

RÉPUBLIQUE DÉMOCRATIQUE ALGÉRIENNE ET POPULAIRE

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

September 27, 2021



Université Mouloud Mammeri
Tizi Ouzou
Faculté des Sciences
Département de Mathématiques



MEMOIRE

THÈME

ESTIMATION ROBUSTE DANS LES PROCESSUS AR

Réalisé par Mr DIALLO Mamadou en vue de l'obtention du diplôme de
Master en Mathématiques

Domaine : Mathématiques et Informatique
Option : Probabilités et Statistiques

TABLE DES MATIÈRES

1) Généralités sur les processus stochastiques

Définitions

Processus Gaussien

Stationnarité stricte (forte)

Stationnarité faible (large)

Bruit Blanc Fort

Bruit Blanc Faible

2) Présentation des processus autorégressifs

Définition 1 (Processus AR(p))

Définition 2 (Équation polynômiale associée aux AR(p))

Théorème 1

Théorème 2

Théorème 3

3) Différentes méthodes d'estimation dans les modèles AR(p)

3.1) Construction des estimateurs

3.1.1) Cas particulier:Processus AR(1)

3.1.2) Estimateurs des moindres carrés

3.1.3) Estimateurs du maximum de vraisemblance

3.2) Cas général:Processus AR(p)

La fonction d'autocovariance

La fonction d'autocorrélation

La fonction d'autocorrélation partielle

3.2.1) Estimateurs des moindres carrés

3.2.2) Estimateurs du maximum de vraisemblance

3.2.3) La méthode de Yule-Walker

4) Premiers pas vers l'estimation robuste

5) Estimation robuste et Régression linéaire

5.1) Quelques méthodes d'estimation robuste

5.1.1) L-estimation

5.1.2) M-estimation

5.2) Application sur les données de ventes

6) Mise en oeuvre pratique avec R

6.1) Étude de la concentration de CO_2

6.2) Programme R

REMERCIEMENTS

Je remercie **DIEU** , le tout puissant de m'avoir donné la patience et le courage durant ces 5 longues années d'études.

Je souhaite adresser mes remerciements les plus sincères aux personnes qui m'ont apporté leur aide et qui ont contribué à l'élaboration de ce mémoire ainsi qu'à la réussite de cette année universitaire.

Ces remerciements vont tout d'abord aux corps **professoral** et **administratif** du Département de Mathématiques de la faculté des Sciences de l'Université Mouloud Mammeri de Tizi Ouzou , pour la richesse et la qualité de leur enseignement durant mon parcours en tant qu'étudiant , vous qui déployez de grands efforts pour assurer une formation actualisée à vos apprenants , veuillez recevoir mes sincères remerciements.

Je tiens également à remercier **Mr MAMOU Mohamed** en tant qu'encadreur de ce thème pour sa disponibilité et son aide précieuse tout au long de la réalisation de ce travail , ainsi que **Mr FELLAG Hocine** pour ses conseils pratiques.

Mes remerciements s'adressent aussi à mes **parents** pour leur soutien sans faille qui a contribué à m'encourager pour réaliser ce mémoire de fin d'études.

Enfin , je remercie tous mes proches et amis qui ont apporté leur aide dans la conception de ce mémoire.

Merci à tous et à toutes !!!

Abstract

The theme of this thesis is on robust estimation in autoregressive processes ie **approach with rigor** the parameters of these models. Before embarking on estimation , we will first recall some basic concepts in stochastic processes of order 1 then generalize these results to order p We will use somme estimation methods such as ordinary least squares (OLS) method , maximum likelihood (MV) method , Yule-Walker equations as well as some robust technicals such as L-estimation and M-estimate.

Subsequently , we will deal with an example on the (OLS) method based on linear regression followed by robust estimation when certain standard estimation assumptions are no longer valid and finally we will use real data to model the evolution of the concentration of carbon dioxide in the air and **estimate the autoregressive process in line with these data.**

The interest of this study is to use all the knowledge acquired in the modeling of chronological series when we have raw on which the classical estimation methods are not efficient.

The real issue of this theme is therefore stochastic modeling based on the application of robust estimation technicals.

NB:

We will start by theoretically looking for the estimators , then the R software will give us the true values for the practical cases.

**Without loss generalities , the terms (MCO) and (MV) will indicate respectively :
(Ordinary Least Squares) and (Likelihood Maximum)**

Keywords:

Stochastic processes , Time series , Chronic , Autoregressive processes ,
Linear regression , Robust estimation , L-estimation , M-estimation ,
Robust regression.

Résumé

Le thème de ce mémoire porte sur l'estimation robuste dans les processus autorégressifs ie **approcher avec rigueur** les paramètres de ces modèles. Avant de se lancer dans l'estimation , nous allons d'abord rappeler quelques concepts de base dans les processus stochastiques en général. Après cela , nous allons commencer par l'estimation paramétrique des processus autorégressifs d'ordre 1 puis généraliser ces résultats à l'ordre p Nous utiliserons quelques méthodes d'estimation usuelles telles que la méthode des moindres carrés ordinaires (MCO) , la méthode du maximum de vraisemblance (MV) , les équations de Yule-Walker ainsi que certaines techniques robustes telles que la L-estimation et la M-estimation.

Par la suite , nous traiterons un exemple sur la méthode des (MCO) basée sur la régression linéaire suivi de l'estimation robuste lorsque certaines hypothèses standards d'estimation ne sont plus valables et enfin nous utiliserons des données réelles pour modéliser l'évolution de la concentration de dioxyde de carbone contenu dans l'air afin d'**estimer le processus autorégressif en adéquation avec ces données**.

L'intérêt de cette étude est de mettre en oeuvre toutes les connaissances acquises dans la modélisation de séries chronologiques quand nous disposons de données brutes sur lesquelles les méthodes classiques d'estimation ne sont pas efficaces.

Le véritable enjeu de ce thème est donc la modélisation stochastique basée sur l'application de techniques robustes d'estimation.

NB:

Nous commencerons par chercher théoriquement les estimateurs, puis le logiciel R nous donnera les vraies valeurs pour les cas pratiques.

**Sans perte de généralités , les termes (MCO) et (MV) désigneront respectivement :
(Moindres Carrés Ordinaires) et (Maximum de Vraisemblance)**

Mots Clés

Processus stochastiques , Séries chronologiques , Chronique , Processus autorégressifs , Régression linéaire , Estimation robuste , L-estimation , M-estimation , Régression robuste.

INTRODUCTION GÉNÉRALE

Pour essayer de comprendre le monde autour de nous , des observations sont souvent faites étalées dans le temps . Les valeurs futures dépendent des observations disponibles actuellement (passé et présent) , généralement de façon stochastique ie aléatoirement.

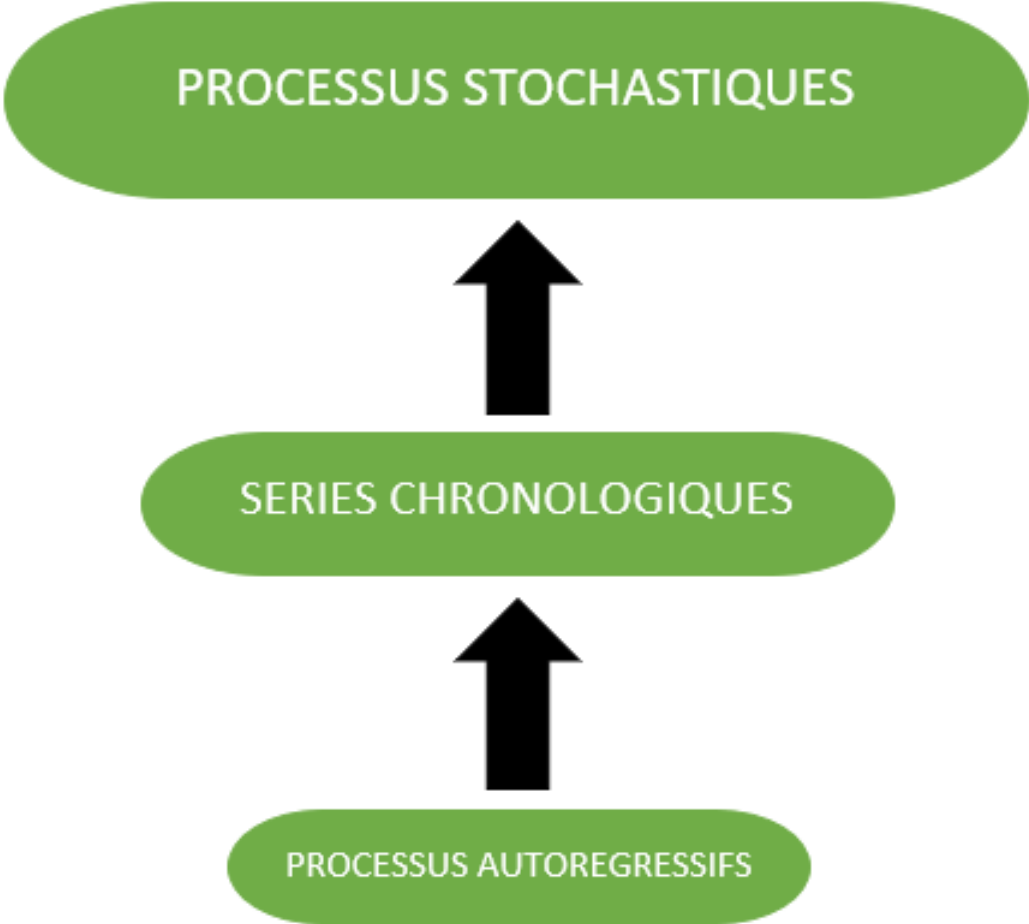
Une telle dépendance est utile pour prévoir le futur en fonction du passé ; les objectifs de l'analyse de séries temporelles sont divers selon le contexte des applications . Pour les statisticiens , une série temporelle est une réalisation d'un processus stochastique et ont pour défi de dévoiler la loi de probabilité de la série pour mieux la contrôler .

