^2}{2\pi |A(e^{-i\lambda})|^2},$$

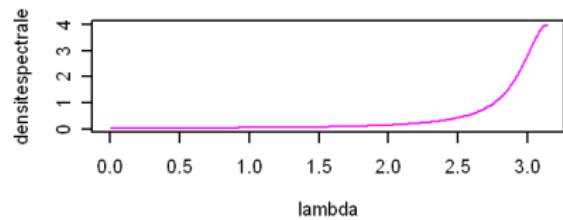
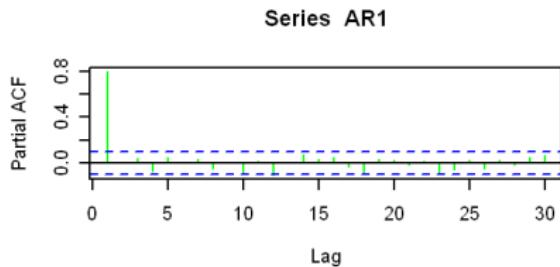
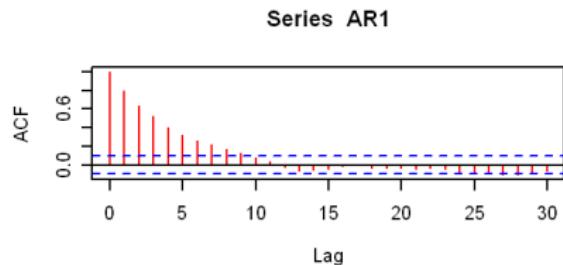
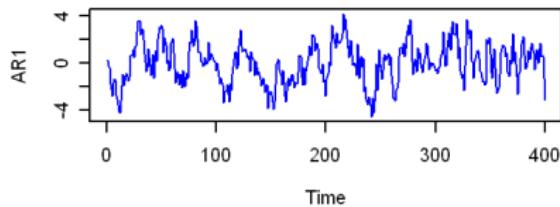
*pour un réel  $\sigma > 0$ , et un polynôme unitaire  $A$  de degré  $p$  dont toutes les racines sont de module  $> 1$ .*

*La représentation (6) du processus  $X$  est alors appelée représentation canonique de  $X$ .*

## 2. Processus autorégressif-moyenne mobile (ARMA)

### Commandes sur R

```
AR1 <- arima.sim(list(order = c(1,0,0), ar = 0.7),400).
```



## 2.Processus autorégressif-moyenne mobile (ARMA)

- Processus ARMA

### Definition

Soit  $(X_t)$  le processus ARMA défini par l'équation (5). On dit que  $(X_t)$  est **causal** si les racines de  $A(z)$  sont toutes strictement à l'extérieur du disque unité.

### Definition

On dit que  $(X_t)$  est **inversible** si les racines de  $C(z)$  sont toutes strictement à l'extérieur du disque unité.

## 2.Processus autorégressif-moyenne mobile (ARMA)

- Processus ARMA

### Definition

Soit  $(X_t)$  le processus ARMA défini par l'équation (5). On dit que  $(X_t)$  est **causal** si les racines de  $A(z)$  sont toutes strictement à l'extérieur du disque unité.

### Definition

On dit que  $(X_t)$  est **inversible** si les racines de  $C(z)$  sont toutes strictement à l'extérieur du disque unité.

## 2.Processus autorégressif-moyenne mobile (ARMA)

### Theorem

(représentation canonique) Soit  $(X_t)$  un processus ARMA stationnaire au second ordre.

i) Sa densité spectrale est donnée par:

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{|C(e^{-i\lambda})|^2}{|A(e^{-i\lambda})|^2} \quad (10)$$

ii) Il existe deux polynômes  $\tilde{A}(z)$  et  $\tilde{C}(z)$  n'ayant que des racines strictement à l'extérieur du disque unité et un bruit blanc  $(\eta_t)$  tel que:

$$\tilde{A}(B)X_t = \tilde{C}(B)\eta_t,$$

cette représentation est unique et dite canonique.

## 2.Processus autorégressif-moyenne mobile (ARMA)

$$A(z) = \prod_{j=1}^p (1 - \alpha_j z); \quad C(z) = \prod_{j=1}^q (1 - \beta_j z);$$

$|\alpha_j| < 1$ , si  $1 \leq j \leq r$ ;  $|\alpha_j| > 1$ , si  $r + 1 \leq j \leq p$ ;

$|\beta_j| < 1$ , si  $1 \leq j \leq s$ ;  $|\beta_j| > 1$ , si  $s + 1 \leq j \leq q$ ;

$$\tilde{A}(z) = \prod_{j=1}^r (1 - \alpha_j z) \prod_{j=r+1}^p \left(1 - \frac{1}{\bar{\alpha}_j} z\right);$$

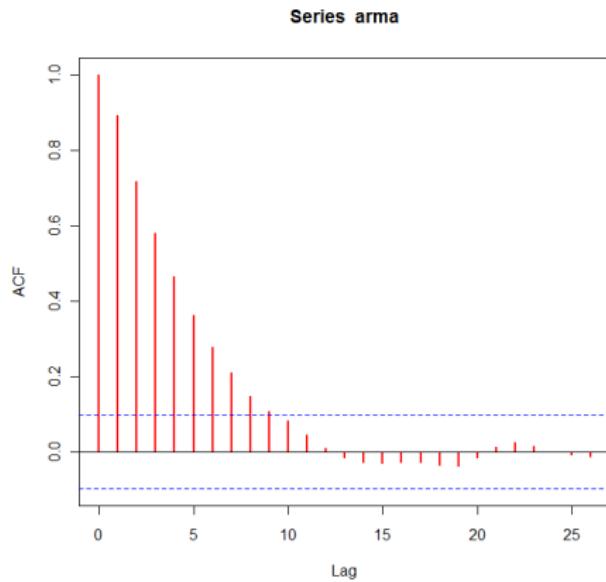
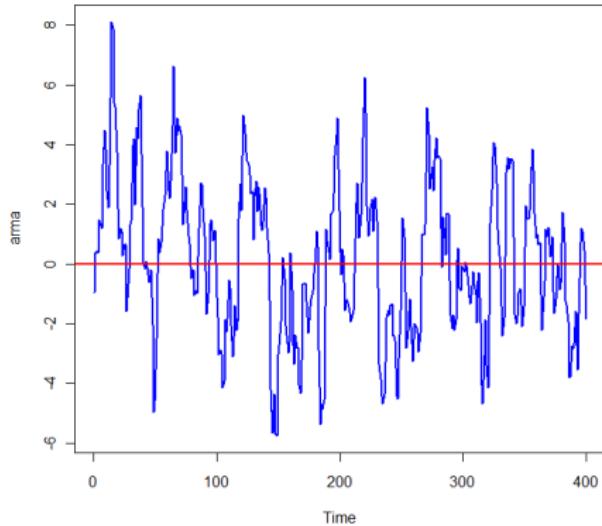
$$\tilde{C}(z) = \prod_{j=1}^s (1 - \beta_j z) \prod_{j=s+1}^q \left(1 - \frac{1}{\bar{\beta}_j} z\right);$$

$$\sigma_\eta^2 = \sigma_\varepsilon^2 \frac{\prod_{j=s+1}^q |\beta_j|^2}{\prod_{j=r+1}^p |\alpha_j|^2}.$$

## 2. Processus autorégressif-moyenne mobile (ARMA)

### Commandes sur R

```
arma <- arima.sim(list(order = c(1,0,1), ar = 0.8,ma=0.6), 400).
```



## 2.Processus autorégressif-moyenne mobile (ARMA)

Modélisation: Trois étapes principales:

- 1. Identification des degrés  $p$  et  $q$  en utilisant la fonction d'auto-corrélation, la fonction d'auto-corrélation partielle, ou encore les critères d'information:

$$\textbf{Akaike} : AIC(p, q) = \ln(\hat{\sigma}^2) + 2(p + q)/T,$$

$$\textbf{Schwarz} : BIC(p, q) = \ln(\hat{\sigma}^2) + 2(p + q) \ln(T)/T;$$

avec  $\hat{\sigma}^2$  un estimateur de  $\sigma^2 = \text{var}(\varepsilon_t)$ .

## 2. Processus autorégressif-moyenne mobile (ARMA)

- 2. Estimation des paramètres  $a_j, 1 \leq j \leq p$  et  $c_k, 1 \leq k \leq q$  en utilisant les méthodes d'estimation (moindres carrés, maximum de vraisemblance)
- 3. Tester la validité du modèle proposé.

## 2.Processus autorégressif-moyenne mobile (ARMA)

- **Estimation d'un AR( $p$ ):** Sur R il existe quatre méthodes d'estimation utilisant le critère d'information AIC pour déterminer un choix optimal du degré  $p$ .

### Commandes sur R

```
ar(x, aic= TRUE, order.max = NULL, method=c("yule-walker",  
"burg", "ols", "mle", "yw"))
```

### Commandes sur R

```
fit4 <- ar(AR1,aic=TRUE,"burg")
```

Call:`ar(x = AR1, aic = TRUE, method = "burg")`

Coefficients:

1

0.6872

## 2. Processus autorégressif-moyenne mobile (ARMA)

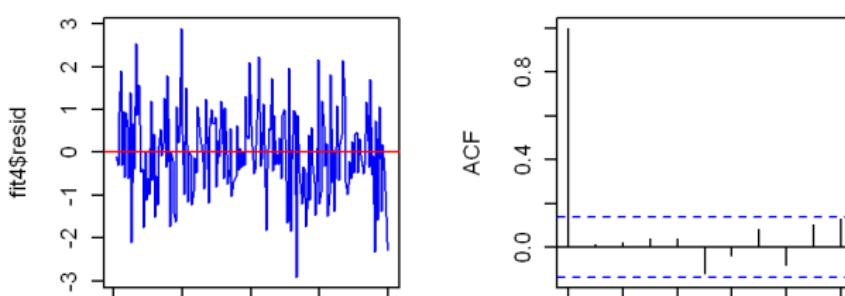
Order selected 1 sigma<sup>2</sup> estimated as 1.030

Interprétation: Le modèle optimal est un AR(1),  $\hat{a}_1 = 0.6872$ .

- Analyse des résidus

Commandes sur R

```
ts.plot(fit4$resid,type="l",col="blue"), abline(0,0,col="red"),
acf(fit4$resid[2:200],10,"correlation")
```



## 2.Processus autorégressif-moyenne mobile (ARMA)

### Estimation d'un ARMA( $p, q$ )

- Plusieurs librairies dans R: stats, FitARMA

Commandes sur R, stats

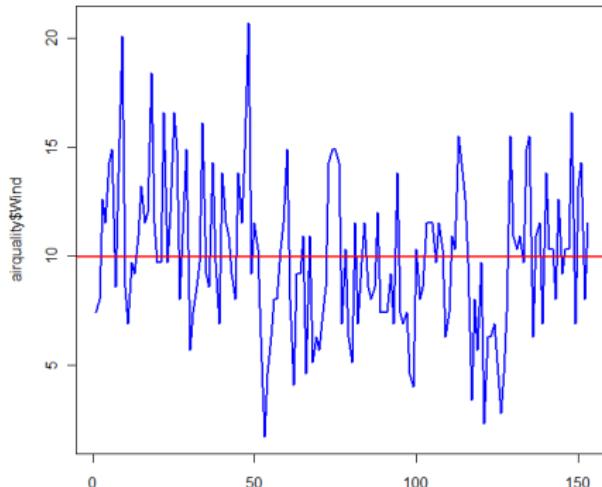
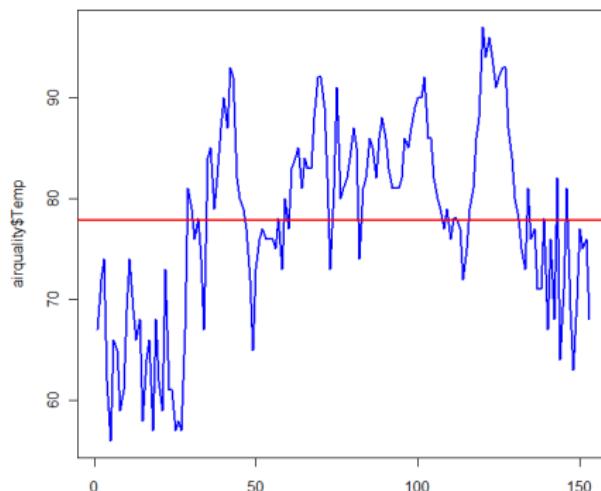
```
library(stats), fit= arima0(data, list(order=c(p,1,q)))
```

## 2. Processus autorégressif-moyenne mobile (ARMA)

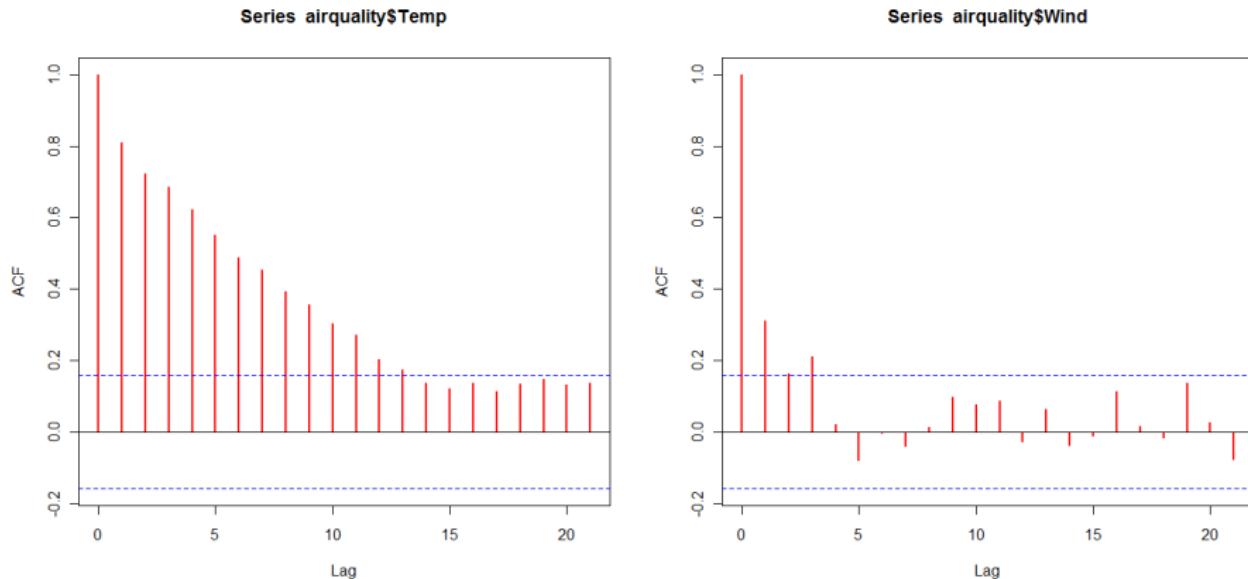
Exemple. Modélisation des séries temporelles airquality

(température et vent, Mai-Septembre). Echantillon d'apprentissage:

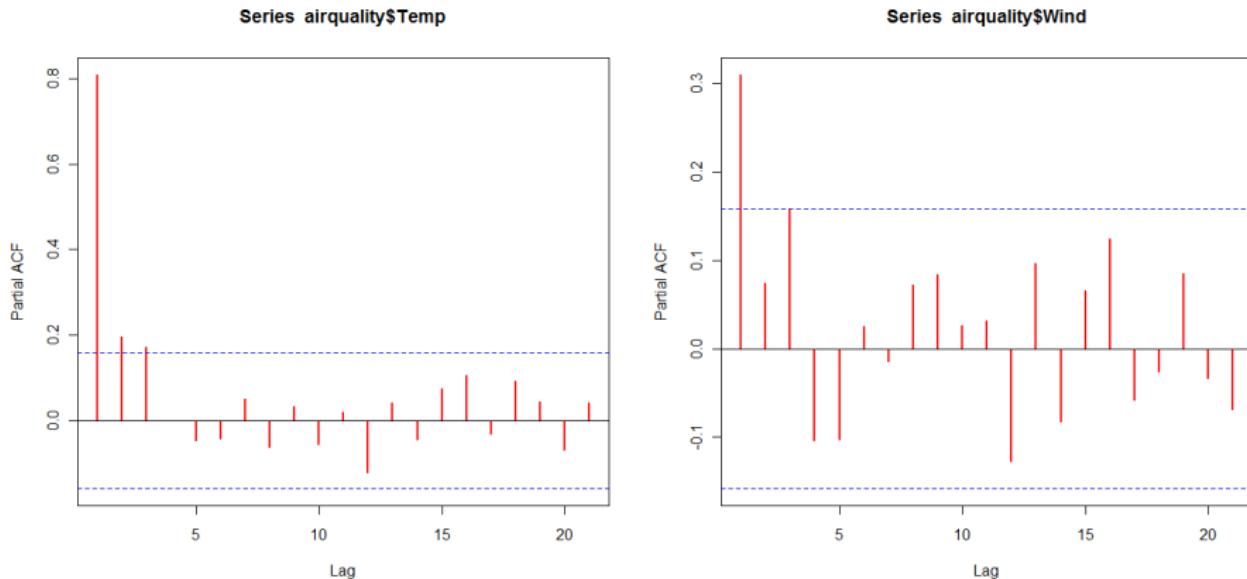
146 jours. Echantillon témoin : 7 jours.