Cependant , il existe une multitude de processus stochastiques et la classe la plus populaire concerne les processus linéaires qui se composent des modèles autorégressifs dont nous parlerons principalement dans la suite de cette étude. Ces processus autorégressifs supposent que chaque point peut être prédit par la somme pondérée d'un ensemble de points précédents , plus un terme d'erreur aléatoire , ils sont utiliser pour modéliser les séries chronologiques dans de nombreux domaines comme la météorologie , la finance , etc ...

Par exemple , en télécommunication , le signal des réseaux téléphoniques ; en météorologie , les prévisions climatologiques en fonction des températures observées précédemment ; en finance , la modélisation du taux d'échange d'une devise , etc ...

Les statisticiens ont toujours travaillé sur l'estimation des paramètres de ces processus et des recherches sont toujours en cours pour répondre aux problèmes pratiques et faire des prévisions.

Pour aider les lecteurs à bien se situer ici , voici un organigramme hiérarchique dans la classification des différentes notions mathématiques qui nous intéressent:

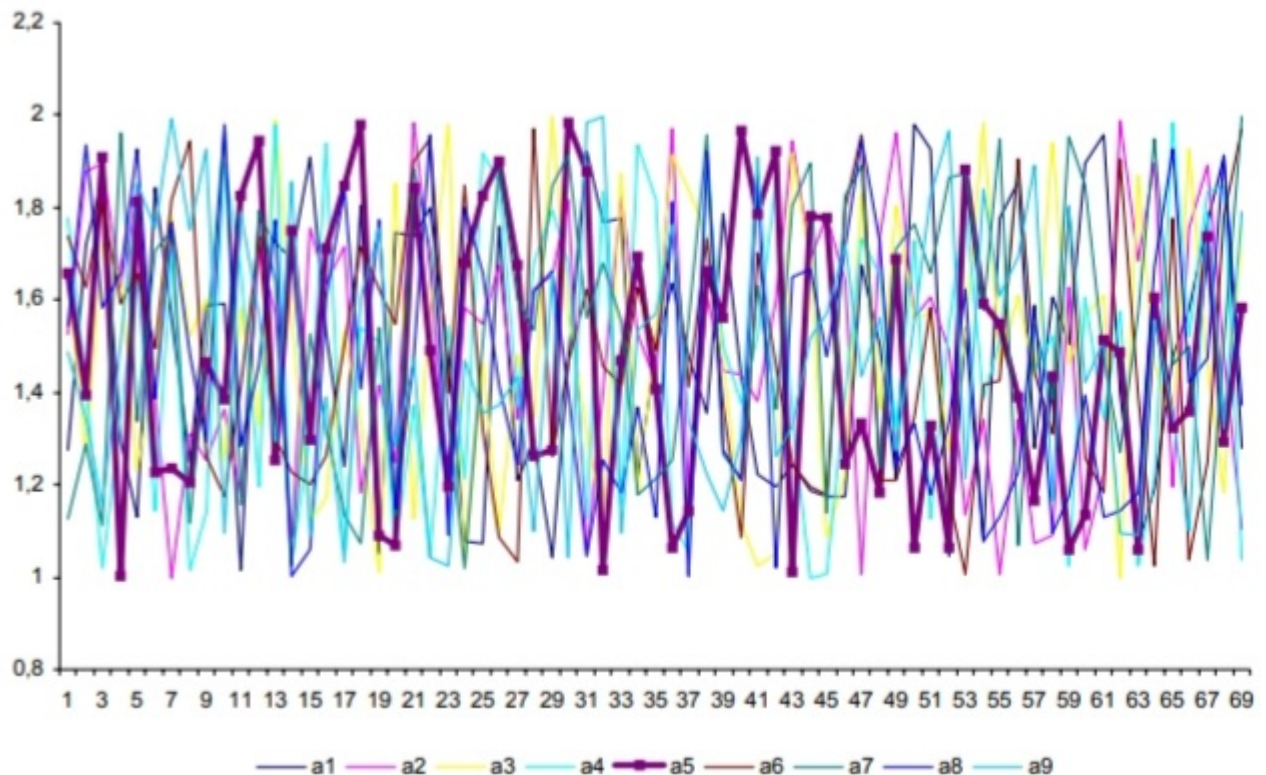


1 Généralités sur les processus stochastiques

Définitions

- Un processus stochastique est un modèle de probabilité permettant d'étudier un phénomène aléatoire au cours du temps.
- Mathématiquement, un processus stochastique noté $X_T = (X_t, t \in T)$ est une famille de v.a définies sur un même espace probabilisé (Ω, A, P) à valeurs dans (R, B_R) où A est une tribu, P une probabilité et B_R la tribu borélienne de R .

Exemple de processus



Remarques

Pour ω fixé, l'application $t \mapsto X_t(\omega)$ est la réalisation du processus au point ω appelée trajectoire du processus.
Le processus $(X_t, t \in T)$ peut être considéré comme une v.a.

$$X_T : (\Omega, A, P) \rightarrow (R^T, B_{R^T})$$

$$\omega \mapsto X_t(\omega), t \in T$$

B_{R^T} est la plus petite tribu contenant tous les cylindres $C_n \times R^{T-T_n}$ avec T_n une partie finie quelconque de T à n éléments et C_n une partie borélienne du produit fini R^{T_n}

Remarques

- La tribu B_{R^T} est aussi la tribu engendrée par les projections canoniques :

$$P_t : (R^T, B_{R^T}) \rightarrow (R, B_R)$$

$$(X_s, s \in T) \rightarrow X_t$$

- Le processus $X_T = (X_t, t \in T)$ a la même loi que le processus canonique P_t et est donnée par :

$$P_{X_T}(s) = P(X_T^{-1}(s)) \quad \text{avec } s \in B_{R^T}$$

- Les lois de dimension finie du processus X_T sont les lois des vecteurs aléatoires $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$, $k \geq 1$ et $t_1, t_2, \dots, t_k \in T$
- La loi du processus $X_T = (X_t, t \in T)$ est entièrement déterminée par les lois de dimension finie du processus (Voir Théorème de Kolmogorov) .

Processus Gaussien

Le processus $X_T = (X_t, t \in T)$ est dit gaussien , si ses lois de dimension finie sont gaussiennes , ie si toute combinaison linéaire finie de $X_t, t \in T$ est une v.a gaussienne.

La loi d'un tel processus est entièrement déterminée par la donnée des fonctions moyenne et covariance:

$$E(X_t) = m(t), t \in T$$

$$Cov(X_s, X_t) = c(s, t), s, t \in T$$

Stationnarité

Il est clair que le concept de stationnarité est fondamental car nous pouvons tomber sur des processus non stationnaires , dans ce cas il faudra les

stationnariser.

Il existe deux types de stationnarité :

Stationnarité stricte (forte)

Un processus $(X_t, t \in Z)$ est dit strictement stationnaire si pour toute partie finie t_1, \dots, t_n de T et pour tout $s > 0$, $L(X_{t_1+s}, \dots, X_{t_n+s}) = L(X_{t_1}, \dots, X_{t_n})$

Stationnarité faible (large)

Un processus $(X_t, t \in Z)$ est dit faiblement stationnaire si tous ses moments d'ordre 2 ($E(X_t^2)$) existent et sa covariance dans Z^2 définie par : $Cov(X_s, X_t) = E(X_s X_t) - E(X_s) E(X_t)$ ne dépend que de la différence $s - t$.

Remarques

- Pour les processus gaussiens, les deux notions sont équivalentes.
- La stationnarité forte implique la stationnarité faible, la réciproque est généralement fausse.

Processus Bruit Blanc (White Noise)

Il existe également deux types de processus bruit blanc :

Bruit Blanc Fort

On dit que le processus $(\epsilon_t, t \in Z)$ est un bruit blanc fort si les v.a $\epsilon_t, t \in Z$ sont centrées, iid, et de variance $\sigma^2 > 0$

Bruit Blanc Faible

$(\epsilon_t, t \in Z)$ est un bruit blanc faible si :

$$E(\epsilon_t) = 0 \quad \text{et} \quad E(\epsilon_t \epsilon_s) = \delta_{st} \sigma^2$$

δ_{st} est le symbole de Kronecker et $\sigma^2 > 0$ ie si :

$$E(\epsilon_t) = 0$$

$$E(\epsilon_t \epsilon_s) = \begin{cases} \sigma^2 & \text{si } s = t \\ 0 & \text{si } s \neq t \end{cases}$$

2 Présentation des processus autorégressifs

Dans toute la suite , nous supposons que le processus $(X_t, t \in Z)$ est à temps discret à valeurs réelles.

Définition 1

Le processus $(X_t, t \in Z)$ satisfait une représentation autorégressive d'ordre p si il vérifie pour tout $p \geq 1$:

$$\mathbf{X}_t = \alpha_0 + \alpha_1 \mathbf{X}_{t-1} + \alpha_2 \mathbf{X}_{t-2} + \dots + \alpha_p \mathbf{X}_{t-p} + \epsilon_t, \mathbf{t} \in \mathbf{Z}, \alpha_p \neq 0 \quad (1.1)$$

Avec $\alpha_0, \alpha_1, \dots, \alpha_p$ des réels , les v.a $\epsilon_t, t \in Z$ sont un bruit blanc faible telles que $E(\epsilon_t X_s) = 0$ pour tout $s < t$

On notera un autorégressif d'ordre p par $AR(p)$

Définition 2

On appelle équation polynômiale associée à un $AR(p)$, l'équation :

$$\Delta(z) = z^p - \sum_{i=1}^p \alpha_i z^{p-i}$$

Théorème 1

Un processus autorégressif vérifiant l'équation (1.1) est faiblement stationnaire si les racines de son équation polynômiale associée $\Delta(z) = 0$ sont en module strictement inférieur à 1

Théorème 2

Si toutes les conditions du théorème 1 sont vérifiées alors le processus X_T peut s'écrire sous la forme :

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \delta_j \epsilon_{t-j}, \mathbf{t} \in \mathbf{Z}$$

Avec $\delta_0 = 1$, la convergence est en moyenne quadratique et δ_j sont les coefficients du développement en série de Taylor de :

$$\frac{1}{1 - \sum_{i=1}^p \alpha_i z^i} = \sum_{j=0}^{\infty} \delta_j z^j$$

X_t non corrélée avec $\epsilon_{t+1}, \epsilon_{t+2}, \dots$

Théorème 3

Soit $(X_t, t \in Z)$ un processus $AR(p)$ faiblement stationnaire tels que les racines de $\Delta(z) = 0$ soient de module strictement inférieur à 1 et $\epsilon_t, t \in z$ soient des v.a gaussiennes iid de loi $N(0, \sigma_\epsilon^2)$, alors le processus $(X_t, t \in Z)$ est gaussien.

Les lois de dimensions finie sont des lois normales de moyenne et de covariance respectives :

$$m(t) = \frac{\alpha_0}{1 - \alpha_1 - \alpha_2 - \dots - \alpha_p}$$

$$C(k) = \sum_{i=1}^p c_i x_i^k$$

avec c_i des constantes et x_i les racines distinctes de $\Delta(z) = 0$

3 Différentes méthodes d'estimation dans les modèles AR(p)

Estimer, c'est trouver une valeur approchée des paramètres observés dans le modèle, pour cela, il existe certaines méthodes que nous connaissons déjà tels que la méthode des moindres carrés ordinaires, la méthode du maximum de vraisemblance, la méthode par résolution des équations de Yule-Walker... Ces méthodes d'estimation constituent la base dans l'assimilation de la théorie de toute estimation.

3.1 Construction des estimateurs

Soit $(X_t, t \in Z)$ un processus $AR(p)$ gaussien et stationnaire vérifiant:

$$\mathbf{X}_t = \alpha_0 + \alpha_1 \mathbf{X}_{t-1} + \alpha_2 \mathbf{X}_{t-2} + \dots + \alpha_p \mathbf{X}_{t-p} + \epsilon_t$$

Les paramètres inconnus à estimer sont : $\alpha_1, \alpha_2, \dots, \alpha_p$ et σ_ϵ^2
 $\epsilon_t, t \in Z$ constituent une suite de v.a iid de loi $N(0, \sigma_\epsilon^2)$.

On dispose de X_1, \dots, X_n observations tels que $X_0 = X_{-1} = \dots = X_{-(p-1)} = 0$

3.1.1 Cas particulier : Processus AR(1)

Soit $X_t = \alpha_0 + \alpha X_{t-1} + \epsilon_t$ un AR(1) avec les hypothèses suivantes :

- $|\alpha| < 1$
- $V(\epsilon_t) < \infty$
- $X_0 = X_{-1} = 0$

3.1.2 Estimateurs des moindres carrés ordinaires

Les premiers estimateurs sont obtenus par régression de X_t sur X_{t-1}
 $t = 1, \dots, n$.

La méthode des moindres carrés ordinaires dans le cas de la régression linéaire simple est un problème d'optimisation sans contraintes qui consiste à minimiser la somme $\epsilon_t, t = 1, \dots, n$

Ecriture matricielle

$$\mathbf{Y} = \alpha_0 \mathbf{1} + \alpha_1 \mathbf{X} + \epsilon$$

$Y = \{y_1, \dots, y_n\}^t$ est aléatoire de dimension n

$\mathbf{1} = \{1, \dots, 1\}^t$ les vecteur unitaire de R^n

$X = \{x_1, \dots, x_n\}^t$ non aléatoire de dimension n

α_0 et α_1 les paramètres inconnus non aléatoires

$\epsilon = \{\epsilon_1, \dots, \epsilon_n\}^t$ aléatoire de dimension n

Preuve

$$\text{Soit } S(\alpha_0, \alpha_1) = \sum_t (y_t - \alpha_0 - \alpha_1 x_t)^2$$

Minimiser la quantité $S(\alpha_0, \alpha_1) = \sum_t \epsilon_t^2$ revient à chercher :

$$\begin{aligned} \alpha_{*,0} &= \operatorname{argmin}_{\alpha_0} \sum_t (y_t - \alpha_0 - \alpha_1 x_t)^2 \\ \alpha_{*,1} &= \operatorname{argmin}_{\alpha_1} \sum_t (y_t - \alpha_0 - \alpha_1 x_t)^2 \end{aligned}$$

Ce qui nous pousse à chercher les solutions du système suivant :

$$\frac{\partial \mathbf{S}}{\partial \alpha_0} = -2 \sum_t (\mathbf{y}_t - \alpha_{*,0} - \alpha_{*,1} \mathbf{x}_t) = \mathbf{0}$$

$$\frac{\partial \mathbf{S}}{\partial \alpha_1} = -2 \sum_t \mathbf{x}_t (\mathbf{y}_t - \alpha_{*,0} - \alpha_{*,1} \mathbf{x}_t) = \mathbf{0}$$

La première équation donne :

$$\alpha_{*,0}n + \alpha_{*,1} \sum_t x_t = \sum_t y_t \quad (1.2)$$

En tirant $\alpha_{*,0}$ de cette dernière on aura:

$$\alpha_{*,0} = \bar{y} - \alpha_{*,1}\bar{x} \quad (1.3)$$

\bar{y} et \bar{x} représentent les moyennes empiriques respectives de y_t et x_t
En résolvant le système formé par les équations (1.2) et (1.3) on obtient :

$$\bar{y} - \alpha_{*,1}\bar{x} = \frac{\sum_t y_t - \alpha_{*,1} \sum_t x_t}{n}$$