## 2. Processus autorégressif-moyenne mobile (ARMA)



## 2. Processus autorégressif-moyenne mobile (ARMA)



## 2.Processus autorégressif-moyenne mobile (ARMA)

Température : un AR(3)

### Commandes sur R

```
fit = arima0(airquality$Temp[1:146], order=c(3,0,0)).
```

Call: arima0(x = airquality\$Temp, order = c(3, 0, 0))

Coefficients:

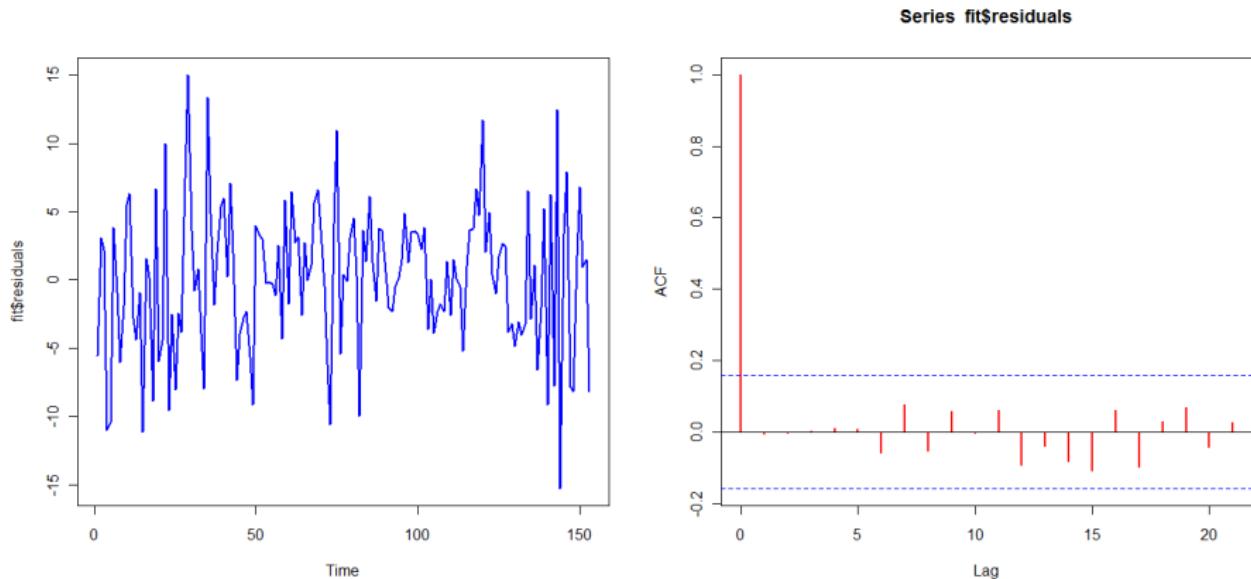
ar1	ar2	ar3	intercept
-----	-----	-----	-----------

0.6302	0.0735	0.1711	76.8909
--------	--------	--------	---------

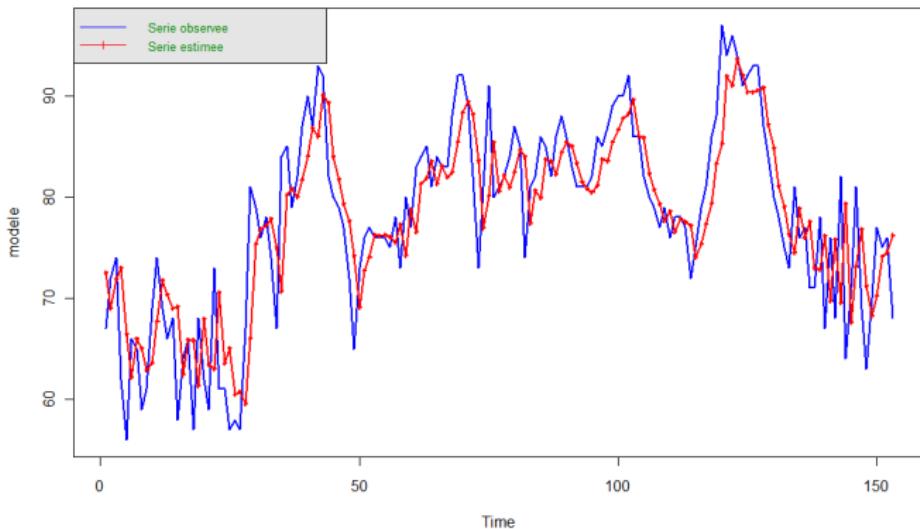
s.e.	0.0801	0.0951	0.0799	3.2257
------	--------	--------	--------	--------

$\sigma^2$  estimated as 27.9: log likelihood = -472.37, aic = 954.73

## 2. Processus autorégressif-moyenne mobile (ARMA)



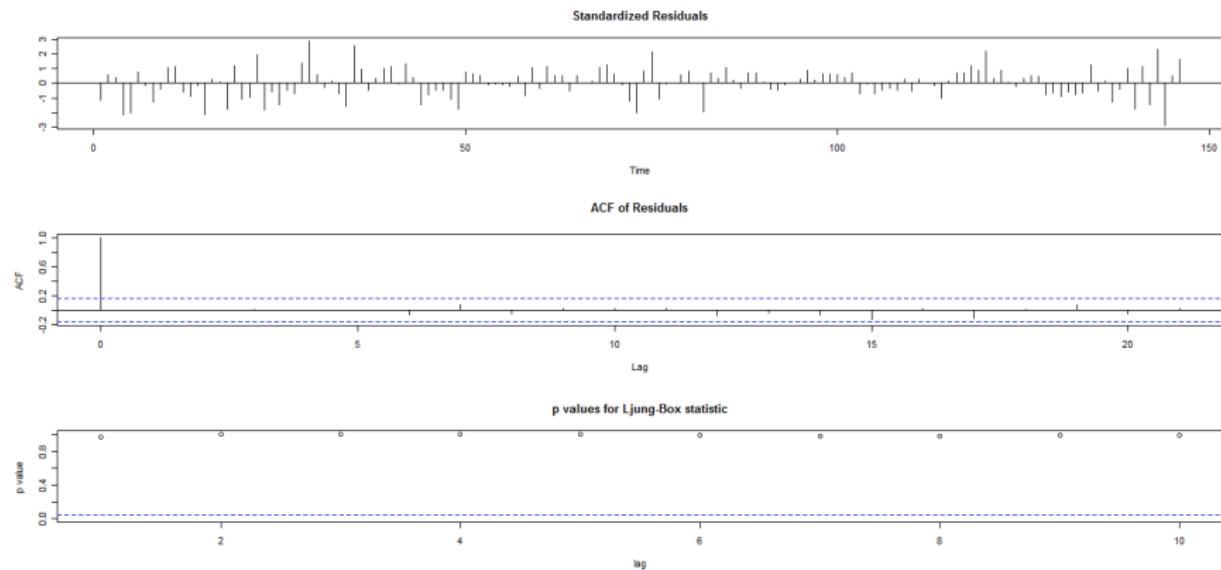
## 2. Processus autorégressif-moyenne mobile (ARMA)



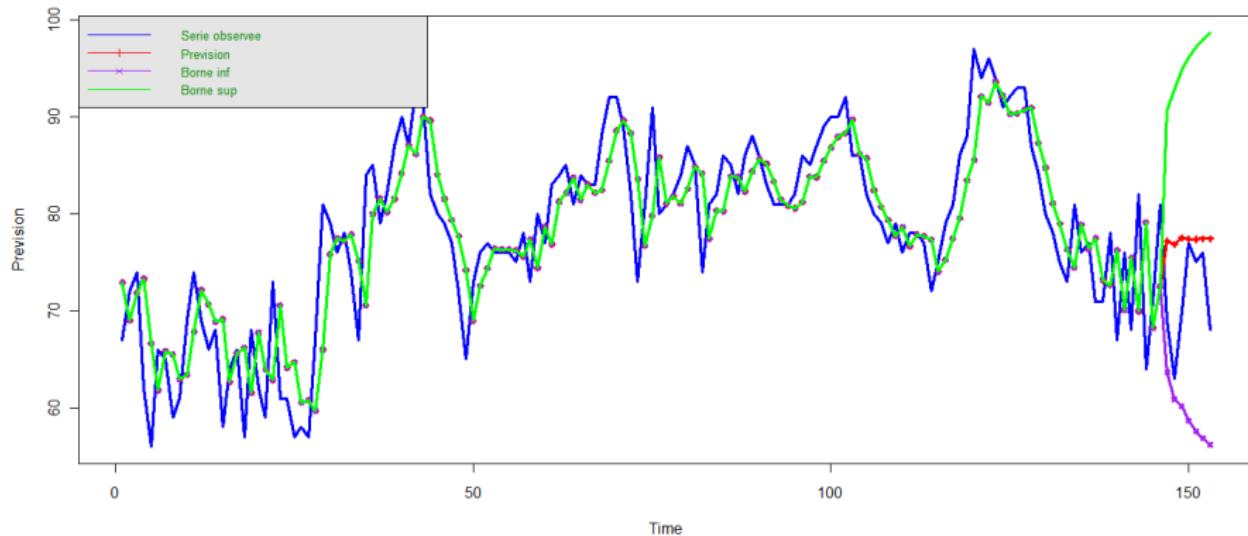
## 2. Processus autorégressif-moyenne mobile (ARMA)

Commandes sur R

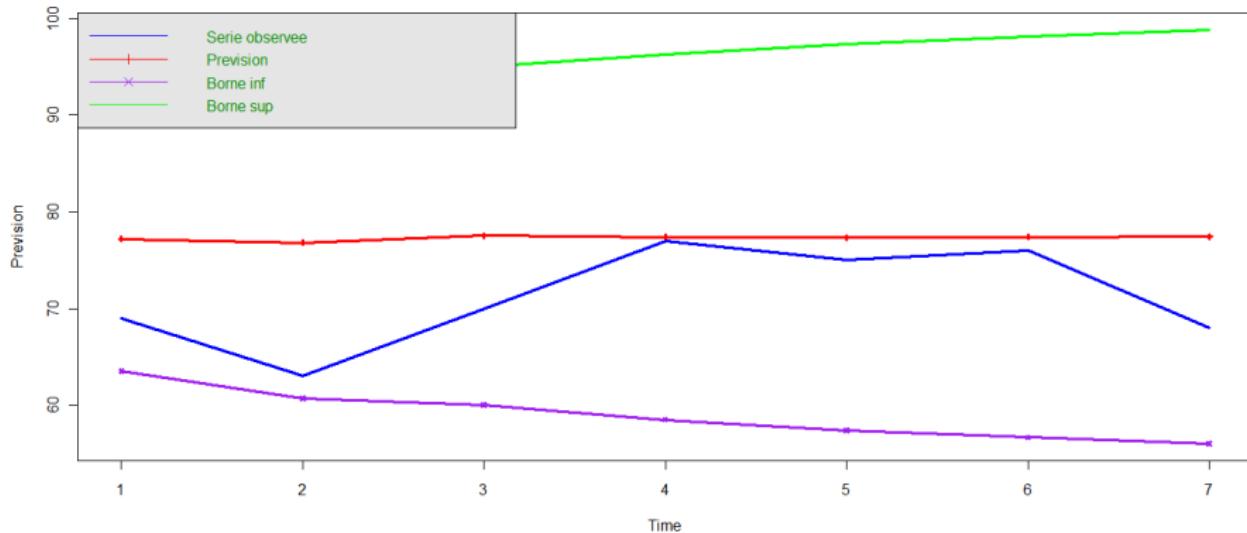
`tsdiag(fit)`



## 2. Processus autorégressif-moyenne mobile (ARMA)



## 2. Processus autorégressif-moyenne mobile (ARMA)



## 2. Processus autorégressif-moyenne mobile (ARMA)

Vent: un AR(1)

Commandes sur R

```
fit = arima0(airquality$Wind[1:146], order=c(1,0,0)).
```

Call:

```
arima0(x = airquality$Wind[1:143], order = c(1, 0, 0))
```

Coefficients:

ar1    intercept

0.3445    9.8577

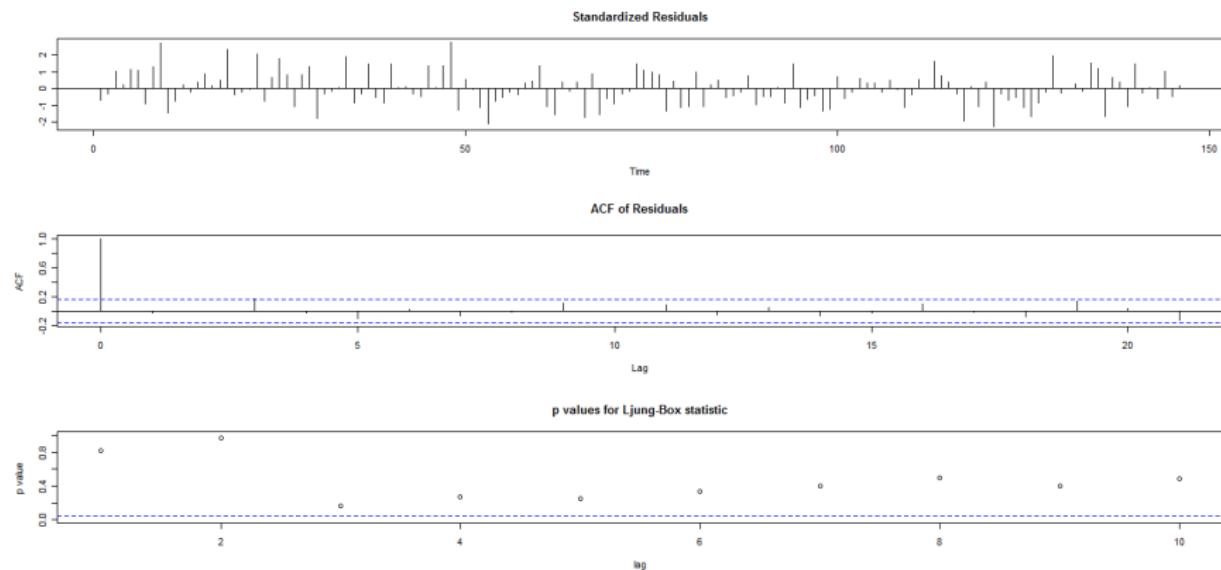
s.e. 0.0783    0.4217

$\sigma^2$  estimated as 11.01: log likelihood = -374.48, aic = 754.96

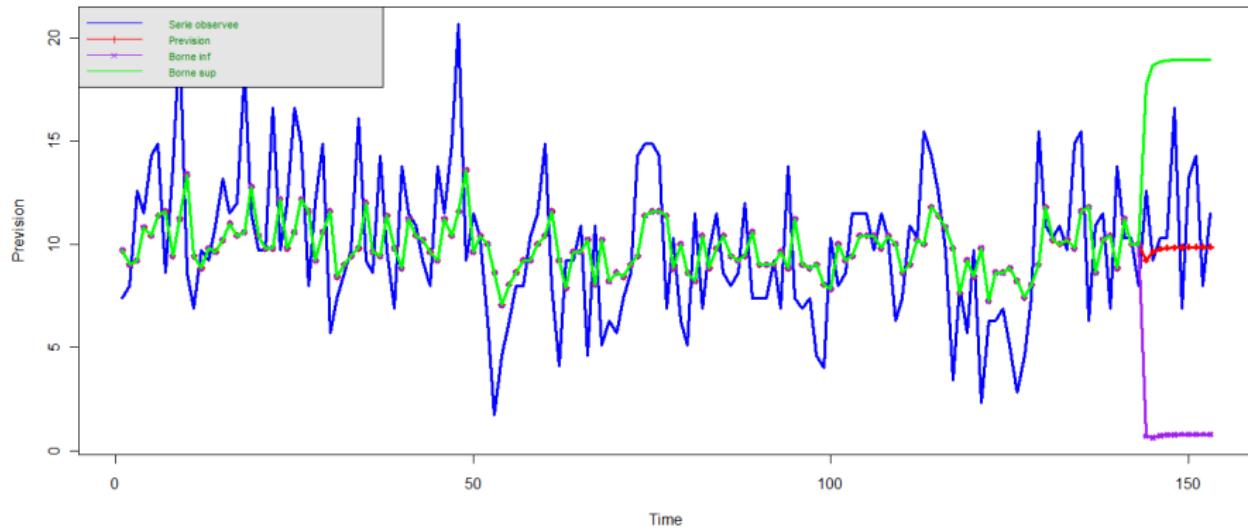
## 2. Processus autorégressif-moyenne mobile (ARMA)

Commandes sur R

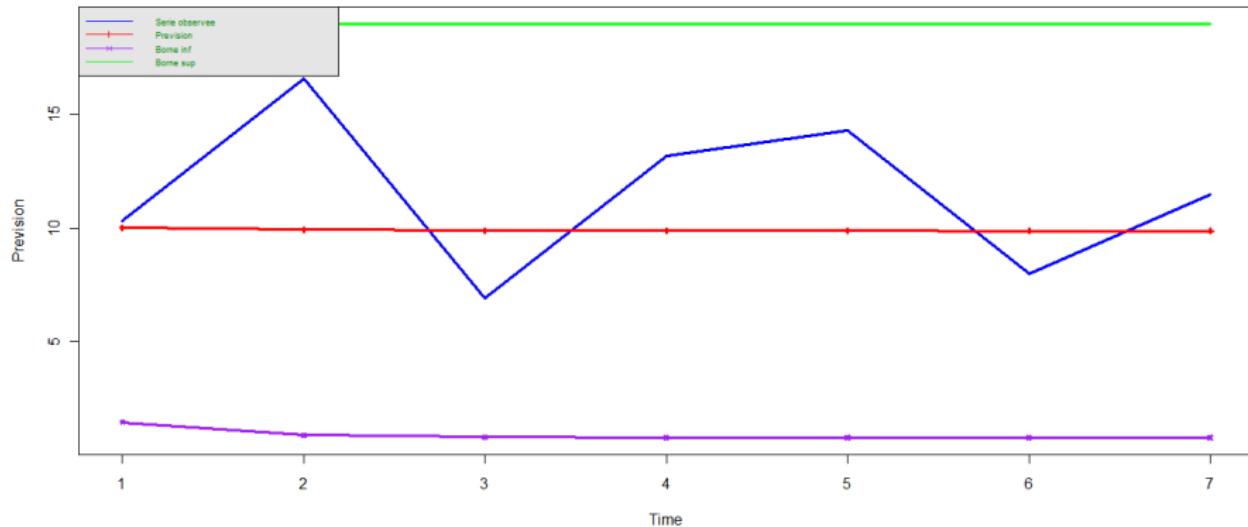
`tsdiag(fit)`



## 2. Processus autorégressif-moyenne mobile (ARMA)



## 2. Processus autorégressif-moyenne mobile (ARMA)



## 2. Processus autorégressif-moyenne mobile (ARMA)

### Mesure de la qualité de prévision

Dans l'échantillon initial,  $(X_1, \dots, X_T)$  on considère seulement ,  $T_1 = [(1 - \varepsilon)T]$  observations avec  $\varepsilon > 0$ . Les  $L = T - [(1 - \varepsilon)T]$  seront à prévoir par le modèle. On peut alors considérer plusieurs critères:

#### 1. Mean Absolute Percentage Error

$$MAPE = \frac{1}{L} \sum_{r=1}^L \left| \frac{X_{T_1+r} - \hat{X}_{T_1+r/T_1}}{X_{T_1+r}} \right|$$

#### 2. Mean Square Error

$$MSE = \left( \sum_{r=1}^L \frac{(X_{T_1+r} - \hat{X}_{T_1+r/T_1})^2}{L} \right)^{1/2}.$$



### 3.Processus non stationnaires ARIMA

- Les séries temporelles sont souvent non stationnaires.