D'où :

$$\alpha_{*,1} = \frac{\sum_t (\mathbf{x}_t - \bar{x})(\mathbf{y}_t - \bar{y})}{\sum_t (\mathbf{x}_t - \bar{x})(\mathbf{x}_t - \bar{x})}$$

Les estimateurs des (MCO) d'un processus AR(1) sont donc :

$$\alpha_{*,0} = \bar{y} - \alpha_{*,1}\bar{x} \quad (1.4)$$

$$\alpha_{*,1} = \frac{\sum_t (\mathbf{x}_t - \bar{x})(\mathbf{y}_t - \bar{y})}{\sum_t (\mathbf{x}_t - \bar{x})^2} \quad (1.5)$$

Propriétés 1

Les estimateurs $\alpha_{*,0}$ et $\alpha_{*,1}$ sont sans biais respectivement pour α_0 et α_1

$$\mathbf{E}(\alpha_{*,0}) = \alpha_0$$

$$\mathbf{E}(\alpha_{*,1}) = \alpha_1$$

Preuve

En remplaçant $(y_t - \bar{y})$ par $\alpha_1(x_t - \bar{x}) + \epsilon_t$ dans (1.5) , on aura :

$$\alpha_{*,1} = \alpha_1 + \frac{\sum_t (x_t - \bar{x}) \epsilon_t}{\sum_t (x_t - \bar{x})^2} \quad (1.6)$$

$$E(\alpha_{*,1}) = \alpha_1 + \frac{\sum_t (x_t - \bar{x}) E(\epsilon_t)}{\sum_t (x_t - \bar{x})^2}$$

D'où :

$$\mathbf{E}(\alpha_{*,1}) = \alpha_1$$

Par ailleurs :

$$\alpha_{*,0} = \bar{y} - \alpha_{*,1}\bar{x} = \alpha_0 + \alpha_1\bar{x} - \alpha_{*,1}\bar{x} = \alpha_0 + (\alpha_1 - \alpha_{*,1})\bar{x} \quad (1.7)$$

$$E(\alpha_{*,0}) = \alpha_0 + E[(\alpha_1 - \alpha_{*,1})\bar{x}] = \alpha_0$$

Nous déduisons donc que :

$$\mathbf{E}(\alpha_{*,0}) = \alpha_0$$

Propriétés 2

$$\mathbf{V}(\alpha_{*,0}) = \frac{(\sigma_\epsilon \sum_t \mathbf{x}_t)^2}{\mathbf{n} \sum_t (\mathbf{x}_t - \bar{\mathbf{x}})^2}$$

$$\mathbf{V}(\alpha_{*,1}) = \frac{\sigma_\epsilon^2}{\sum_t (\mathbf{x}_t - \bar{\mathbf{x}})^2}$$

$$\mathbf{Cov}(\alpha_{*,0}, \alpha_{*,1}) = -\frac{\sigma_\epsilon^2 \bar{\mathbf{x}}}{\sum_t (\mathbf{x}_t - \bar{\mathbf{x}})^2}$$

Preuve

Ces propriétés se démontrent facilement en utilisant les équations (1.6) et (1.7) :

$$V(\alpha_{*,1}) = V\left[\alpha_1 + \frac{\sum_t (x_t - \bar{x}) \epsilon_t}{\sum_t (x_t - \bar{x})^2}\right] = \frac{\sum_t (x_t - \bar{x})^2 V(\epsilon_t)}{[\sum_t (x_t - \bar{x})^2]^2}$$

D'où le résultat :

$$\mathbf{V}(\alpha_{*,1}) = \frac{\sigma_\epsilon^2}{\sum_t (\mathbf{x}_t - \bar{\mathbf{x}})^2}$$

Pour $V(\alpha_{*,0})$ on sait que $Cov(\bar{y}, \alpha_{*,1}) = 0$ ie que \bar{y} et $\alpha_{*,1}$ ne sont pas corrélés , alors :

$$V(\alpha_{*,0}) = V(\bar{y}) + \bar{x}^2 V(\alpha_{*,1}) = \frac{\sigma_\epsilon^2}{n} + \frac{(\sigma_\epsilon \sum_t x_t)^2}{[n \sum_t (x_t - \bar{x})]^2}$$

En cherchant un dénominateur commun , on obtient :

$$\mathbf{V}(\alpha_{*,0}) = \frac{(\sigma_\epsilon \sum_t \mathbf{x}_t)^2}{\mathbf{n} \sum_t (\mathbf{x}_t - \bar{\mathbf{x}})^2}$$

Finalement , pour la covariance de $\alpha_{*,0}$ et $\alpha_{*,1}$:

$$Cov(\alpha_{*,0}, \alpha_{*,1}) = Cov(\bar{y} - \alpha_{*,1}\bar{x}, \alpha_{*,1}) = Cov(\bar{y}, \alpha_{*,1}) - \bar{x}V(\alpha_{*,1})$$

En remplaçant $Cov(\bar{y}, \alpha_{*,1})$ et $V(\alpha_{*,1})$ par leurs valeurs , on aura :

$$\mathbf{Cov}(\alpha_{*,0}, \alpha_{*,1}) = -\frac{\sigma_\epsilon^2 \bar{\mathbf{x}}}{\sum_t (\mathbf{x}_t - \bar{\mathbf{x}})^2}$$

3.1.3 Estimateurs du maximum de vraisemblance

Les seconds estimateurs sont obtenus en maximisant le log de la fonction de vraisemblance qu'on désignera par $l(\cdot)$

$$\epsilon = \mathbf{Y} - \alpha_0 - \alpha_1 \mathbf{X} \sim \mathbf{BB}(\mathbf{0}, \sigma_\epsilon^2)$$

$(\epsilon_t, t \in Z)$ est un bruit blanc gaussien $\Rightarrow (Y_t, t \in Z)$ est un processus gaussien
Soit l'équation de la fonction de vraisemblance sous l'hypothèse de la normalité des $\epsilon_t, t \in Z$:

$$\begin{aligned} \mathbf{L}(\alpha, \sigma_\epsilon^2) &= \prod_{t=1}^n \frac{1}{\sigma_\epsilon \sqrt{2\pi}} \exp\left(\frac{-\epsilon_t^2}{2\sigma_\epsilon^2}\right) \\ \mathbf{L}(\alpha, \sigma_\epsilon^2) &= \left(\frac{1}{\sigma \sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma_\epsilon^2} \sum_t (y_t - \alpha_0 - \alpha_1 \mathbf{x}_t)^2\right\} \end{aligned}$$

Maximiser $l(\alpha, \sigma_\epsilon^2)$ revient à minimiser la quantité :

$$Q = S(\alpha_0, \alpha_1)$$

Ce qui a déjà été fait , alors on conclut que les estimateurs du maximum de vraisemblance sont les même que ceux retrouvés par la méthode des moindres carrées et sont donnés par :

$$\alpha_{0,MV} = \bar{y} - \alpha_{*,1}\bar{x}$$

$$\alpha_{1,MV} = \frac{\sum_t (\mathbf{x}_t - \bar{\mathbf{x}}) (\mathbf{y}_t - \bar{y})}{\sum_t (\mathbf{x}_t - \bar{\mathbf{x}}) (\mathbf{x}_t - \bar{\mathbf{x}})}$$

Proposition 1

L'estimateur des moindres carrées α_* de α est consistant ie que :

$$\alpha_* \xrightarrow{p} \alpha$$

Nous avons aussi la normalité asymptotique :

$$\sqrt{\mathbf{n}}(\alpha_* - \alpha) \xrightarrow{\ell} \mathbf{N}(\mathbf{0}, \mathbf{1} - \alpha^2)$$

Remarque

Pour trouver l'estimateur du maximum de vraisemblance de σ_ϵ^2 , on procède de la même façon en considérant $\alpha_{0,MV}$ et $\alpha_{1,MV}$ comme constantes et on aura :

$$\sigma_{*,MV}^2 = \frac{1}{\mathbf{n}} \sum_t \epsilon_{*,t}^2 = \frac{1}{\mathbf{n}} \sum_t (\mathbf{y}_t - \alpha_{*,0} - \alpha_{*,1}\mathbf{x}_t)^2$$

$$\mathbf{E}(\sigma_{*,MV}^2) = \frac{\mathbf{n} - 2}{\mathbf{n}} \sigma_\epsilon^2$$

Alors $\sigma_{*,MV}^2$ est asymptotiquement sans biais donc c'est un estimateur biaisé de σ_ϵ^2 .

L'estimateur empirique sans biais de σ_ϵ^2 est égal à :

$$\sigma_*^2 = \frac{1}{\mathbf{n} - 2} \sum_t \epsilon_{*,t}^2 = \frac{1}{\mathbf{n} - 2} \sum_t (\mathbf{y}_t - \alpha_{*,0} - \alpha_{*,1}\mathbf{x}_t)^2$$

Nous déduisons de la normalité asymptotique que :

$$\frac{\sqrt{\mathbf{n}}(\alpha_{*,1} - \alpha_1)}{\sqrt{1 - \alpha_{*,1}^2}} \text{ converge en loi vers } \mathbf{N}(0, 1)$$

Proposition 2

$\alpha_{*,1}^2$ est un estimateur consistant de α_1^2 (Voir Lemme de Slutsky)
Cela entraîne qu'une région de confiance asymptotique de niveau $1 - \theta$ pour α_1 est donné par:

$$\left[\alpha_{*,1} - \frac{u_{1-\frac{\theta}{2}}}{\sqrt{\mathbf{n}}} \sqrt{1 - \alpha_{*,1}^2} \quad , \quad \alpha_{*,1} + \frac{u_{1-\frac{\theta}{2}}}{\sqrt{\mathbf{n}}} \sqrt{1 - \alpha_{*,1}^2} \right]$$

Avec $u_{1-\frac{\theta}{2}}$ le quantile d'ordre $1 - \frac{\theta}{2}$ de la loi $\mathbf{N}(0, 1)$

3.2 Cas général : Processus $\mathbf{AR}(p)$

Soit $X_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p} + \epsilon_t, t \in \mathbf{Z}, \alpha_p \neq 0$ un $\mathbf{AR}(p)$ telle que :

$$\mu = \mathbf{E}(\mathbf{X}_t) = \frac{\alpha_0}{1 - \alpha_1 - \alpha_2 - \dots - \alpha_p}$$

Un estimateur naturel de cette moyenne est la moyenne empirique :

$$\bar{\mathbf{X}}_n = \frac{1}{\mathbf{n}} \sum_t \mathbf{X}_t \quad , \quad t \in \{p+1, \dots, n\}$$

La fonction d'autocovariance

La fonction d'autocovariance d'un processus $(X_t, t \in \mathbf{Z})$ est définie par :

$$\gamma(\mathbf{h}) = \mathbf{Cov}(\mathbf{X}_t, \mathbf{X}_{t+\mathbf{h}})$$

La fonction d'autocorrélation

La fonction d'autocorrélation est donnée par :

$$\rho(\mathbf{h}) = \mathbf{Corr}(\mathbf{X}_t, \mathbf{X}_{t+\mathbf{h}}) = \frac{\mathbf{Cov}(\mathbf{X}_t, \mathbf{X}_{t+\mathbf{h}})}{\mathbf{V}(\mathbf{X}_t)} = \frac{\gamma(\mathbf{h})}{\gamma(\mathbf{0})}$$

L'autocorrélation partielle

L'autocorrélation (ou le coefficient d'autocorrélation) partielle d'ordre q de X_t est égale au coefficient de corrélation entre :

$$X_t - E(X_t/X_{t-1}, X_{t-2}, \dots, X_{t-q+1}) \quad \text{et} \quad X_{t-q} - E(X_{t-q}/X_{t-1}, X_{t-2}, \dots, X_{t-p+1})$$

3.2.1 Estimateurs des moindres carrés

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p} + \epsilon_t, t = 1, \dots, n, \alpha_p \neq 0$$

L'écriture sous forme matricielle de cette équation est donnée par :

$$\mathbf{Y} = \mathbf{X}\alpha + \epsilon$$

- $Y = (y_1, \dots, y_n)^t$ est une matrice de dimension $(n \times 1)$ ie les valeurs de la variable à expliquer.
- X est une matrice de dimension $n \times (p + 1)$, ie les valeurs des variables explicatives X_j

$$X = \begin{pmatrix} 1 & x_0 & \dots & x_{1-p} \\ & & \vdots & \\ & & \vdots & \\ & & \vdots & \\ 1 & x_{n-1} & \dots & x_{n-p} \end{pmatrix}$$

$\alpha = (\alpha_0, \dots, \alpha_p)^t$ est un vecteur de $p + 1$ paramètres inconnus.

$\epsilon = (\epsilon_1, \dots, \epsilon_n)^t$ est une matrice de dimension $n \times 1$ des erreurs du modèle de régression de variance inconnue et constante.

Pour trouver le meilleur ajustement linéaire possible, on estime les valeurs de α_j qui minimiseront les erreurs alors nous cherchons la valeur estimée du vecteur α :

$$\alpha_{*,n} = \mathbf{argmin}_{\alpha} \sum_t (y_t - \alpha_0 - \alpha_1 y_{t-1} - \dots - \alpha_p y_{t-p})^2$$

Soit $S(\alpha) = \sum_t (y_t - \alpha_0 - \alpha_1 y_{t-1} - \dots - \alpha_p y_{t-p})^2$

$$\begin{aligned}
\frac{\partial S(\alpha)}{\partial \alpha_0} &= -2 \sum_t (y_t - \alpha_0 - \alpha_1 y_{t-1} - \dots - \alpha_p y_{t-p}) \\
\frac{\partial S(\alpha)}{\partial \alpha_1} &= -2 \sum_t y_{t-1} (y_t - \alpha_0 - \alpha_1 y_{t-1} - \dots - \alpha_p y_{t-p}) \\
&\vdots \\
&\vdots \\
&\vdots \\
\frac{\partial S(\alpha)}{\partial \alpha_p} &= -2 \sum_t y_{t-p} (y_t - \alpha_0 - \alpha_1 y_{t-1} - \dots - \alpha_p y_{t-p})
\end{aligned}$$

Les estimateurs $\alpha_{*,j}$ sont les valeurs telles que :

$$\frac{\partial \mathbf{S}(\alpha)}{\partial \alpha_j} = \mathbf{0} \quad , \quad \mathbf{j} = \mathbf{0}, \dots, \mathbf{p}$$

En substituant les valeurs des variables expliquée et explicative par les réalisations du processus $(X_t, t \in Z)$, cela nous ramène à la résolution du système d'équations linéaires suivant :

$$\begin{aligned}
\sum_t y_t &= n\alpha_{*,0} + \sum_t x_{t-1}\alpha_{*,1} + \sum_t x_{t-2}\alpha_{*,2} + \dots + \sum_t x_{t-p}\alpha_{*,p} \\
\sum_t y_t x_{t-1} &= \sum_t x_{t-1}\alpha_{*,0} + \sum_t x_{t-1}^2\alpha_{*,1} + \sum_t x_{t-1}x_{t-2}\alpha_{*,2} + \dots + \sum_t x_{t-1}x_{t-p}\alpha_{*,p} \\
&\vdots \\
&\vdots \\
\sum_t y_t x_{t-p} &= \sum_t x_{t-p}\alpha_{*,0} + \sum_t x_{t-1}x_{t-p}\alpha_{*,1} + \sum_t x_{t-2}x_{t-p}\alpha_{*,2} + \dots + \sum_t x_{t-p}^2\alpha_{*,p}
\end{aligned}$$

Sous forme matricielle ce système s'écrit de manière suivante :

$$\begin{pmatrix} \sum_t y_t \\ \sum_t y_t x_{t-1} \\ \vdots \\ \sum_t y_t x_{t-p} \end{pmatrix} = \begin{pmatrix} n & \sum_t x_{t-1} & \sum_t x_{t-2} & \cdot & \cdot & \sum_t x_{t-p} \\ \sum_t x_{t-1} & \sum_t x_{t-1}^2 & \sum_t x_{t-2}x_{t-1} & \cdot & \cdot & \sum_t x_{t-p}x_{t-1} \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ \sum_t x_{t-p} & \sum_t x_{t-1}x_{t-p} & \sum_t x_{t-2}x_{t-p} & \cdot & \cdot & \sum_t x_{t-p}^2 \end{pmatrix} \begin{pmatrix} \alpha_{*,0} \\ \alpha_{*,1} \\ \vdots \\ \alpha_{*,p} \end{pmatrix}$$

Ce qui est équivalent à :

$$X^t Y = X^t X \alpha_{*,n}$$

Si $X^t X$ est inversible alors l'estimateur $\alpha_{*,n}$ de α par la méthode des moindres carrés ordinaires est donnée par :

$$\alpha_{*,n} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Y} \quad \text{tel que} \quad \frac{\partial^2 \mathbf{S}(\alpha)}{\partial \alpha_j^2} > 0$$