#### Modèles ARIMA

##### Definition

Nous dirons que  $X = (X_t)$  est un processus ARIMA( $p,d,q$ ), s'il existe un entier naturel  $d$  tel que le processus  $Y_t = (1 - B)^d X_t$  est un ARMA( $p,q$ ),

$$(1 - B)^d A(B)X_t = C(B)\varepsilon_t \quad (11)$$

### 3.Processus non stationnaires ARIMA

MODELISATION ARIMA: Méthode de Box et Jenkins.

Etapes

- **Stationnarisation par différentiation.**
  - L'opérateur de différentiation  $\Delta = 1 - B$ , ( ou  $\Delta^d$ ) élimine les tendances,  $d$  est estimé en effectuant des tests de stationnarité sur la série brute puis sur les séries résiduelles. Un estimateur de  $d$  est le nombre total de fois où on rejette la stationnarité.
  - La série résiduelle stationnaire sera modélisée par un ARMA.
- **Estimation:** R offre la possibilité d'utiliser deux méthodes: la méthode de maximum de vraisemblance "ML" et la méthode des moindres carrés conditionnels "CSS"
- **Validation:** Le modèle retenu est-il compatible?: Tests d'adéquation sur les résidus: absence d'autocorrélation,

### 3.Processus non stationnaires ARIMA

1

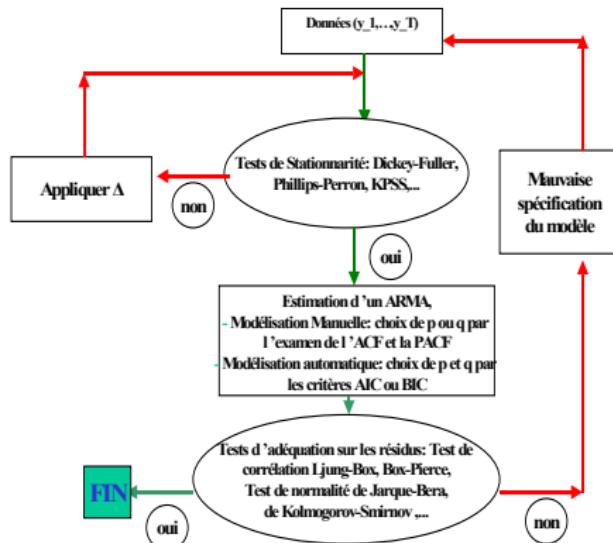


Schéma général de la modélisation  
d'une série temporelle par un modèle ARIMA.

### 3.Processus non stationnaires ARIMA

Test de non-stationnarité

Test de Dickey et Fuller augmenté:

Robuste à l'autocorrélation par rapport au test de Dickey-Fuller

$$\begin{cases} H_0 : y_t \sim I(1) \\ H_1 : y_t \text{ n'est pas } I(1) \end{cases}$$

On exécute la régression

$$\Delta y_t = \beta' D_t + \pi y_{t-1} + \sum_{j=1}^p \psi_j \Delta y_{t-j} + u_t,$$

$$ADF = T\hat{\pi}/(1 - \hat{\psi}_1 - \dots - \hat{\psi}_p).$$

### 3.Processus non stationnaires ARIMA

#### Test de Phillips-Perron

Robuste à l'hétérosécédasticité. On effectue la régression

$$\Delta y_t = \beta' D_t + \pi y_{t-1} + u_t,$$

$$PP = T\hat{\pi} - \frac{1}{2} T^2 \frac{SE(\hat{\pi})}{\hat{\sigma}^2} (\hat{\lambda}^2 - \hat{\sigma}^2),$$

$\hat{\sigma}^2 = \frac{1}{T-k} \sum_{t=1}^T \hat{u}_t^2$ ,  $\hat{u}_t$  : résidus de la régression,

$SE(\hat{\pi})$  = écart-type de  $\hat{\pi}$ ,

$\hat{\lambda}^2 = \hat{\sigma}^2 + 2 \sum_{j=1}^q (1 - \frac{j}{q+1}) \hat{\gamma}(j)$ , (*Newey – West*),

$\hat{\gamma}(j) = \frac{1}{T} \sum_{t=j+1}^T \hat{u}_t \hat{u}_{t-j}$ .

### 3.Processus non stationnaires ARIMA

Valeurs critiques de ADF et PP :  $P(ADF < c) = \alpha$

- $D_t = 0$  (processus centré)

$\alpha$	T	25	50	100	250	500	$\infty$
0.025		-9.3	-9.9	-10.2	-10.3	-10.4	-10.5
0.05		-7.3	-7.7	-7.9	-8	-8	-8.1
0.95		1.40	1.35	1.31	1.28	1.28	1.28
0.975		1.79	1.70	1.65	1.62	1.61	1.60

### 3.Processus non stationnaires ARIMA

- $D_t = 1$  (processus avec drift)

$\alpha$	T	25	50	100	250	500	$\infty$
0.025		-14.6	-15.7	-16.3	-16.6	-16.8	-16.9
0.05		-12.5	-13.3	-13.7	-14	-14	-14.1
0.95		0.01	-0.07	-0.10	-0.12	-0.13	-0.13
0.975		0.65	0.53	0.47	0.43	0.42	0.41

- $D_t = (1, t)'$  (processus avec tendance)

$\alpha$	T	25	50	100	250	500	$\infty$
0.025		-19.9	-22.4	-23.6	-24.6	-24.8	-25.1
0.05		-17.9	-19.8	-20.7	-21.3	-21.5	-21.8
0.95		-2.51	-2.60	-2.62	-2.64	-2.65	-2.66
0.975		-1.53	-1.66	-1.73	-1.78	-1.78	-1.79

### 3.Processus non stationnaires ARIMA

#### Tests de stationnarité

##### Test de KPSS

##### Modèle

$$y_t = \beta' D_t + \mu_t + u_t, \mu_t = \mu_{t-1} + \varepsilon_t, \text{var}(\varepsilon_t) = \sigma_\varepsilon^2.$$

$$\begin{cases} H_0 : \sigma_\varepsilon^2 = 0 \\ H_1 : \sigma_\varepsilon^2 \neq 0 \end{cases}$$

$$KPSS = \left( T^{-2} \sum_{t=1}^T \hat{S}_t^2 \right) / \hat{\lambda}^2.$$

$\hat{S}_t = \sum_{j=1}^t \hat{u}_j$ ,  $\hat{u}_t$  : résidus de la régression de  $y_t$  sur  $D_t$

### 3.Processus non stationnaires ARIMA

Valeurs critiques du KPSS:  $P(KPSS > c) = \alpha$

$D_t$	$\alpha$	0.10	0.05	0 .025	0.01
$D_t = 1$		0.347	0.436	0.574	0.739
$D_t = (1, t)'$		0.119	0.146	0.176	0.216

#### Tests d'adéquations

Modèle adéquat  $\implies$  la suite des résidus  $\hat{u}_t$  s'approche d'un bruit blanc.

#### Tests d'autocorrélation des résidus.

##### i. Tests Portmanteau

$$\begin{cases} H_0 : \hat{u}_t \text{ est un bruit blanc} \\ H_1 : \hat{u}_t \text{ n'est pas un bruit blanc} \end{cases}$$

### 3.Processus non stationnaires ARIMA

- Test de Box-Pierce (1970):

$$Q_{BP}(m) = T \sum_{j=1}^m \hat{\rho}^2(j),$$

- Test de Ljung-Box (1978):

$$Q_{LB}(m) = T(T+2) \sum_{j=1}^m \hat{\rho}^2(j)/(T-m),$$

$\hat{\rho}(j)$ : autocorrélation estimée de  $\hat{u}_t$

- Sous  $H_0$   $Q_{BP}(m)$  et  $Q_{LB}(m)$  suivent une loi  $X^2(m)$ .
- Si  $Q_{LB}(m) > X^2_{1-\alpha}(m)$  on rejette  $H_0$

### 3.Processus non stationnaires ARIMA

#### Tests de normalité

$$\left\{ \begin{array}{l} H_0 : \hat{u}_t \text{ est gaussien} \\ H_1 : \hat{u}_t \text{ n'est pas gaussien} \end{array} \right.$$

Test QQ-PLOT : (méthode graphique) : Le nuage de point est formé par ( quantiles de  $N(0,1)$ , quantiles empiriques réduits de  $\hat{u}_t$ ) , sous  $H_0$  le nuage est rectiligne sur la droite  $y=x$ ).

### 3.Processus non stationnaires ARIMA

#### Test de Jarque-Bera

$$JB = \frac{T}{6} \left( \hat{SK}^2 + \frac{(\hat{KU} - 3)^2}{4} \right).$$

$$\hat{SK} = \frac{\frac{1}{T} \sum_{t=1}^T (\hat{u}_t - \bar{u})^3}{\left( \frac{1}{T} \sum_{t=1}^T (\hat{u}_t - \bar{u})^2 \right)^{3/2}}, \text{ estimateur du Skewness}.$$

$$\hat{KU} = \frac{\frac{1}{T} \sum_{t=1}^T (\hat{u}_t - \bar{u})^4}{\left( \frac{1}{T} \sum_{t=1}^T (\hat{u}_t - \bar{u})^2 \right)^2}, \text{ estimateur du Kurtosis},$$

$$\bar{u} = \frac{1}{T} \sum_{t=1}^T \hat{u}_t. \text{ Sous } H_0 : JB \sim X^2(2).$$

### 3.Processus non stationnaires ARIMA

Test de Kolmogorov-Smirnov:

$$\begin{aligned} KS_T &= \sup_x |F_T(x) - \Phi(x)| \\ &= \sup_i \left\{ \left| \Phi(\hat{u}_t^{(i)}) - \frac{i}{T} \right| ; \left| \Phi(\hat{u}_t^{(i)}) - \frac{i-1}{T} \right| \right\} \end{aligned}$$

$$F_T(x) = \frac{1}{T} \sum_{t=1}^T 1_{\{\hat{u}_t \leq x\}} = \frac{\text{nombre de } \hat{u}_t \text{ inférieur } x}{T},$$

fonction de répartition empirique de  $(\hat{u}_t)$ ,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du,$$

fonction de répartition de la loi normale  $N(0, 1)$ .

### 3.Processus non stationnaires ARIMA

$\hat{u}_t^{(1)} \leq \dots \leq \hat{u}_t^{(T)}$ , statistiques d'ordre.

Valeurs critiques de  $KS_T$   $P(KS_T > c) = \alpha$ .

$\alpha$	T	25	50	100	$T > 100$
0.01		0.3165	0.2260	0.1608	$1.629/\sqrt{T}$
0.05		0.2640	0.1884	0.134	$1.358/\sqrt{T}$

### 3.Processus non stationnaires ARIMA

Exemples. 1. Etude du prix hébdomadaire du pétrole Brent.

Echantillon d'apprentissage : 1275 semaines.

Echantillon témoin : 4 semaines.



### 3.Processus non stationnaires ARIMA

- Tests de Stationnarité

Commandes sur R

```
library(tseries), adf.test(Brent)
```

Augmented Dickey-Fuller Test

data: Brent

Dickey-Fuller = -3.1179, Lag order = 10, p-value = 0.1051

alternative hypothesis: stationary

**On ne rejette pas l'hypothèse nulle de non stationnarité.**

### 3.Processus non stationnaires ARIMA

#### Commandes sur R

```
library(tseries), pp.test(brent)
```

Phillips-Perron Unit Root Test

data: Brent

Dickey-Fuller Z(alpha) = -10.7538, Truncation lag parameter = 7,

p-value = 0.5099

alternative hypothesis: stationary

### 3.Processus non stationnaires ARIMA

#### Commandes sur R

```
library(tseries), kpss.test(Brent)
```

KPSS Test for Level Stationarity

data: Brent

KPSS Level = 9.6879, Truncation lag parameter = 8, p-value =  
0.01

Message d'avis :

In kpss.test(Brent) : p-value smaller than printed p-value.

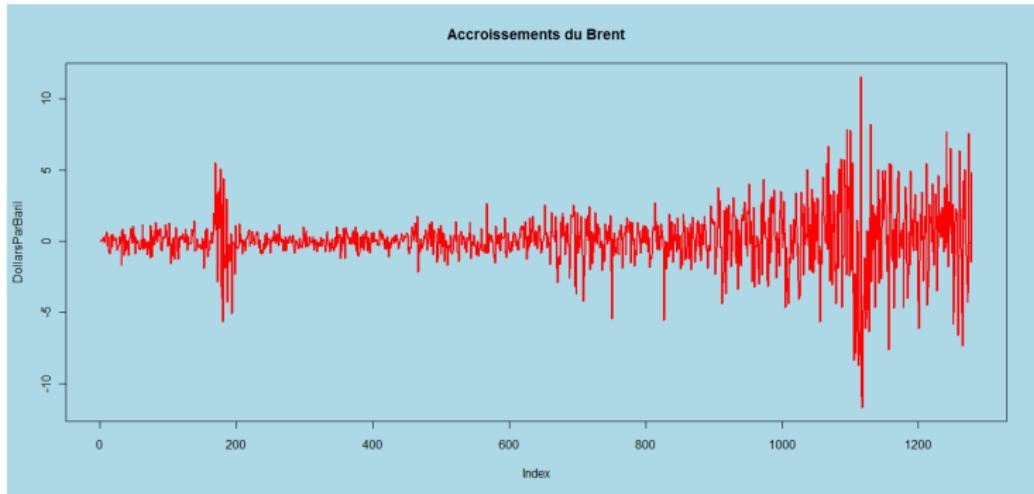
Conclusion: **Tous les tests confirment la non stationnarité du  
prix du pétrole.**

### 3.Processus non stationnaires ARIMA

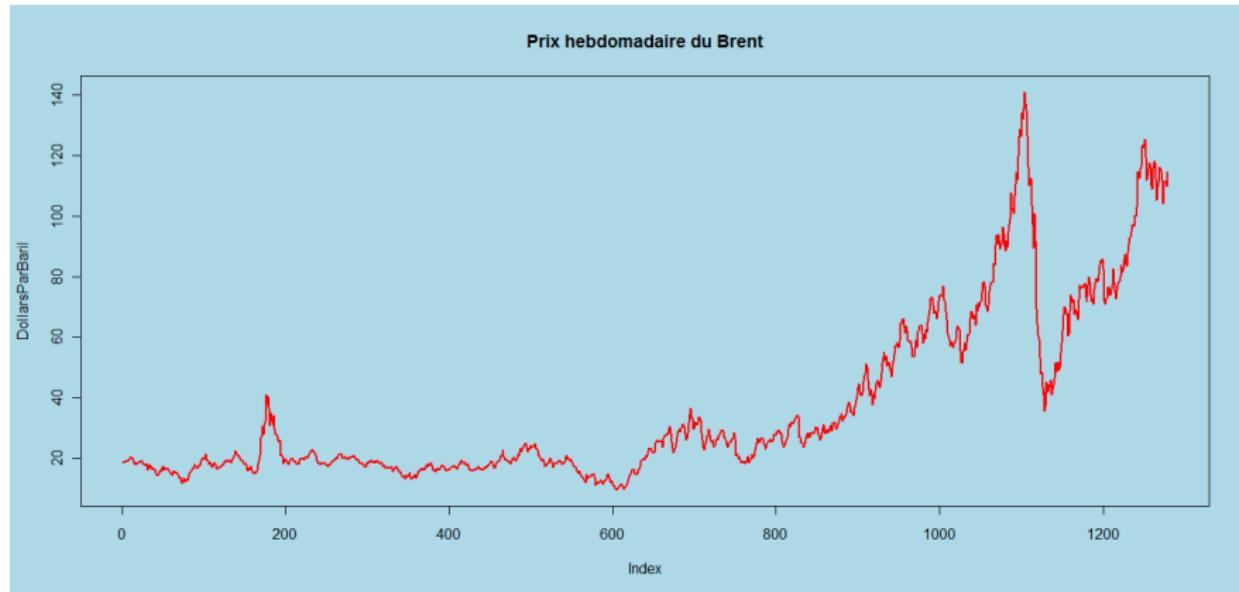
- Stationnarisation

#### Commandes sur R

```
dBrent <- diff(Brent), plot(dBrent)
```



### 3.Processus non stationnaires ARIMA



### 3.Processus non stationnaires ARIMA

- Tests de Stationnarité de la série différentiée

Commandes sur R

```
library(tseries), adf.test(dBrent)
```

Augmented Dickey-Fuller Test

data: dBrent

Dickey-Fuller = -8.8055, Lag order = 10, p-value = 0.01

alternative hypothesis: stationary

Message d'avis :

In adf.test(dBrent) : p-value smaller than printed p-value.