Remarque

Si le processus est centré ie si $\alpha_0 = \mathbf{0}$, nous cherchons à minimiser :

$$\sum_t \epsilon_t^2 = \sum_t (\mathbf{X}_t - \alpha_1 \mathbf{X}_{t-1} - \alpha_2 \mathbf{X}_{t-2} - \dots - \alpha_p \mathbf{X}_{t-p})^2$$
$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_p \end{pmatrix}$$

α est le vecteur des paramètres

$$X = \begin{pmatrix} X_p & X_{p-1} & \dots & X_1 \\ X_{p+1} & X_p & \dots & X_2 \\ & & \cdot & \\ & & \cdot & \\ & & \cdot & \\ X_{n-1} & X_{n-1} & \dots & X_{n-p} \end{pmatrix}$$

X regroupe les variables explicatives

$$Y = \begin{pmatrix} X_{p+1} \\ X_{p+2} \\ \cdot \\ \cdot \\ \cdot \\ X_n \end{pmatrix}$$

Y les variables à expliquer

$$\epsilon = \begin{pmatrix} \epsilon_{p+1} \\ \epsilon_{p+2} \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_n \end{pmatrix}$$

ϵ est le vecteur aléatoire

Nous avons un résultat qui découle de la régression linéaire multiple qui dit que l'estimateur des moindres carrés est donné par :

$$\alpha_{*,n} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Y}$$

Théorème 4 (Gauss Markov)

L'estimateur $\alpha_{*,n}$ obtenu par la méthode des (MCO) est un estimateur BLUE (Best Linear Unbiased Estimator) ie le meilleur estimateur linéaire possible pour le vecteur α

3.2.2 Estimateurs du maximum de vraisemblance

D'après certaines hypothèses sur le modèle :

$$\mathbf{Y} - \mathbf{X}\alpha = \epsilon \sim \mathbf{N}(\mathbf{0}_{n \times 1}, \Sigma = \sigma^2 \mathbf{1}_{n \times n})$$

Alors la vraisemblance (Likelihood) est donnée par :

$$\mathbf{L}(\alpha, \sigma_\epsilon^2) = \left(\frac{\mathbf{1}}{\sigma \sqrt{2\pi}} \right)^n \exp \left(-\frac{\mathbf{1}}{2\sigma_\epsilon^2} \sum_t \epsilon_t^2 \right)$$

$$\mathbf{L}(\alpha, \sigma_\epsilon^2) = \left(\frac{\mathbf{1}}{\sigma \sqrt{2\pi}} \right)^n \exp \left(-\frac{(\mathbf{Y} - \mathbf{X}\alpha)^t (\mathbf{Y} - \mathbf{X}\alpha)}{2\sigma_\epsilon^2} \right)$$

En utilisant les résultats précédents , l'estimateur du (MV) du vecteur inconnu α est:

$$\alpha_{*,n} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Y}$$

En effet un maximum de $l(\alpha, \sigma_\epsilon^2)$ est exactement un minimum par rapport à α de $S(\alpha) = \sum_t (Y - X\alpha)^2$, pour tout $t = 1, \dots, n$.

Preuve

$$\log L(\alpha, \sigma_\epsilon^2) = -\frac{n}{2} \log(2\pi\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2} S(\alpha)$$

Notre but est de maximiser cette log-vraisemblance par rapport au vecteur α et σ_ϵ^2

Les composants du vecteurs α n'apparaissent que dans le terme en $-S(\alpha)$

, or maximiser $-S(\alpha)$ équivaut à minimiser son contraire $S(\alpha)$, d'où les estimateurs des (MCO) et du (MV) de α sont égaux.

L'estimateur de σ_ϵ^2 est retrouvé par maximisation de la log-vraisemblance :

$$\frac{\partial \log \mathbf{L}(\alpha_{*,\mathbf{n}}, \sigma_\epsilon^2)}{\partial \sigma_\epsilon^2} = -\frac{\mathbf{n}}{2\sigma_\epsilon^2} + \frac{1}{2\sigma_\epsilon^4} \mathbf{S}(\alpha_{*,\mathbf{j}}) = -\frac{\mathbf{n}}{2\sigma_\epsilon^2} + \frac{1}{2\sigma_\epsilon^4} \sum_{\mathbf{t}} (\mathbf{Y} - \mathbf{X}\alpha_{*,\mathbf{q}})^2, \quad \mathbf{q} = \overline{0, \dots, \mathbf{n}}$$

Ainsi, l'estimateur du (MV) de σ_ϵ^2 est :

$$\sigma_{*,\text{MV}}^2 = \frac{1}{\mathbf{n}} \sum_{\mathbf{t}} \epsilon_{*,\mathbf{t}}^2$$

- Nous savons que cet estimateur est biaisé mais ce biais semble négligeable quand la taille de l'échantillon devient très grande.
- Lorsque le processus $(X_t, t \in Z)$ est centré, les estimateurs des (MCO) et du (MV) sont les mêmes et possèdent les mêmes propriétés de convergence et de normalité asymptotique, ces estimateurs sont obtenus par la résolution des équations de Yule Walker.

3.2.3 La méthode de Yule Walker

Supposons que le processus $(X_t, t \in Z)$ est stationnaire et centré.

La méthode de Yule Walker repose sur l'étude de la fonction d'autocovariance du processus stationnaire sur Z .

En multipliant de chaque côté de l'égalité par X_{t-1} et X_{t-2} respectivement on a :

$$\begin{aligned} X_{t-1}X_t &= \alpha_1 X_{t-1}^2 + \alpha_2 X_{t-1}X_{t-2} + \dots + \alpha_p X_{t-1}X_{t-p} + X_{t-1}\epsilon_t \\ X_{t-2}X_t &= \alpha_1 X_{t-2}X_{t-1} + \alpha_2 X_{t-2}^2 + \dots + \alpha_p X_{t-2}X_{t-p} + X_{t-2}\epsilon_t \end{aligned}$$

Par passage à l'espérance on obtient les équations :

$$\begin{aligned} \gamma(1) &= \alpha_1 \gamma(0) + \alpha_2 \gamma(1) + \dots + \alpha_p \gamma(p-1) \\ \gamma(2) &= \alpha_1 \gamma(1) + \alpha_2 \gamma(0) + \dots + \alpha_p \gamma(p-2) \end{aligned}$$

En itérant cette équations p fois , on aura un système de p équations à p inconnus donné par :

$$\begin{pmatrix} \gamma(0) & \dots & \gamma(p-1) \\ \gamma(1) & \dots & \gamma(p-2) \\ \vdots & & \vdots \\ \vdots & & \vdots \\ \gamma(p-1) & \dots & \gamma(0) \end{pmatrix} \begin{pmatrix} \alpha(1) \\ \alpha(2) \\ \vdots \\ \vdots \\ \alpha(p) \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \vdots \\ \gamma(p) \end{pmatrix}$$

La matrice des variables explicative est symétrique et constante par diagonales appelée **matrice de Toeplitz** inversible lorsque le processus est stationnaire.

Les valeurs de l'autocovariance étant inconnues il est donc impossible d'utiliser ce système pour estimer le vecteur des paramètres , nous allons donc estimer ces valeurs.

Estimation des valeurs de la fonction d'autocovariance

Pour tout $h \in Z$, $t \in \{p+1, \dots, n\}$:

$$\gamma_*(\mathbf{h}) = \frac{1}{n} \sum_{t=p+1}^{n-|h|} \mathbf{X}_t \mathbf{X}_{t+|h|} \xrightarrow{p} \gamma(\mathbf{h})$$

Ce qui signifie que γ_* est un estimateur consistant de γ et l'estimateur du vecteur paramétrique α est donné par :

$$\alpha_{*,\mathbf{n}} = \mathbf{\Gamma}_*^{-1} \gamma_*$$

Avec :

$$\mathbf{\Gamma}_* = \begin{pmatrix} \gamma_*(0) & \dots & \gamma_*(p-1) \\ \gamma_*(1) & \dots & \gamma_*(p-2) \\ \vdots & & \vdots \\ \vdots & & \vdots \\ \gamma_*(p-1) & \dots & \gamma_*(0) \end{pmatrix} \quad \text{et} \quad \gamma_* = \begin{pmatrix} \gamma_*(1) \\ \gamma_*(2) \\ \vdots \\ \vdots \\ \gamma_*(p) \end{pmatrix}$$

Remarques

- Si le processus $(X_t, t \in Z)$ n'est pas centré , ie $(\mu \neq 0)$ alors on remplace X_t et $X_{t+|h|}$ respectivement par $X_t - \bar{X}_n$ et $X_{t+|h|} - \bar{X}_n$

- Le dernier terme obtenu en multipliant de chaque côté de l'égalité par X_t est :

$$\gamma(0) = \alpha_1\gamma(1) + \alpha_2\gamma(2) + \dots + \alpha_p\gamma(p) + \sigma^2$$

Ce qui donne l'estimateur consistant de la variance $\sigma_{*,\epsilon}^2$ de σ_ϵ^2 défini par :

$$\sigma_{*,n,\epsilon}^2 = \gamma_*(\mathbf{0}) - \alpha_{*,1}\gamma_*(\mathbf{1}) - \alpha_{*,2}\gamma_*(\mathbf{2}) - \dots - \alpha_{*,p}\gamma_*(\mathbf{p})$$

Avec $\alpha_{*,i}$, $i = 1, \dots, p$ les composantes de $\alpha_{*,n}$

- Les estimateurs des moindres carrés et de Yule Walker sont quasi-équivalents dans une situation de stationnarité et possèdent les mêmes propriétés de convergence et de normalité asymptotique.
- $\alpha_{*,n}$ est un estimateur sans biais de α

4 Premiers pas vers l'estimation robuste

Les estimateurs des (MCO) sont-ils robustes?

Lorsque toutes les données sont entachées d'une petite erreur ou seulement quelques unes sont entachées d'une très grosse erreur , les estimateurs des (MCO) sont très sensibles à la présence de ces données , ie qu'ils ne tiennent pas compte des valeurs aberrantes du modèle.

Les valeurs aberrantes ne sont pas les seules qui peuvent biaiser un modèle homoscédastique , il existe un autre cas dans lequel un estimation robuste peut être envisagée , c'est lorsqu'il existe une forte suspicion d'hétéroscédasticité des erreurs , cela signifie clairement que la variance des erreurs n'est pas finie, ie que l'hypothèse sur laquelle est basée la méthode des (MCO) est violée , dans ce cas il faut forcément utiliser d'autres techniques plus robustes pour estimer ces modèles.

Exemple

Un commerçant de bandes dessinées pense que le montant de ses **ventes** mensuelles est corrélé avec 3 variables , **le prix** , le budget alloué à la **publicité TV** et celui alloué à la **publicité internet**.

Il a donc collecté ces données mensuelles ci-dessous :

<i>Ventes</i>	<i>Prix</i>	<i>PubTV</i>	<i>PubInt</i>
73959	8.75	50.04	61.13
71544	8.99	50.74	60.19
78587	7.50	50.14	59.16
80364	7.25	50.27	60.38
78771	7.40	51.25	59.71
71986	8.50	50.65	59.88
74885	8.40	50.87	60.14
73345	7.90	50.15	60.08
76659	7.25	48.24	59.90
71880	8.70	50.19	59.68
73598	8.40	51.11	59.83
74893	8.10	51.49	59.77
69003	8.40	50.10	59.29
78542	7.40	49.24	60.40
72543	8	50.04	59.89
74247	8.30	49.46	60.06
76253	8.10	51.62	60.51
72582	8.20	49.78	58.93
69022	8.99	48.60	60.09
76200	7.99	49	61
69701	8.50	48	59
77005	7.90	54	59.50
70987	7.99	48.70	58

Objectif :

Nous allons tenter de répondre à la question précédente en effectuant une première régression linéaire des ventes sur les 3 autres variables pour faire une **estimation** des ventes par la méthode des (MCO) étudiée plus haut; puis réaliser une seconde régression lorsque nous modifions l'échantillon original en introduisant des valeurs aberrantes et comparer les deux résultats obtenus. L'algorithme suivant sous le logiciel R nous donne l'équation linéaire :

NB : Avant d'exécuter cet algorithme sous le logiciel *R* , assurez vous que le fichier de données des ventes est d'abord enregistré dans votre répertoire de travail accessible par la commande `getwd()`.

Importation des données

```
drv = read.table("drv.txt", header = TRUE)  
drv  
attach(drv)
```

Identification des variables

```
names(drv)
```

Tracé des nuages de points

```
Y = Ventes  
X = Prix + PubTV + PubInt  
plot(X, Y, pch = 20)  
pairs(drv)
```

Modèle de régression linéaire multiple

```
reg = lm(Ventes ~ ., data = drv)  
reg  
summary(reg)
```

Analyse des résidus

```
res = residuals(reg)  
res  
res = rstudent(reg)  
res
```

Ventes estimés en fonction des résidus studentisés

```
fitted_values = predict(reg)  
fitted_values
```

Diagnostic des ventes estimés

```
plot(fitted_values, res, xlab = "fitted_values", ylab = "Residuals")  
abline(h = c(-2, 0, 2), lty = 1, col = c(1, 2, 1))  
plot(reg)
```