**On ne rejette pas la stationnarité.**

### 3.Processus non stationnaires ARIMA

#### Commandes sur R

```
library(tseries), pp.test(dBrent)
```

Phillips-Perron Unit Root Test

data: dBrent

Dickey-Fuller Z(alpha) = -1095.299, Truncation lag parameter = 7, p-value = 0.01

alternative hypothesis: stationary

Message d'avis :

In pp.test(dBrent) : p-value smaller than printed p-value

### 3.Processus non stationnaires ARIMA

#### Commandes sur R

```
library(tseries), kpss.test(dBrent)
```

KPSS Test for Level Stationarity

data: dBrent

KPSS Level = 0.155, Truncation lag parameter = 8, p-value = 0.1

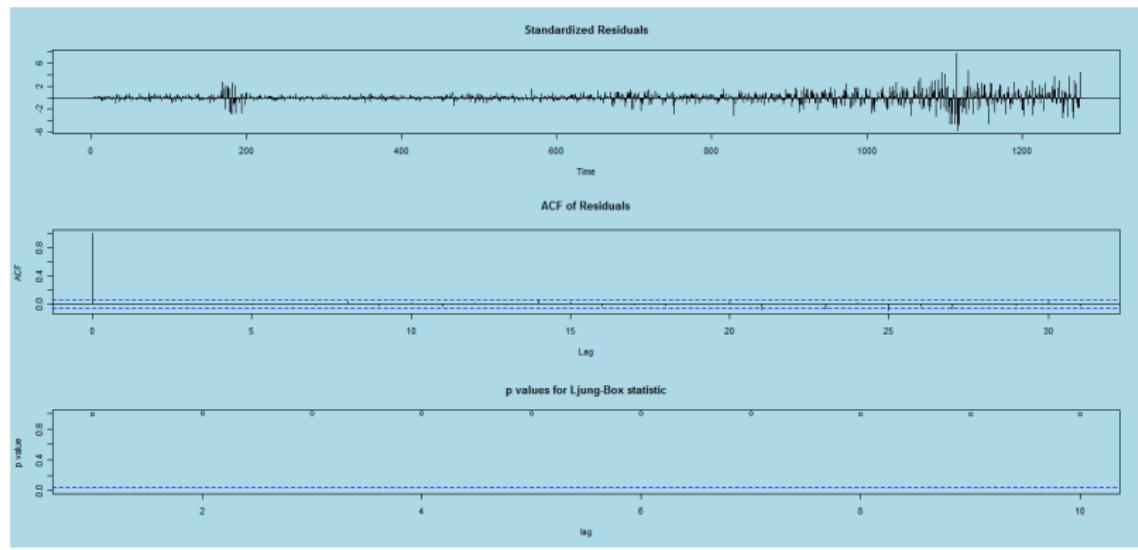
Message d'avis :

In kpss.test(dBrent) : p-value greater than printed p-value

- Conclusion:**Tous les tests confirment la stationnarité des accroissements du prix du pétrole dBrent.**
- **La série dBrent peut être modélisée alors par un ARMA.**

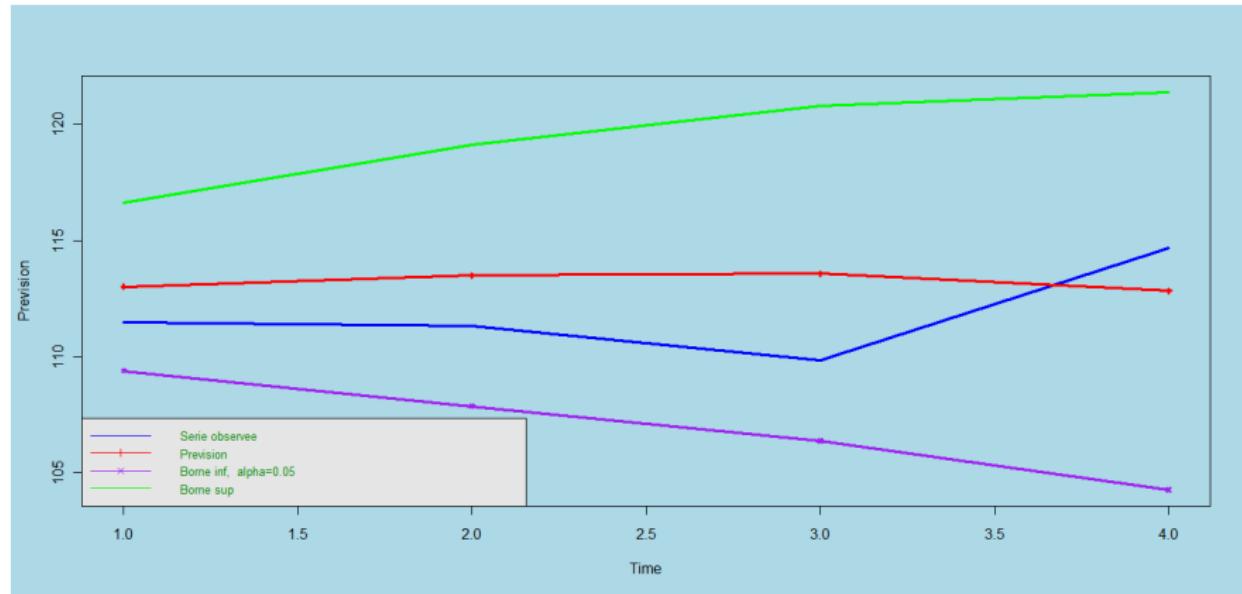
### 3.Processus non stationnaires ARIMA

- En utilisant le critère d'information AIC, on obtient un ARMA(5,7) pour dBrent soit un ARIMA(5,1,7) pour le prix du pétrole Brent sur l'échantillon  $(X_1, \dots, X_{1275})$ .



### 3.Processus non stationnaires ARIMA

Prévision de l'échantillon témoin  $X_{1276}, \dots, X_{1279}$  :



### 3.Processus non stationnaires ARIMA

- Test d'adéquation

#### 1. Test d'autocorélation

##### Commandes sur R

```
fit7 <- arima0(Brent[1:1275], order = c(5, 1, 7)),  
Box.test(fit7$residuals, lag = 10, type = "Box-Pierce")
```

Box-Pierce test

data: fit7\$residuals

X-squared = 2e-04, df = 1, p-value = 0.9885

- **La p-value est grande donc les résidus ne sont pas corrélés .**

### 3.Processus non stationnaires ARIMA

#### 2. Test de normalité

Commandes sur R

```
library(nortest), lillie.test(fit7$residuals)
```

Lilliefors (Kolmogorov-Smirnov) normality test

data: fit7\$residuals

D = 0.1362, p-value < 2.2e-16

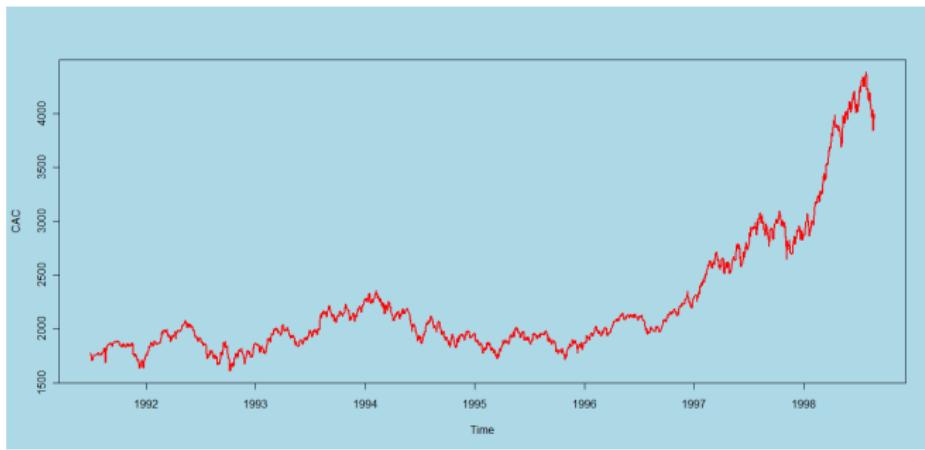
- **Donc les résidus ne suivent pas une loi de Gauss.**

### 3.Processus non stationnaires ARIMA

Exemple 2: Prix de clôture de l'indice boursier CAC40 sur la période 1991-1998 (1860 observations).

#### Commandes sur R

```
data(EuStockMarkets), CAC <- EuStockMarkrets[,3],  
ts.plot(CAC, type="l", col="red", lwd=2)
```



### 3.Processus non stationnaires ARIMA

#### Commandes sur R

```
library(tseries), adf.test(CAC)
```

Augmented Dickey-Fuller Test: data: CAC

Dickey-Fuller = -0.249, Lag order = 12, p-value = 0.99

alternative hypothesis: stationary

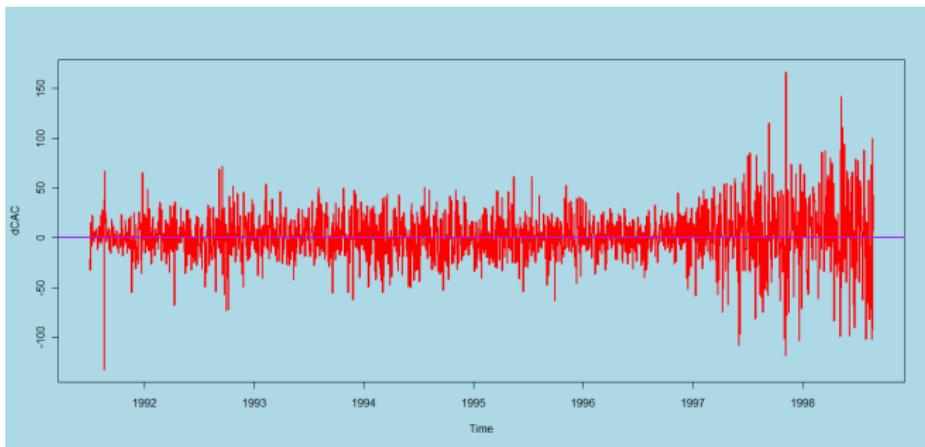
- **On ne rejette pas l'hypothèse de la non stationnarité de l'indice CAC40.**

### 3.Processus non stationnaires ARIMA

#### Stationnarisation

#### Commandes sur R

```
dCAC <- diff(CAC), plot(dCAC)
```



### 3.Processus non stationnaires ARIMA

Tests de Stationnarité de la série différentiée:

Commandes sur R

```
library(tseries), adf.test(dCAC)
```

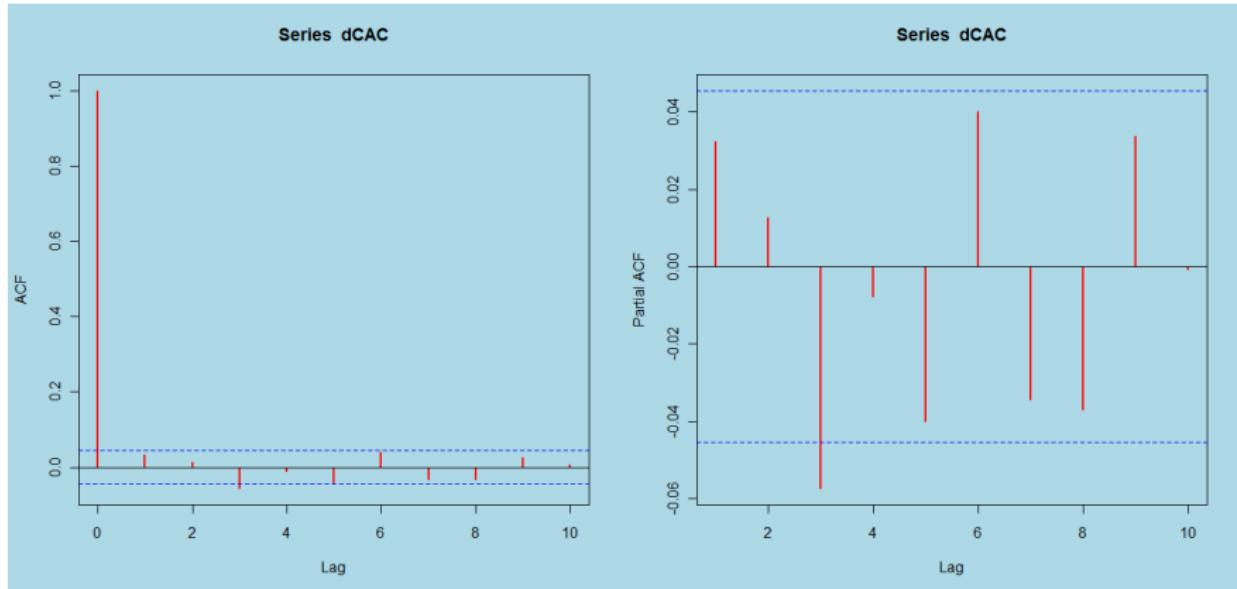
Augmented Dickey-Fuller Test: data: dCAC

Dickey-Fuller = -11.4471, Lag order = 12, p-value = 0.01

alternative hypothesis: stationary

- **La série différentiée est donc stationnaire.**

### 3.Processus non stationnaires ARIMA



- La fonction d'autocorrélation ACF nous indique que la série différentiée est un bruit blanc.

### 3.Processus non stationnaires ARIMA

Test d'autocorrélation:

Commandes sur R

```
Box.test(dCAC, lag = 10, type = "Box-Pierce")
```

Box-Pierce test : data: dCAC

X-squared = 19.9932, df = 10, p-value = 0.02932

- **Avec un risque de 1% on ne rejette pas l'hypothèse que les accroissements du CAC40 est un bruit blanc.**

### 3.Processus non stationnaires ARIMA

#### Test de normalité:

##### Commandes sur R

```
library(nortest), pearson.test(dCAC)
```

Pearson chi-square normality test:  
data: dCAC

P = 293.1802, p-value <2.2e-16

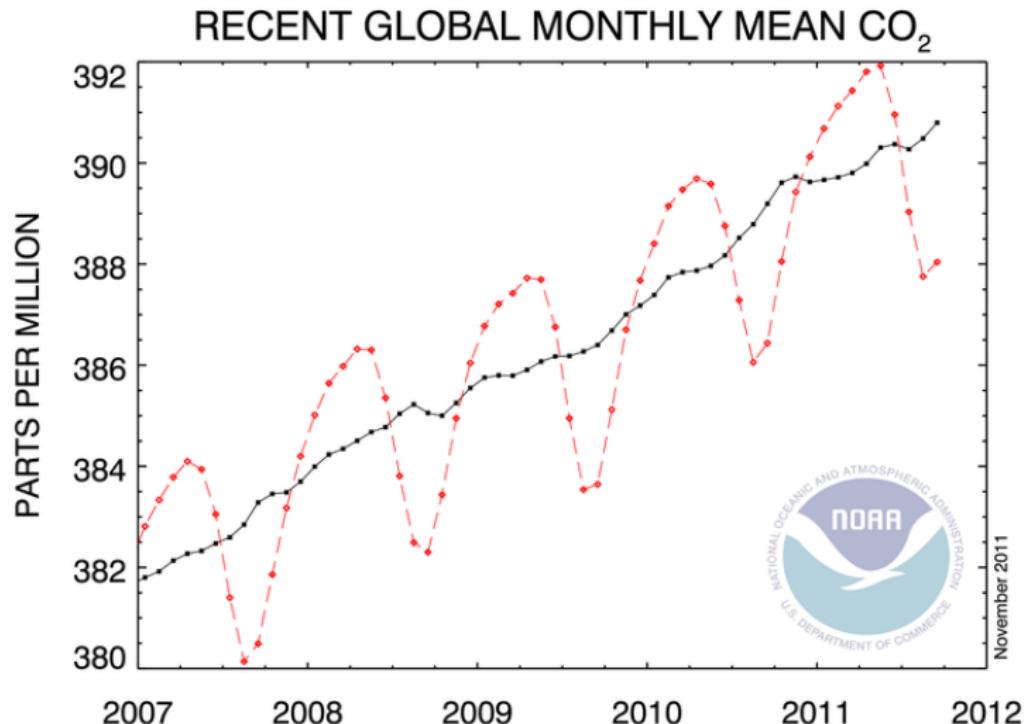
- **La normalité des accroissements est donc rejetée.**
- Le CAC40 est une marche aléatoire (donc une martingale),
- Les lois des accroissements ont des queues épaisses (Student, GEV,...),
- Les accroissements sont hétérosclélastiques (Engel (1982), modèle ARCH).

## 4.Processus non stationnaires SARIMA

SARIMA = "Seasonal-AutoRegressive-Integrated-Moving-Average"

- Ils sont utilisés pour modéliser les données saisonnières (mensuelles, trimestrielles) ou cycliques.
- Saisonnalité  $s$ : les portions de trajectoires obtenues par la translation  $X_t \rightarrow X_{t+s}$  ont de fortes similarités.

## 4.Processus non stationnaires SARIMA



## 4.Processus non stationnaires SARIMA

### Definition

Nous dirons que  $X = (X_t)$  est un processus SARIMA(p,d,s,q), s'il vérifie l'équation:

$$(1 - B)^d (1 - B^s) A(B) X_t = C(B) \varepsilon_t \quad (12)$$

### Modélisation par les modèles SARIMA

Souvent on utilise une structure plus générale, on suppose que la série (ou le processus la modélisant) satisfait l'équation:

$$(1 - B)^d (1 - B^s)^D \phi(B) \Phi(B^s) X_t = \theta(B) \Theta(B^s) \varepsilon_t \quad (13)$$

$p, P, d, s, D, q, Q$  sont des entiers naturels inconnus; les polynômes  $\phi(z), \Phi(z), \theta(z)$  et  $\Theta(z)$  sont à coefficients inconnus.

## 4.Processus non stationnaires SARIMA

$$\Phi(B^s) = 1 + \Phi_1 B^s + \dots + \Phi_P B^{sP}, \Theta(B^s) = 1 + \Theta_1 B^s + \dots + \Theta_Q B^{sQ}$$

Dans la structure (13), on appelle:

- $d$  : Différence,
- $s$  : saison,
- $D$  : différence saisonnière,
- $\phi(z)$  la partie autorégressive,
- $\Phi(z)$  la partie autorégressive saisonnière,
- $\Theta(z)$  la partie moyenne-mobile,
- $\theta(z)$  la partie moyenne-mobile saisonnière.

## 4.Processus non stationnaires SARIMA

### Commandes sur R

```
arima0(x, order = c(p, d, q), seasonal = list(order = c(P, D, Q),  
period = s), xreg = NULL, include.mean = TRUE, delta = 0.01,  
transform.pars = TRUE, fixed = NULL, init = NULL, method =  
c("ML", "CSS"), n.cond, optim.control = list())
```

- On utilise souvent  $\Theta(B^s)$  si on applique aux données le filtre  $(1 - B^s)^D$ .

## 4.Processus non stationnaires SARIMA

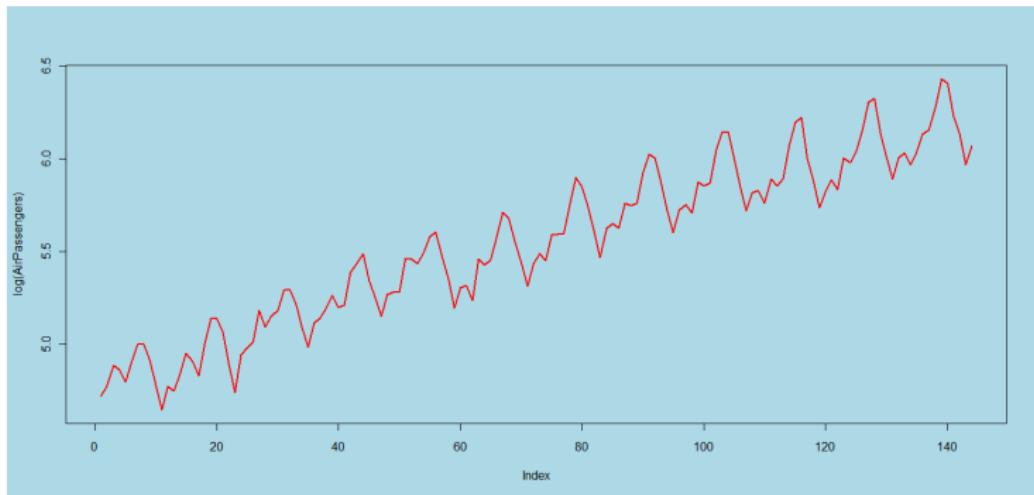
- **Modélisation:** identique à celle de la modélisation ARIMA sauf la première étape de stationnarisation est remplacée par la suivante:

### Stationnarisation par filtrage :

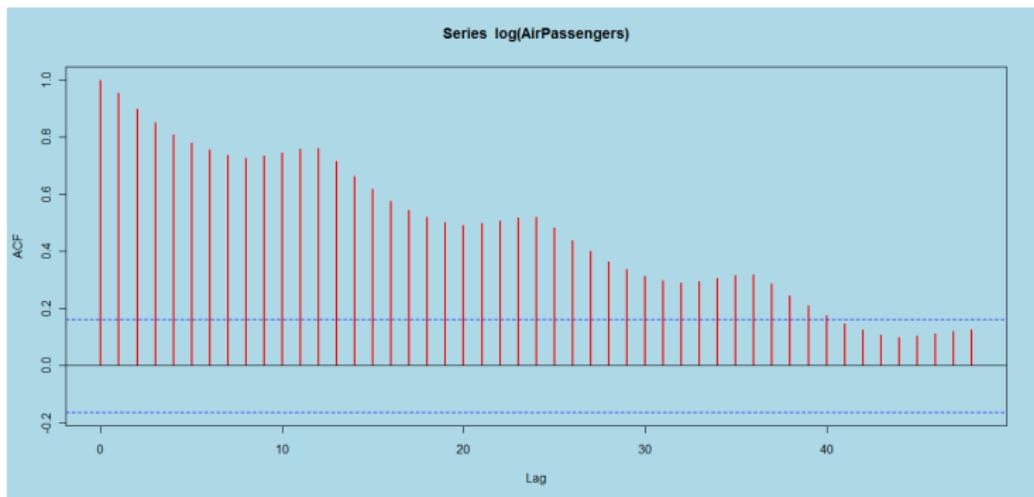
- L'opérateur  $(1 - B^s)^D$ , élimine la périodicité (désaisonalisation).
- L'opérateur de différentiation  $\Delta = 1 - B$ , ( ou  $\Delta^d$ ) élimine les tendances.

## 4.Processus non stationnaires SARIMA

Exemple: Modélisation de la série des voyageurs:



## 4.Processus non stationnaires SARIMA

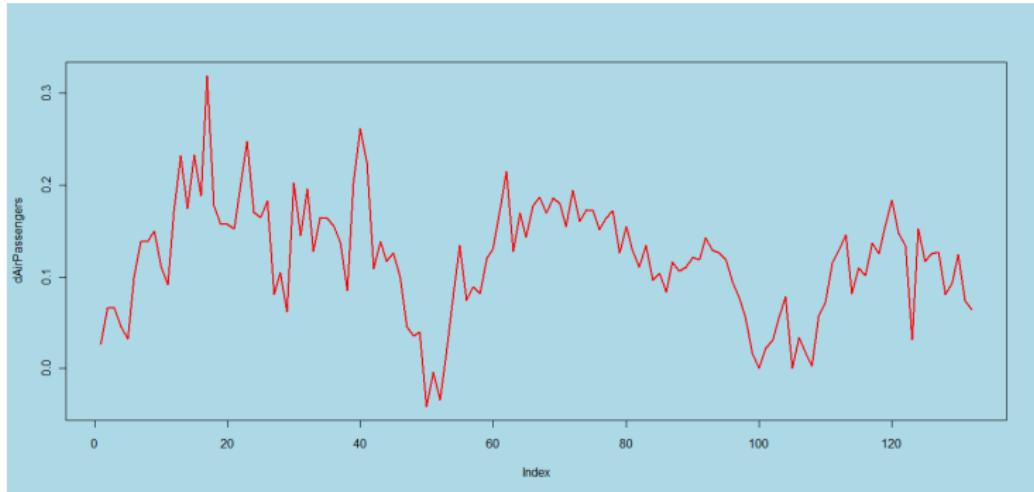


## 4.Processus non stationnaires SARIMA

Désaionalisation:

Commandes sur R

```
dAirPassengers <- diff(log(AirPassengers), lag=12)
```



## 4.Processus non stationnaires SARIMA

### Test de stationnarité

#### Commandes sur R

```
library(tseries), adf.test(dAirPassengers)
```

Augmented Dickey-Fuller Test:  
data: dAirPassengers

Dickey-Fuller = -3.1519, Lag order = 5, p-value = 0.09899

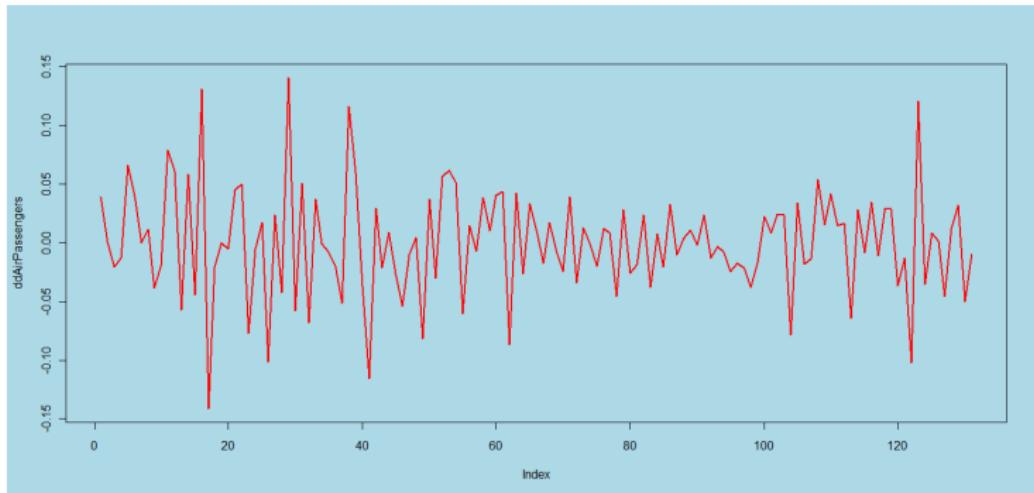
alternative hypothesis: stationary

- **La série dAirPassengers est non stationnaire.**

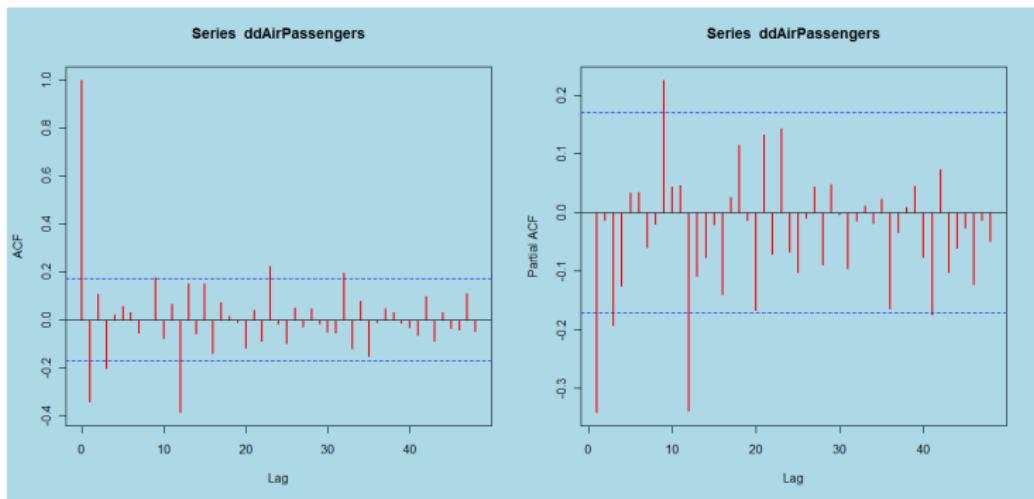
## 4.Processus non stationnaires SARIMA

### Commandes sur R

```
ddAirPassengers = diff(dAirPassengers), plot(ddAirPassengers)
```



## 4.Processus non stationnaires SARIMA



## 4.Processus non stationnaires SARIMA

### Commandes sur R

```
fit8= arima0(log(AirPassengers)[1:132],c(0,1,1), seasonal =  
list(order=c(0, 1 ,1), period=12))
```

Call:arima0(x =log(AirPassengers)[1:132], order = c(0, 1, 1),  
seasonal = list(order = c(0,1, 1), period = 12))

Coefficients:

ma1	sma1
-----	------

-0.3479	-0.5594
---------	---------

s.e.	0.0897	0.0713
------	--------	--------

$\sigma^2$  estimated as 0.001315: log likelihood = 223.62, aic = -441.24.

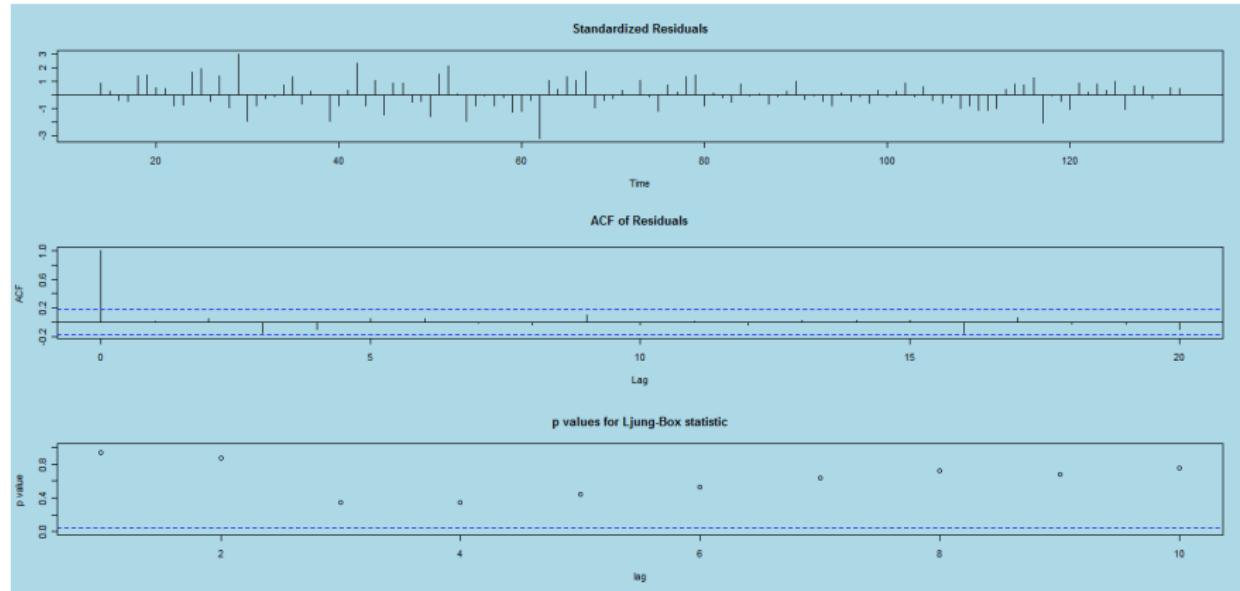
## 4.Processus non stationnaires SARIMA

Modèle proposé pour la série :

$$\begin{aligned} & (1 - B)(1 - B^{12}) \log(AirPassengers_t) \\ = & (1 - 0.3479B)(1 - 0.05594B^{12})\varepsilon_t \end{aligned}$$

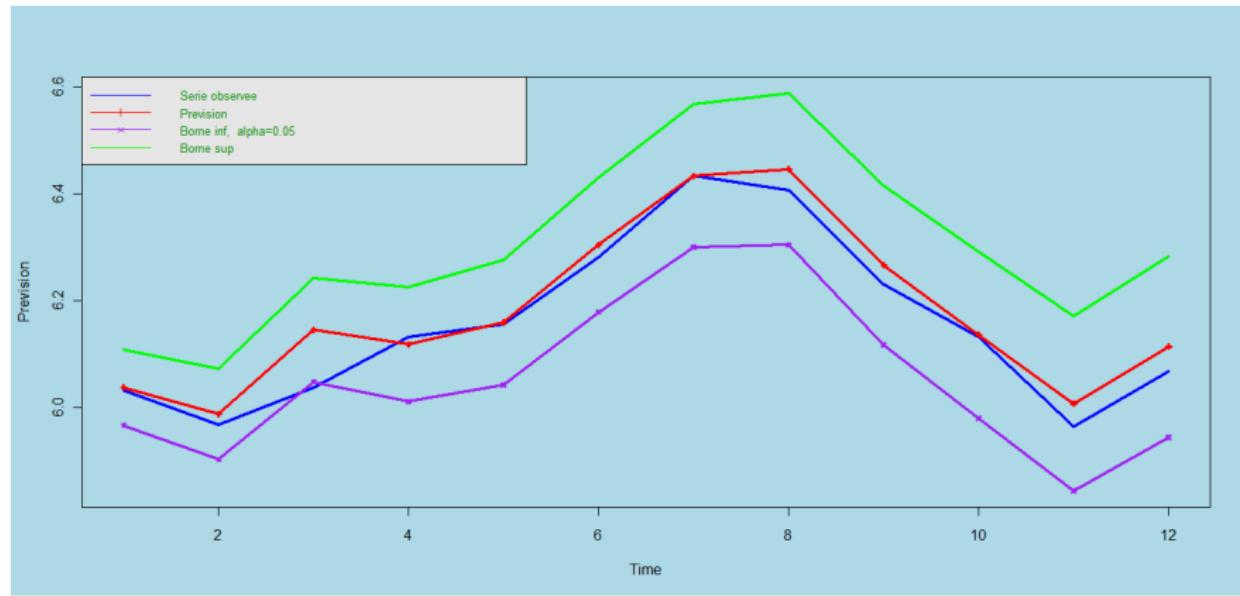
# 4.Processus non stationnaires SARIMA

Tests d'adéquation:



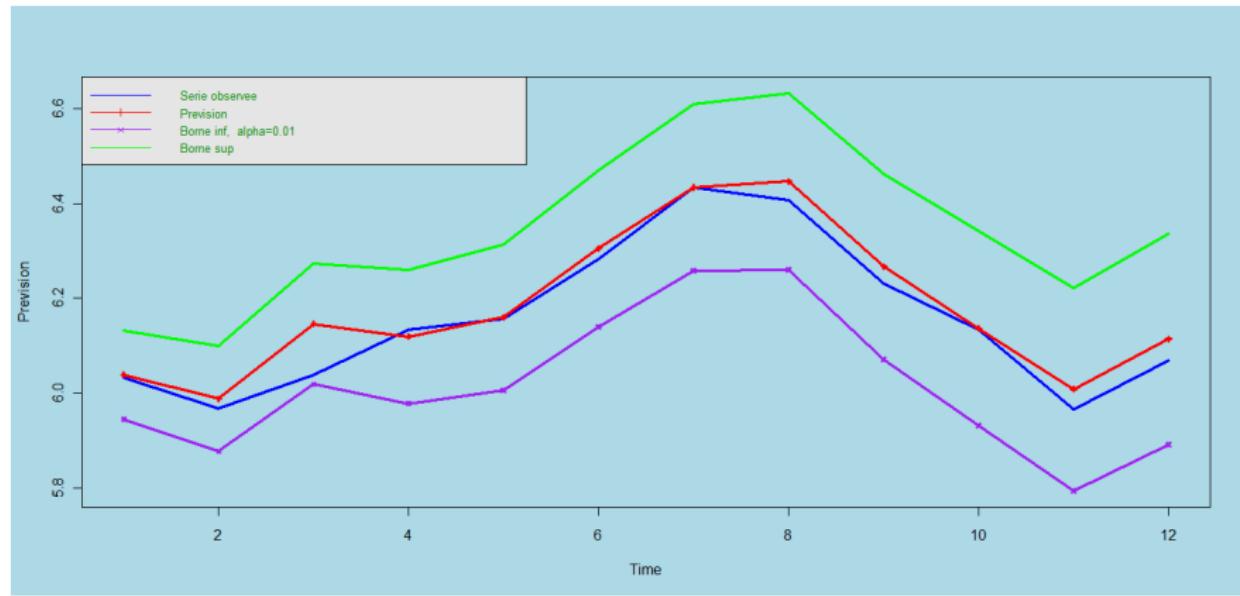
# 4.Processus non stationnaires SARIMA

## Prévision d'une année



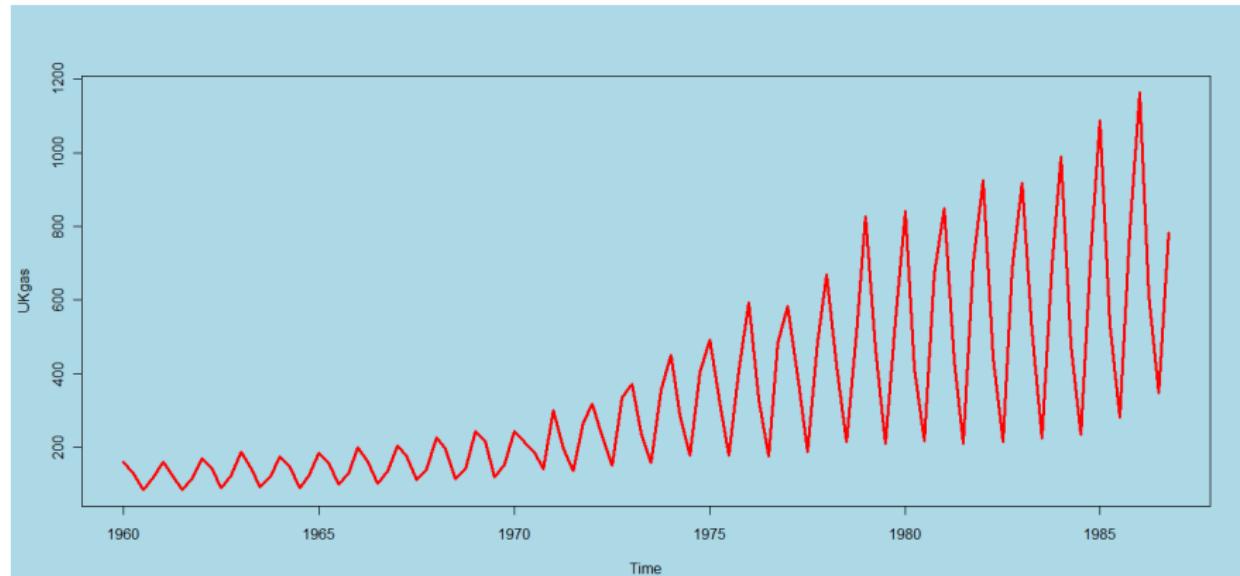
# 4.Processus non stationnaires SARIMA

## Prévision d'une année



## 4.Processus non stationnaires SARIMA

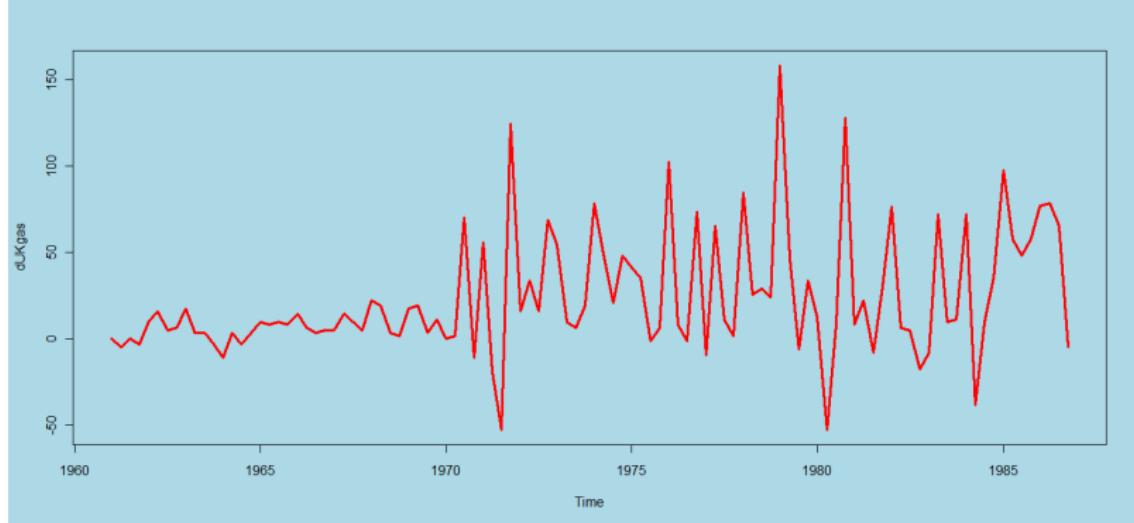
Exemple 2: Consommation trimestrielle du gaz. Echantillon d'apprentissage : 104 trimestres. Echantillon témoin : 4 trimestres.



## 4. Processus non stationnaires SARIMA

Commandes sur R

```
dUKgas <- diff(UKgas, lag=4)
```



- L'effet saisonnier semble disparaître

## 4.Processus non stationnaires SARIMA

### Commandes sur R

```
library(tseries), adf.test(dUKgas)
```

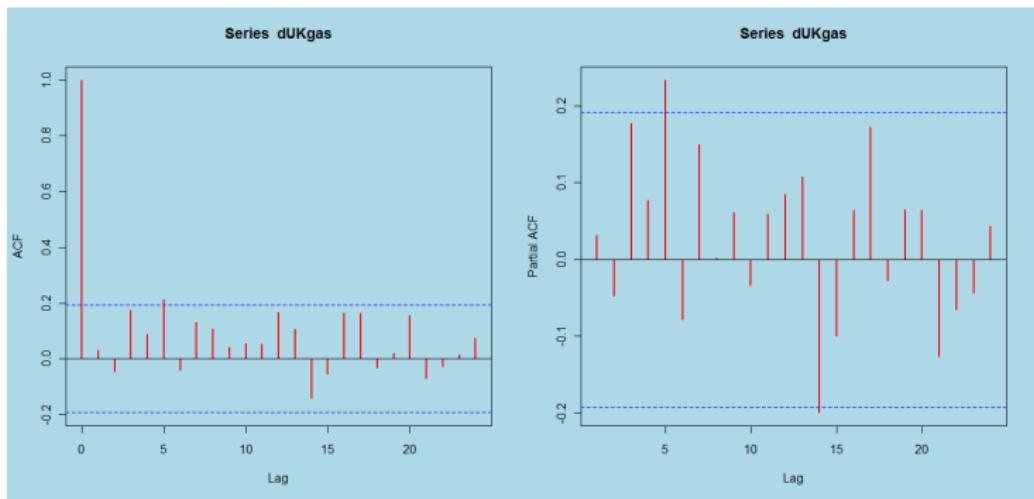
Augmented Dickey-Fuller Test: data: dUKgas

Dickey-Fuller = -3.7419, Lag order = 4, p-value = 0.02438

alternative hypothesis: stationary

- **Avec un risque de 5% on rejette la non stationnarité de dUKgas**

## 4.Processus non stationnaires SARIMA

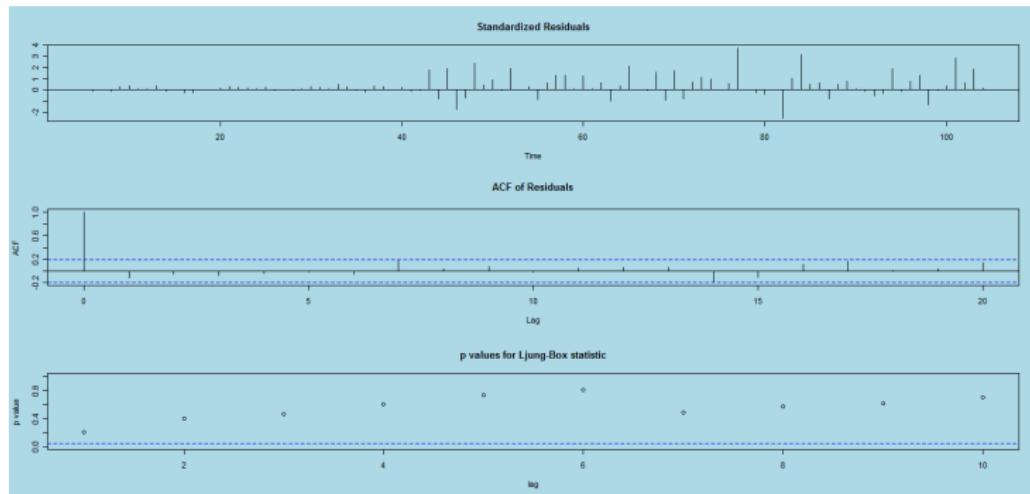


## 4.Processus non stationnaires SARIMA

dUKgas peut être modélisée par un AR(5).

### Commandes sur R

```
fit9= arima0(UKgas[1:104],c(0,0,5), seasonal = list(order=c(0, 1, 1), period=4)), tsdiag(fit9)
```



## 4.Processus non stationnaires SARIMA

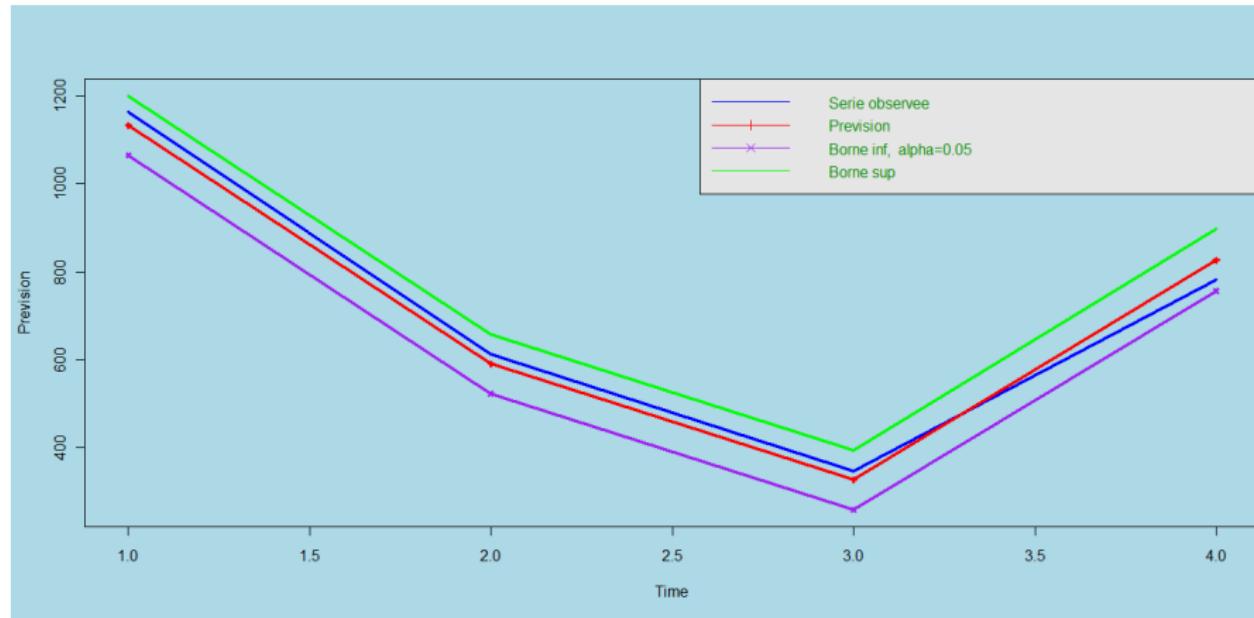
Donc le modèle pour la consommation du gaz est le suivant

$$A(B) (1 - B^4) UKgas_t = \varepsilon_t.$$

$$A(B) = 1 - 0.0547B + 0.0157B^2 - 0.2930B^3 - 0.1614B^4 - 0.3167B^5.$$

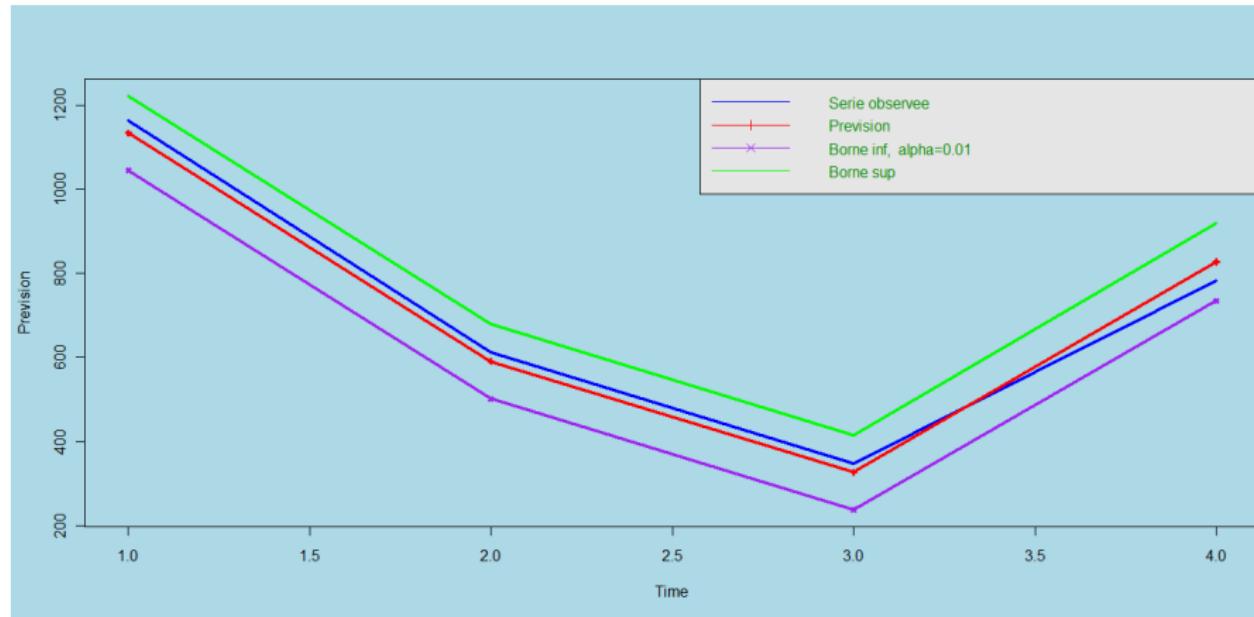
# 4.Processus non stationnaires SARIMA

## Prévision d'une année



## 4.Processus non stationnaires SARIMA

### Prévision d'une année



## Définitions

Soit  $X_t = (X_t^1, \dots, X_t^d)$  un processus vectoriel.

### Definition

On dit que  $(X_t)$  est au second ordre si

$$E(||X_t||^2) < \infty \quad \forall t. \quad (14)$$

### Definition

On dit que  $(X_t)$  est stationnaire si

- $E(X_t) = \mu \quad \forall t,$
- $\Gamma(j) = E(X_t - \mu)(X_{t-j} - \mu)' \quad \forall t.$

## 5.PROCESSUS VECTORIELS VARMA

Modèle  $VARMA_d(p, q)$  : Le modèle AutoRégressif-Moyenne mobile vectoriel est donné par

$$X_t + A_1 X_{t-1} + \dots + A_p X_{t-p} = \varepsilon_t + \Theta_1 \varepsilon_{t-1} + \dots + \Theta_q \varepsilon_{t-q}. \quad (15)$$

$(\varepsilon_t)$  est un bruit blanc centré de variance  $\Sigma$  :

$$E(\varepsilon_t) = 0, E\varepsilon_t\varepsilon_s' = \delta_t^s \Sigma,$$

$A_i, 1 \leq i \leq p, \Theta_j, 1 \leq j \leq q$ , sont des matrices  $(d, d)$ .

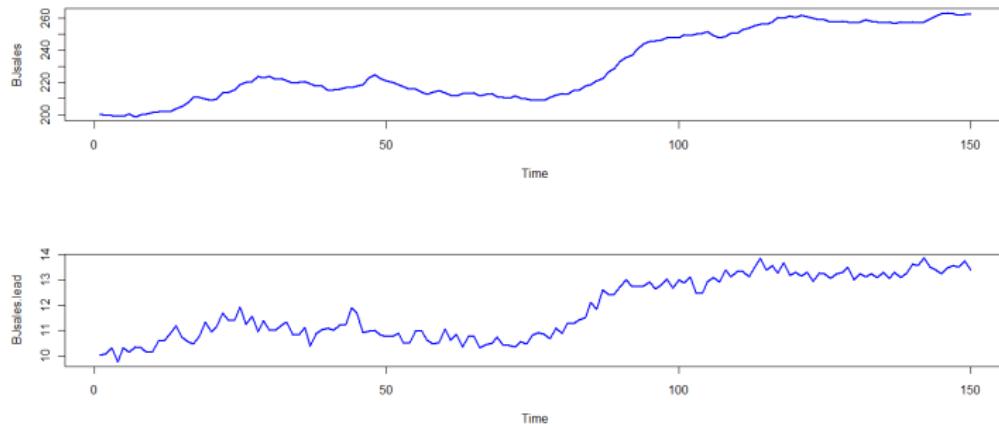
## 5.PROCESSUS VECTORIELS VARMA

Pour un modèle  $MA_d(q)$  on a

$$\Gamma(j) = \begin{cases} \sum_{k=j}^q \Theta_k \Sigma \Theta'_{k-j} & \text{si } 0 \leq j \leq q, (\Theta_0 = \mathbf{I}_d) \\ \Gamma'(-j) & \text{si } -q \leq j \leq 0, \\ 0 & \text{si } |j| > q. \end{cases}$$

## 5.PROCESSUS VECTORIELS VARMA

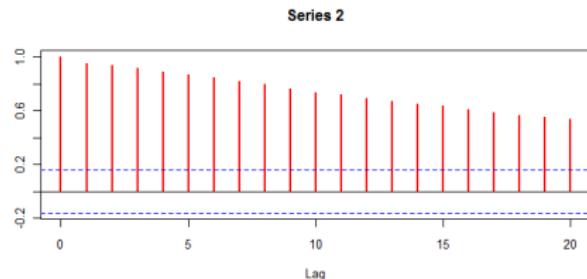
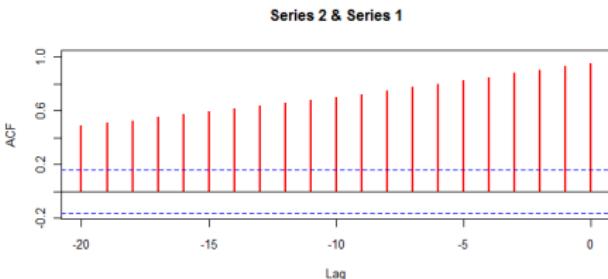
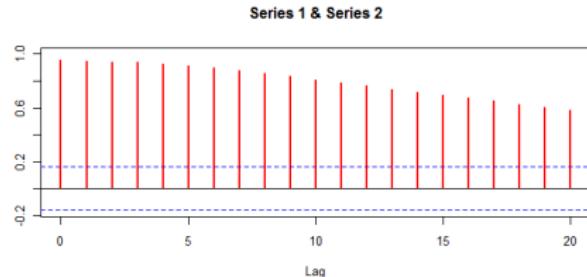
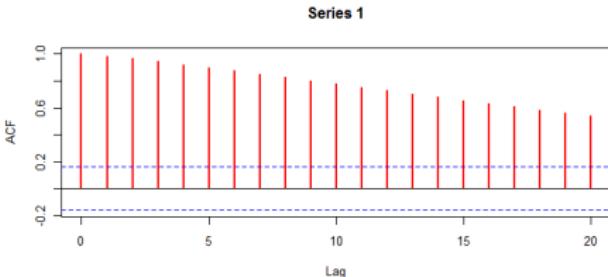
Estimation: Sur R on peut utiliser la fonction ar basée sur le critère AIC. Exemple:



### Commandes sur R

```
z <- matrix(rep(0,1),150,2), z[,1] <- BJsales , z[,2] <-  
BJsales.lead, acf(z,20,"correlation")
```

# 5.PROCESSUS VECTORIELS VARMA



Les deux séries semblent être non stationnaires.

## 4.PROCESSUS VECTORIELS VARMA

Commandes sur R

```
adf.test(z[,1]), adf.test(z[,2]) acf(z,20,"correlation")
```

Augmented Dickey-Fuller Test :data: z[, 1]

Dickey-Fuller = -2.1109, Lag order = 5, p-value = 0.5302

alternative hypothesis: stationary

Augmented Dickey-Fuller Test

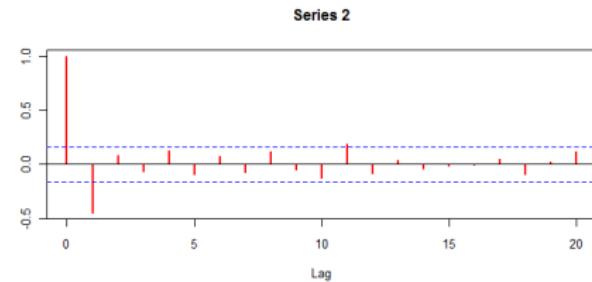
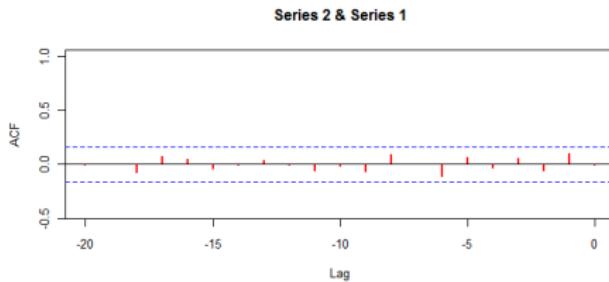
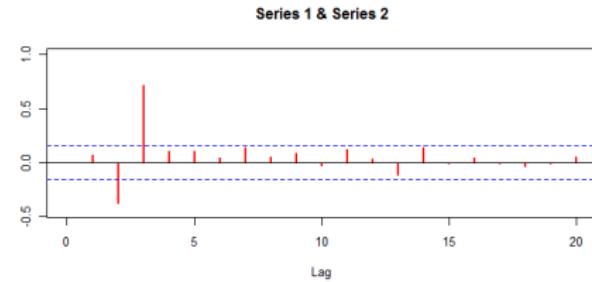
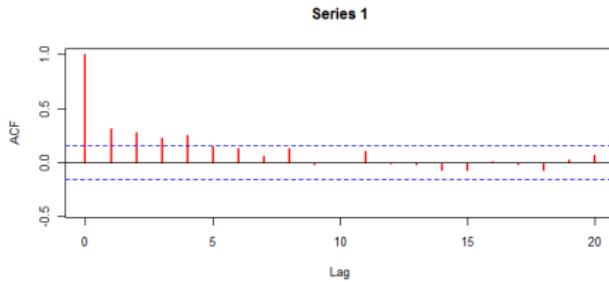
data: z[, 2]

Dickey-Fuller = -1.7237, Lag order = 5, p-value = 0.6915

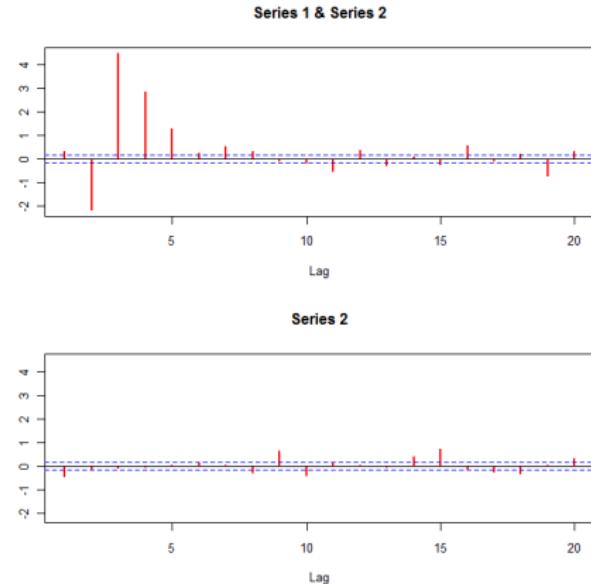
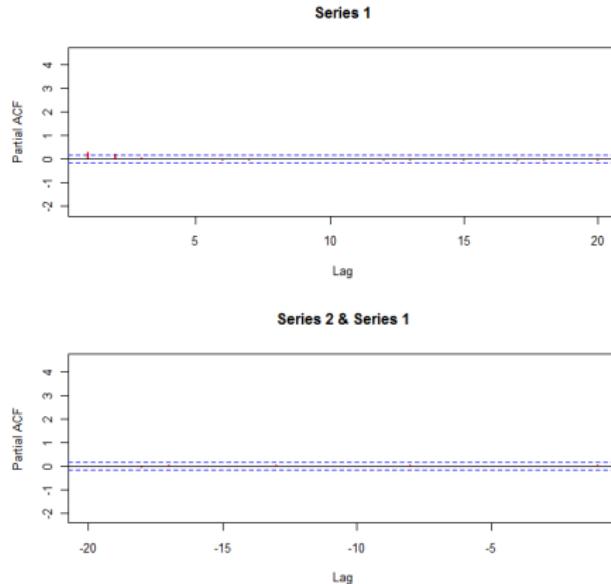
alternative hypothesis: stationary

**La présence de racine unitaire est confirmé par le test ADF.**

# 5.PROCESSUS VECTORIELS VARMA

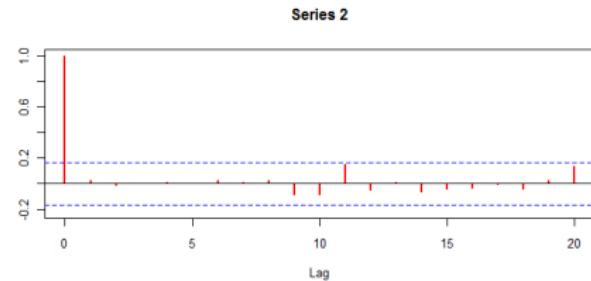
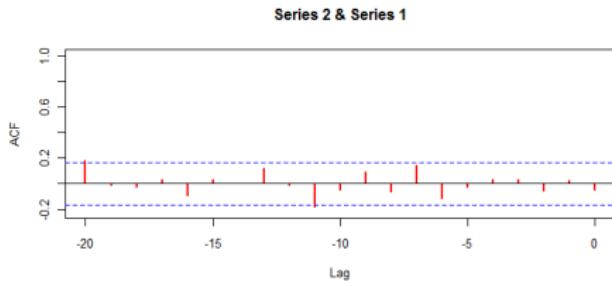
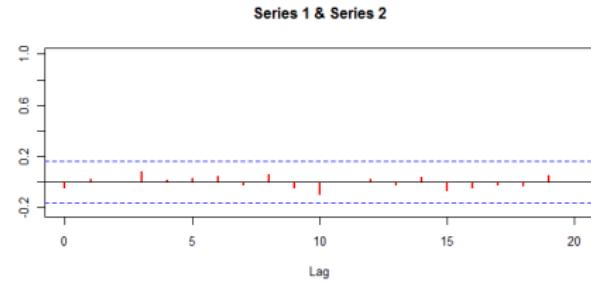
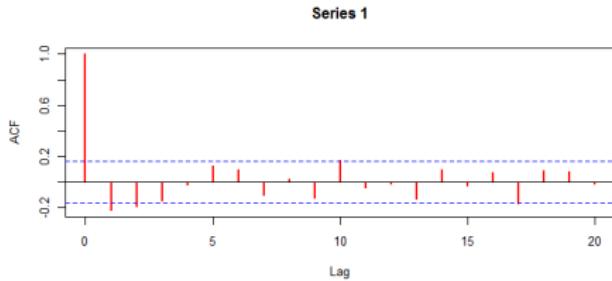


# 5.PROCESSUS VECTORIELS VARMA



## 5.PROCESSUS VECTORIELS VARMA

La fonction ar donne un VAR(5), on obtient les acfs des résidus suivantes:



## 6. Estimation de la densité de probabilité

**Problème** : Comment faire une inférence statistique quand on ne connaît pas la loi des observations  $X_1, \dots, X_T$ ?

### 1. Estimation par histogramme.

Une première solution est de reproduire la représentation en escalier de la fonction de répartition empirique  $F_n$  pour  $f$

$$\hat{f}(x) = \sum_{i=1}^k \omega_i \mathbf{1}_{[a_i, a_{i+1}[}(x), a_1 < \dots < a_{k+1}. \quad (16)$$

## 6. Estimation de la densité de probabilité

En choisissant les  $(\omega_i, a_i)$  telles que

$$\sum_{i=1}^k \omega_i(a_{i+1} - a_i) = 1 \text{ et } \omega_i(a_{i+1} - a_i) = \hat{P}_F(X \in [a_i, a_{i+1}[).$$

Par exemple,

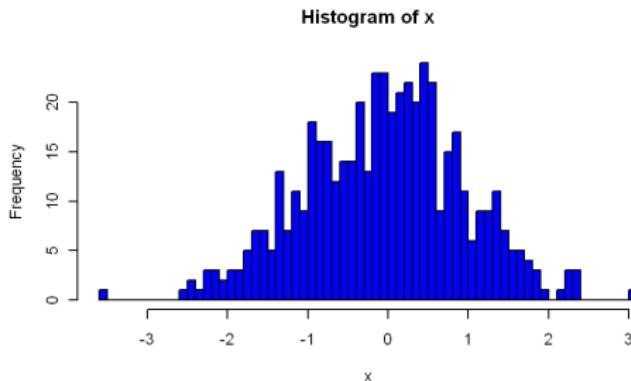
$$\omega_i(a_{i+1} - a_i) = \frac{1}{T} \sum_{j=1}^T 1_{[a_i, a_{i+1}[}(X_j)$$

est un estimateur convergent de  $P_F(X \in [a_i, a_{i+1}[)$ .

## 6. Estimation de la densité de probabilité

### Commandes sur R

```
x <- rnorm(500), hist(x,breaks=50,col="blue")
```



`hist(x)$density` donne les valeurs des  $\omega_i$  et `hist(x)$breaks` les valeurs des  $a_i$ .

## 6. Estimation de la densité de probabilité

Inconvénients:

- L'estimateur dépend du choix de la partition ( $a_i$ ), souvent construite en fonction des données (comme dans R)
- Problème des extrémités  $a_1$  et  $a_{k+1}$  : elles ne peuvent pas être infinis mais doivent suffisamment approcher le support de  $f$ .

## 6. Estimation de la densité de probabilité

- $k$  et  $(a_i)$  doivent dépendre de  $T$  pour que  $\hat{f}$  converge vers  $f$  mais  $a_{i+1} - a_i$  ne doit pas décroître trop vite vers 0 pour que l'estimation soit convergente : il faut suffisamment d'observations par intervalle  $[a_i, a_{i+1}[$ .
- L'histogramme est une fonction discontinue.

## 6. Estimation de la densité de probabilité

### 2. Estimateur à noyau

Au lieu de considérer une approximation uniforme autour de chaque  $X_i$ , on peut utiliser une fonction plus lisse :

$$\hat{f}(x) = \frac{1}{Th} \sum_{i=1}^T K\left(\frac{x - X_i}{h}\right), \quad (17)$$

où  $K$  est un noyau (par exemple une densité de probabilité) et  $h$  un facteur d'échelle.

## 6. Estimation de la densité de probabilité

Sur R on peut utiliser les noyaux suivants

- Le noyau normal [kernel="gaussian" ou "g"]
- Le noyau d'Epanechnikov [kernel="epanechnikov" ou "e"]

$$K(y) = C (1 - y^2)^2 \mathbf{1}_{[-1,1]}(y)$$

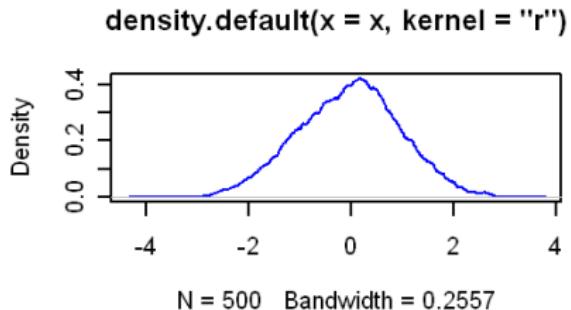
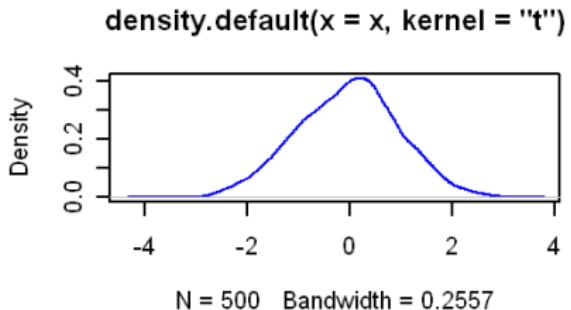
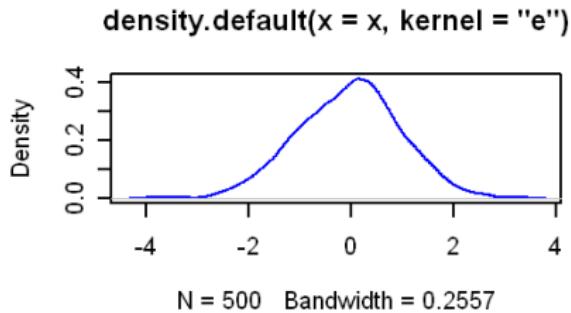
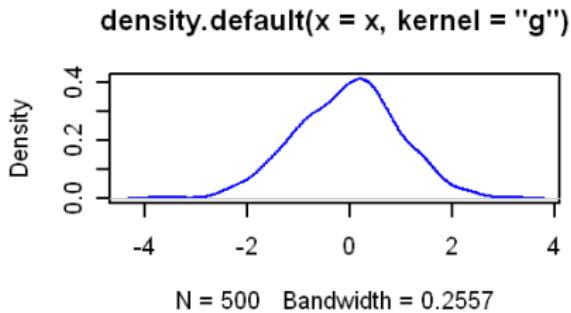
- Le noyau triangulaire [kernel="triangular" ou "t"]

$$K(y) = (1 + y)/\mathbf{1}_{[-1,0]}(y) + (1 - y)\mathbf{1}_{[0,1]}(y).$$

- Les noyaux "rectangular", "biweight", "cosine", "optcosine".

L'estimation est influencée par le choix du noyau.

## 6. Estimation de la densité de probabilité

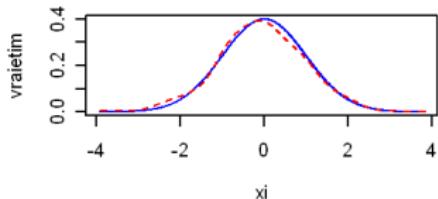
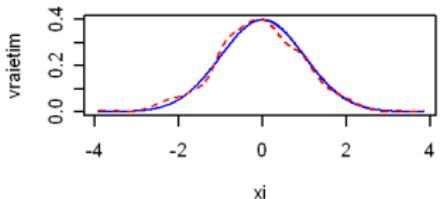
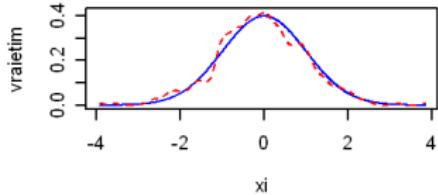
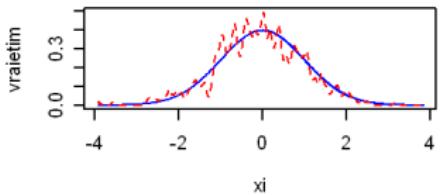


## 6. Estimation de la densité de probabilité

Le choix de la fenêtre  $h$  est crucial,

- Si  $h$  est grande : un grand nombre des  $X_i$  contribuent à l'estimation de  $f(x)$ , on obtient un estimateur avec un biais très grand et une variance très petite.
- Si  $h$  est petite : peu de  $X_i$  contribuent à l'estimation de  $f(x)$ , on obtient un estimateur avec un biais très petit et une variance très grande

## 6. Estimation de la densité de probabilité



Le choix  $h_T = T^{-1/5}$  (en bas à droite) semble être meilleur.

## 6. Estimation de la densité de probabilité

### 3. Fenêtre optimale. En étudiant l'erreur moyenne intégrée

$$d(f, \hat{f}) = E \int (f(x) - \hat{f}(x))^2 dx,$$

on peut trouver un choix optimal pour la fenêtre h.

De la décomposition

$$d(f, \hat{f}) = \int (f(x) - Ef(\hat{x}))^2 dx + \int \text{var}\hat{f}(x)dx = (\text{biais}^2 + \text{var})$$

et les approximations

$$f(x) - Ef(\hat{x}) \simeq \frac{f''(x)}{2} h_T^2$$

## 6. Estimation de la densité de probabilité

$$E \left[ \exp \left( -\frac{(X_i - x)^2}{2h_T^2} \right) \right] \simeq f(x)\sqrt{2\pi}h_T,$$

on en déduit que le carré du biais est de l'ordre de

$$\left( \frac{f''(x)}{2} \right)^2 h_T^4,$$

et que le terme de variance est approximativement  $\frac{1}{Th_T\sqrt{2\pi}}$ .

Par conséquent, l'erreur moyenne intégrée tend vers 0 quand  $T$  tend vers l'infini si

$$h_T \rightarrow 0 \text{ et } Th_T \rightarrow \infty$$

## 6. Estimation de la densité de probabilité

Une fenêtre optimale est donnée par

$$h_{opt} = \left( T \sqrt{2\pi} (f''(x))^2 \right)^{-1/5}.$$

$h_{opt}$  dépend de la dérivée seconde qui elle-même inconnue. Pour résoudre ce problème il existe plusieurs méthodes. La fenêtre optimale basée sur "rule of thumb", avec le noyau de Gauss, a la forme

$$h_{opt} = \frac{0.9 \min(\hat{\sigma}, \hat{q}_{75} - \hat{q}_{25})}{1.34 T^{1/5}},$$

où  $\sigma$  est l' écart-type estimé et  $\hat{q}_{25}$  et  $\hat{q}_{75}$  sont les quantiles à 25% et à 75% estimés (Silverman (1986, page 48, eqn (3.31))).

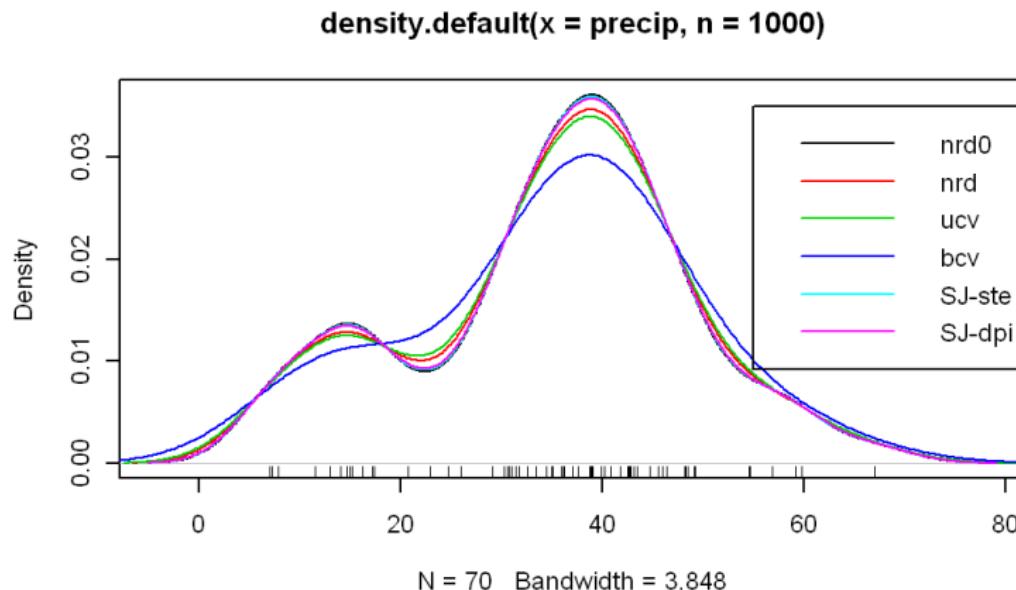
## 6. Estimation de la densité de probabilité

Le logiciel R contient les méthodes suivantes:

- `bw.nrd0` : implémente la "rule-of-thumb" de Silverman.
- `bw.nrd` : une variation de la précédente Scott (1992), utilisant 1.06 au lieu de 1.34.
- `bw.ucv` et `bw.bcv` utilisent la validation croisée non biaisée et biaisée respectivement.
- `bw.SJ` : implémente la méthode de Sheather & Jones (1991).

## 6. Estimation de la densité de probabilité

On considère la série des précipitations annuelles "precip".



## 7. Régression non paramétrique

- Le modèle (1) de régression exprime la variable  $y_t$  comme une combinaison linéaire en fonction des variables explicatives  $x_{tj}$ .
- Ce modèle est incapable de décrire une relation non linéaire entre  $y_t$  et  $x_{tj}$  si celle-ci existe.
- L'estimation non paramétrique offre l'avantage d'être plus flexible: ce sont les données qui déterminent la relation fonctionnelle entre  $y_t$  et  $x_{tj}$ .

## 7. Régression non paramétrique

D'une manière générale on cherche une fonction  $r$  telle que

$$y_t = r(x_{t,1}, \dots, x_{t,p}) + u_t, \quad 1 \leq t \leq T, \quad (18)$$

où  $u_t$  est un bruit blanc.

La fonction  $r$  peut être estimée par plusieurs méthodes:

- le réssogramme.
- la méthode des  $k$  voisins les plus proches.
- la méthode basée sur les splines.
- la méthode du noyau.

## 1. Estimateur à noyau

Posons  $x_t = (x_{t,1}, \dots, x_{t,p})$ .

On dispose d'un échantillon  $(x_t, y_t)_{1 \leq t \leq T}$  et on cherche à identifier la fonction  $r$  telle que

$$y_t = r(x_t) + u_t, \quad 1 \leq t \leq T, \quad (19)$$

## 7. Régression non paramétrique

- Watson (1964) et Nadaraya (1964) ont proposé, indépendamment et simultanément, l'estimateur

$$\hat{r}(x) = \begin{cases} \frac{\sum_{i=1}^T y_i K((x-x_i)/h_T)}{\sum_{i=1}^T K((x-x_i)/h_T)} & \text{si } \sum_{i=1}^T K((x-x_i)/h_T) \neq 0 \\ 0 & \text{sinon} \end{cases}$$

- Collomb (1976 ou 1977a) donne une évaluation asymptotique du biais et de la variance:  $\exists$  deux fonctions  $a(x)$  et  $b(x)$  telles que:

$$r(x) - E\hat{r}(x) \simeq \frac{h_T^2}{2} a(x),$$

$$E((r(x) - E\hat{r}(x))^2) \simeq \frac{1}{Th_T^p} b(x)$$

## 7. Régression non paramétrique

Ces formules permettent de montrer que

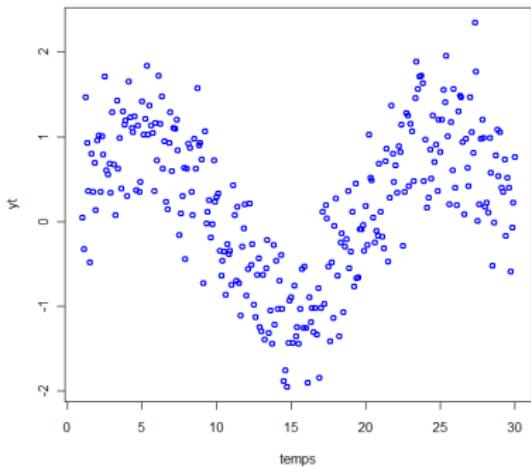
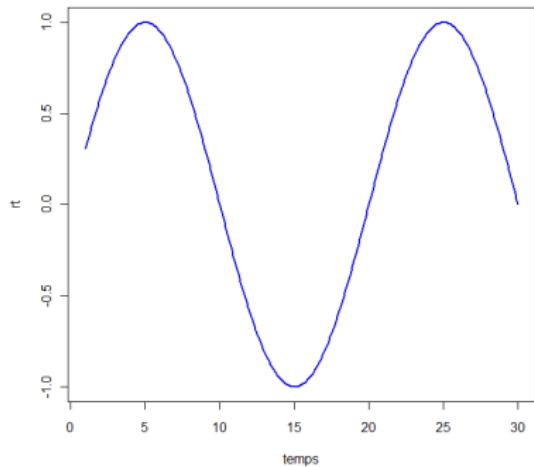
$$\min_{h_T \in \mathbb{R}^+} E (\hat{r}(x) - r(x)^2) \simeq c T^{-4/(p+4)}.$$

- Comme pour l'estimation de la densité le choix de la fenêtre  $h_T$  est crucial.
- Sur R la fonction `glkerns` permet l'estimation de la fonction  $r$  et ses dérivées avec un choix adaptative pour la fenêtre  $h_T$ .

## 7. Régression non paramétrique

### 2. Exemple. Simulation et estimation d'une fonction nonlinéaire.

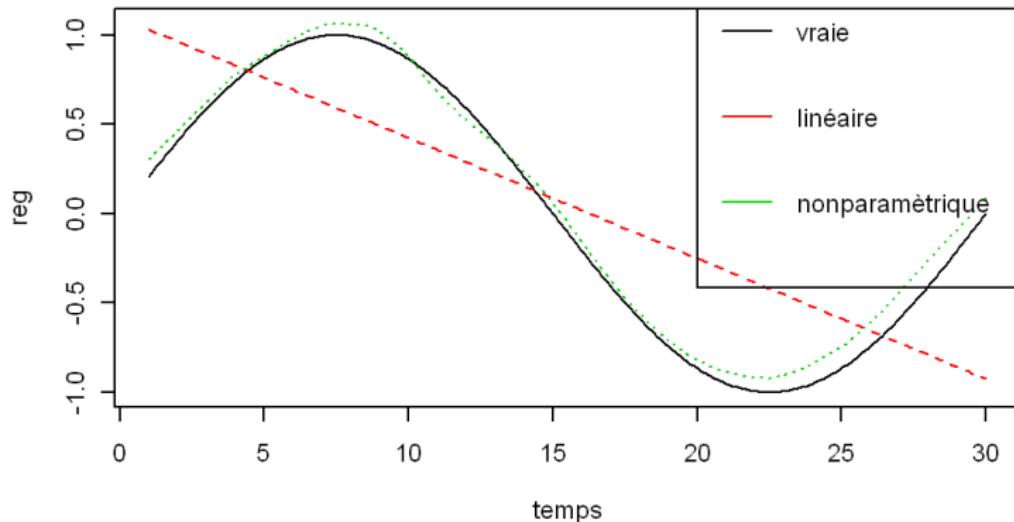
$$y_t = \sin(\pi t/10) + u_t, t = 1, 1.1, \dots, 30, u_t \sim N(0, 0.5)$$



## 7. Régression non paramétrique

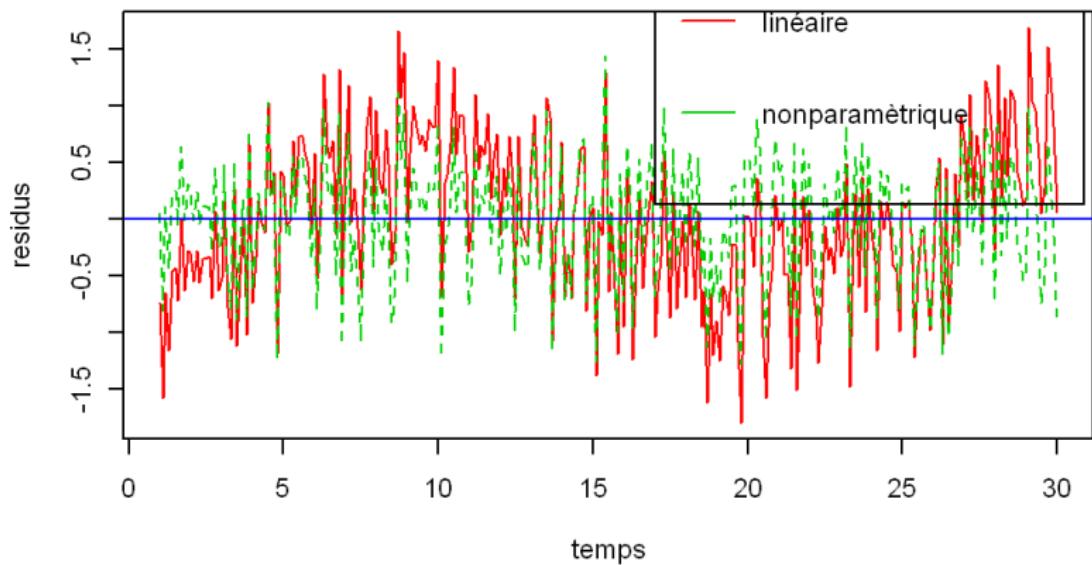
On identifie l'évolution de  $y_t$  par deux modèles:

1. Linéaire:  $y_t = b_1 + b_2 t + u_t$ .
2. Non paramétrique:  $y_t = r(t) + u_t$ .



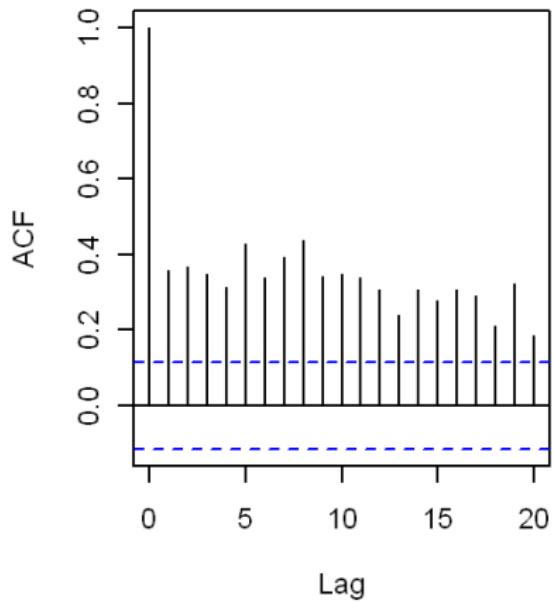
## 7. Régression non paramétrique

Analyse des résidus:

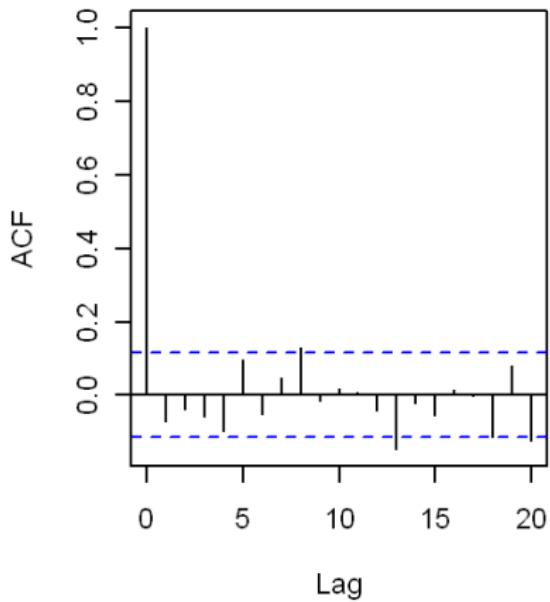


## 7. Régression non paramétrique

**Series reslineaire**



**Series resnparametrique**



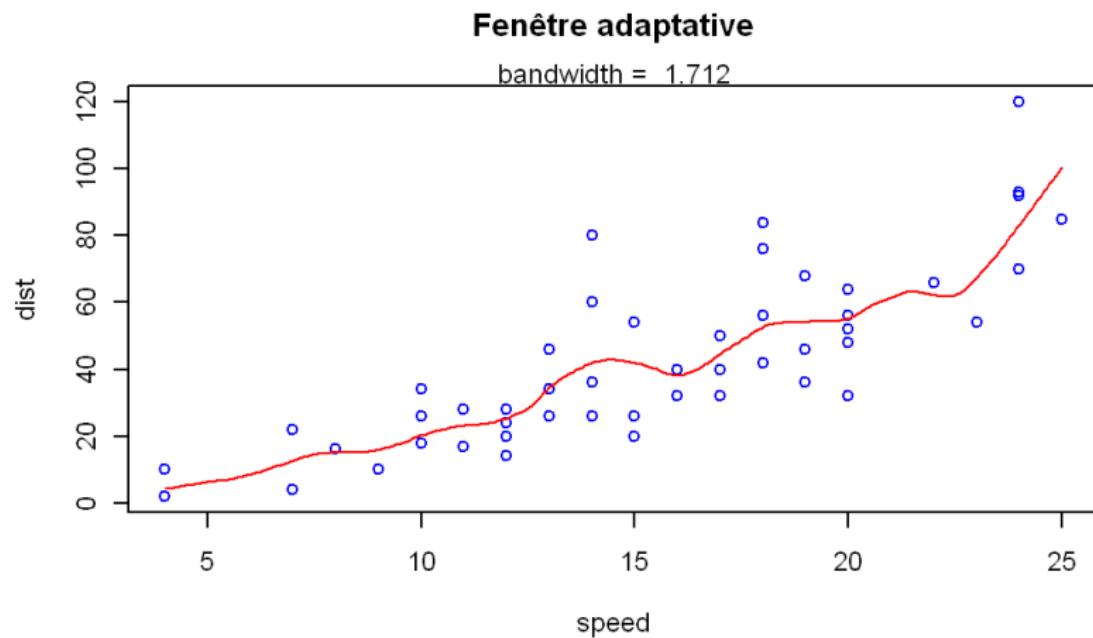
## 7. Régression non paramétrique

2. Exemple. On reprend les données sur la distance de freinage:

### Commandes sur R

```
library(lokern)
data(cars)
plot(cars, main = "Fenetre adaptative")
fit13 <- glkerns(cars$speed, cars$dist)
lines(fit13$x.out, fit13$est, col=2)
```

## 7. Régression non paramétrique



## 7. Régression non paramétrique

### Comparaison paramétrique-non paramétrique