Test de normalité des résidus

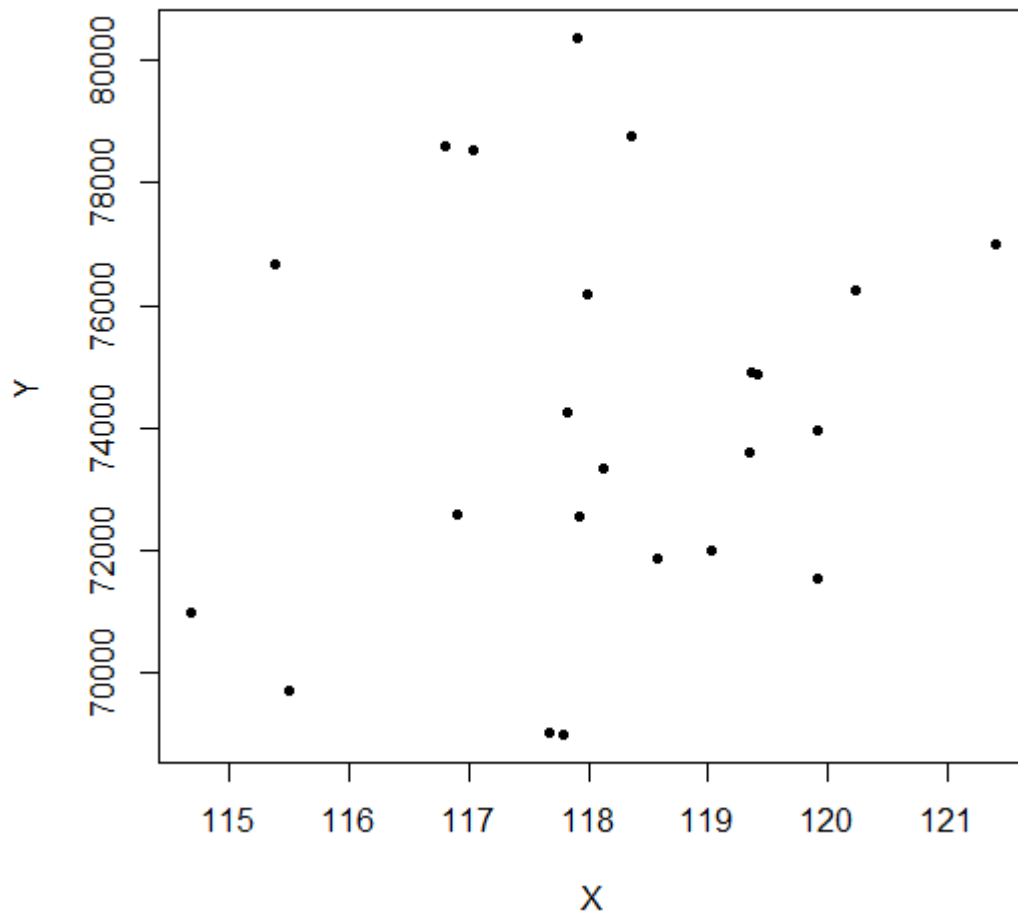
```
shapiro.test(res)
```

Histogramme des résidus

```
hist(res)
```

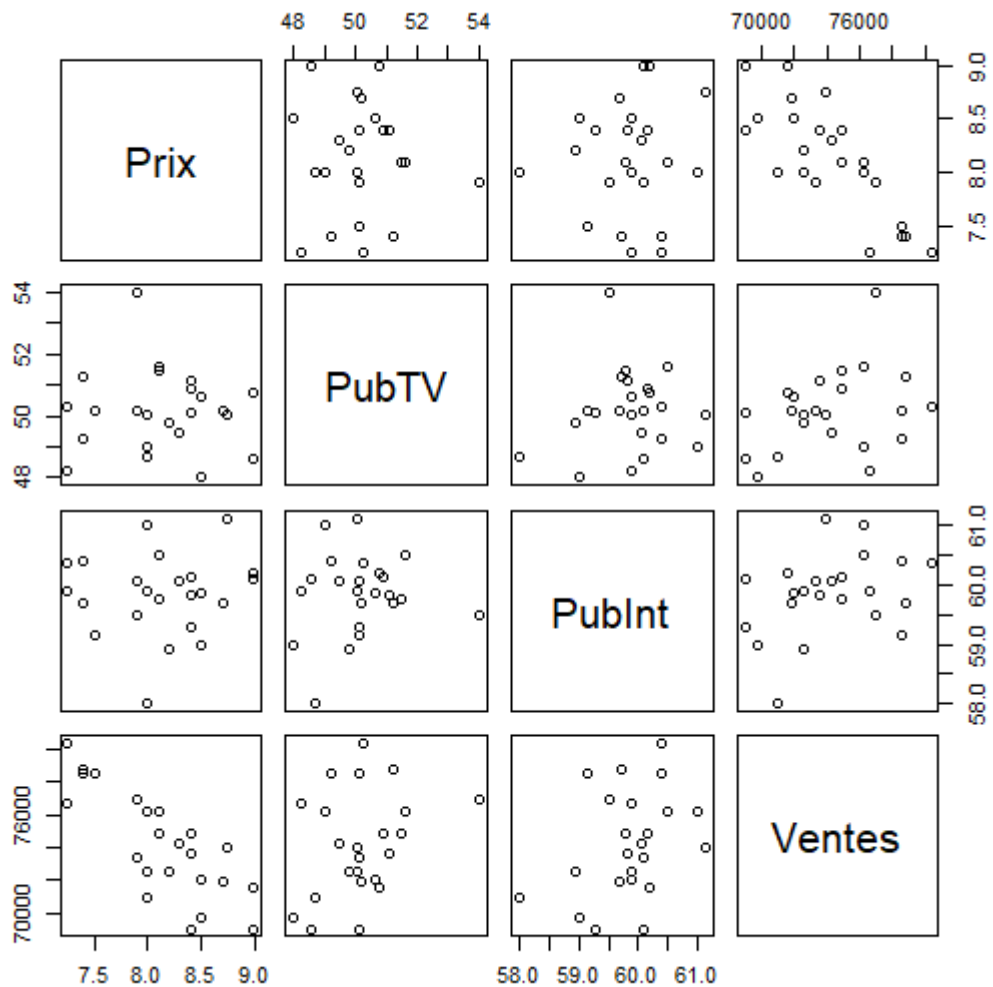
Résultats de R

Nuage de points



Les observations sont aléatoirement distribuées sur le nuage de point sauf que quelques unes d'entre elles s'éloignent de la majorité. Ces observations distantes peuvent présenter des aspects assez intéressants dans l'étude statistique du modèle et le statisticien doit les prendre en compte pendant l'analyse des données.

Nuages de points par paires



Ce schéma donne une idée sur le degré de la relation entre les variables prises deux à deux, en visualisant de plus près cette figure, on constate qu'il y a une bonne corrélation entre les ventes de notre commerçant et le prix de sa marchandise car le croisement entre ces deux variables montre un nuage de point formant une droite.

Il existe aussi une corrélation entre les deux autres variables restantes mais ces corrélations sont moins bonnes qu'avec le prix des bandes dessinées.

Après avoir compiler la suite du programme dans R , les résultats obtenus sont les suivants :

```
Call:
lm(formula = Ventes ~ ., data = drv)

Residuals:
    Min       1Q   Median       3Q      Max
-2796.5  -224.4   273.6   736.8  2425.0

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -21721.9     25485.6  -0.852  0.404651
Prix         -5069.6       532.0  -9.528  1.14e-08 ***
PubTV         656.0       211.3   3.104  0.005840 **
PubInt        1741.2       404.6   4.303  0.000384 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1288 on 19 degrees of freedom
Multiple R-squared:  0.862,    Adjusted R-squared:  0.8402
F-statistic: 39.55 on 3 and 19 DF,  p-value: 2.283e-08
```

Examen des résidus

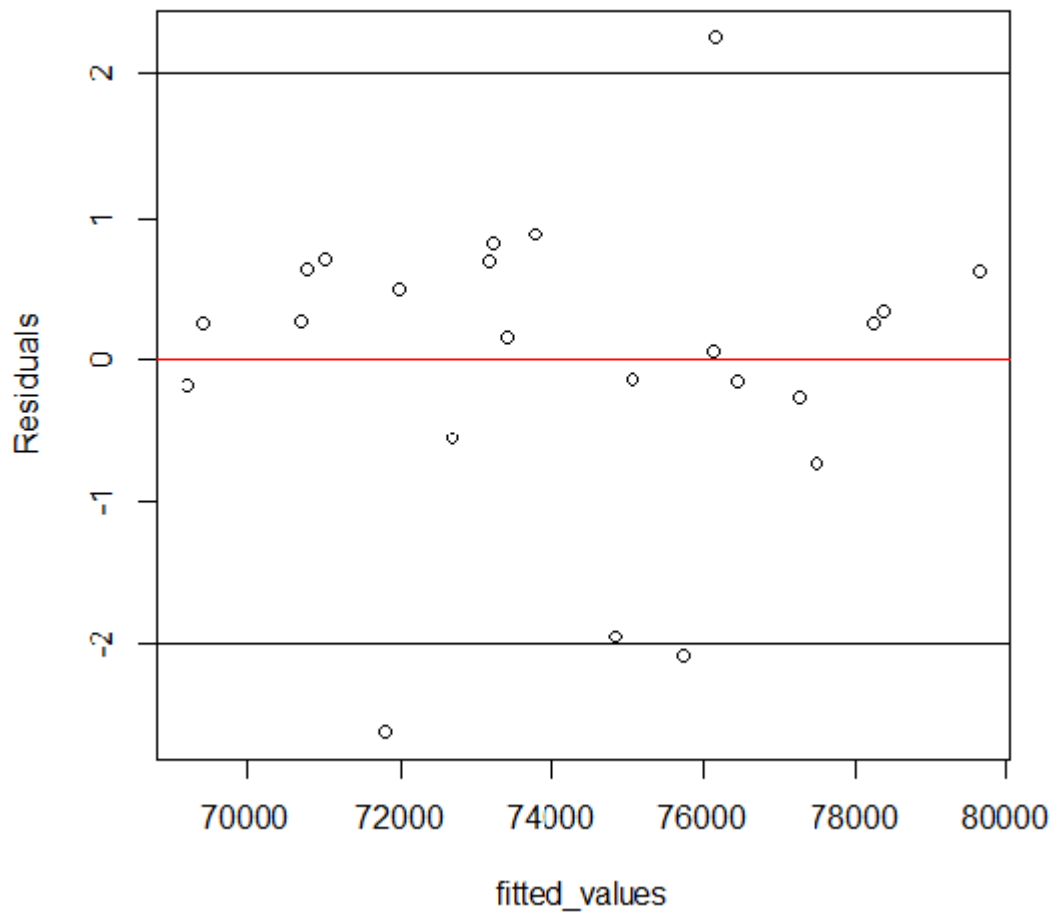
```
> res
      1          2          3          4          5          6
 769.33645  748.58676 2425.02548  725.02368  416.18621 -694.67492
      7          8          9         10         11         12
1100.31592 -2397.65409 -812.43320  863.26430  195.65565 -175.03378
     13         14         15         16         17         18
-2796.47819  304.34717 -2289.69580 1019.66247 -188.84223  605.38653
     19         20         21         22         23
 -195.38137  66.09441  291.10727 -253.45003  273.65130
```

Résidus studentisés

```
> rstd
      1          2          3          4          5          6
 0.68442336  0.63397660  2.24942706  0.62314270  0.34519506 -0.55022965
      7          8          9         10         11         12
 0.88330159 -2.07989497 -0.73589171  0.69866709  0.15443231 -0.13896776
     13         14         15         16         17         18
-2.59687323  0.25796094 -1.95112521  0.81221720 -0.15341550  0.49348796
     19         20         21         22         23
-0.16920180  0.05701829  0.25377257 -0.26402530  0.26915577
```

Valeurs ajustées des ventes

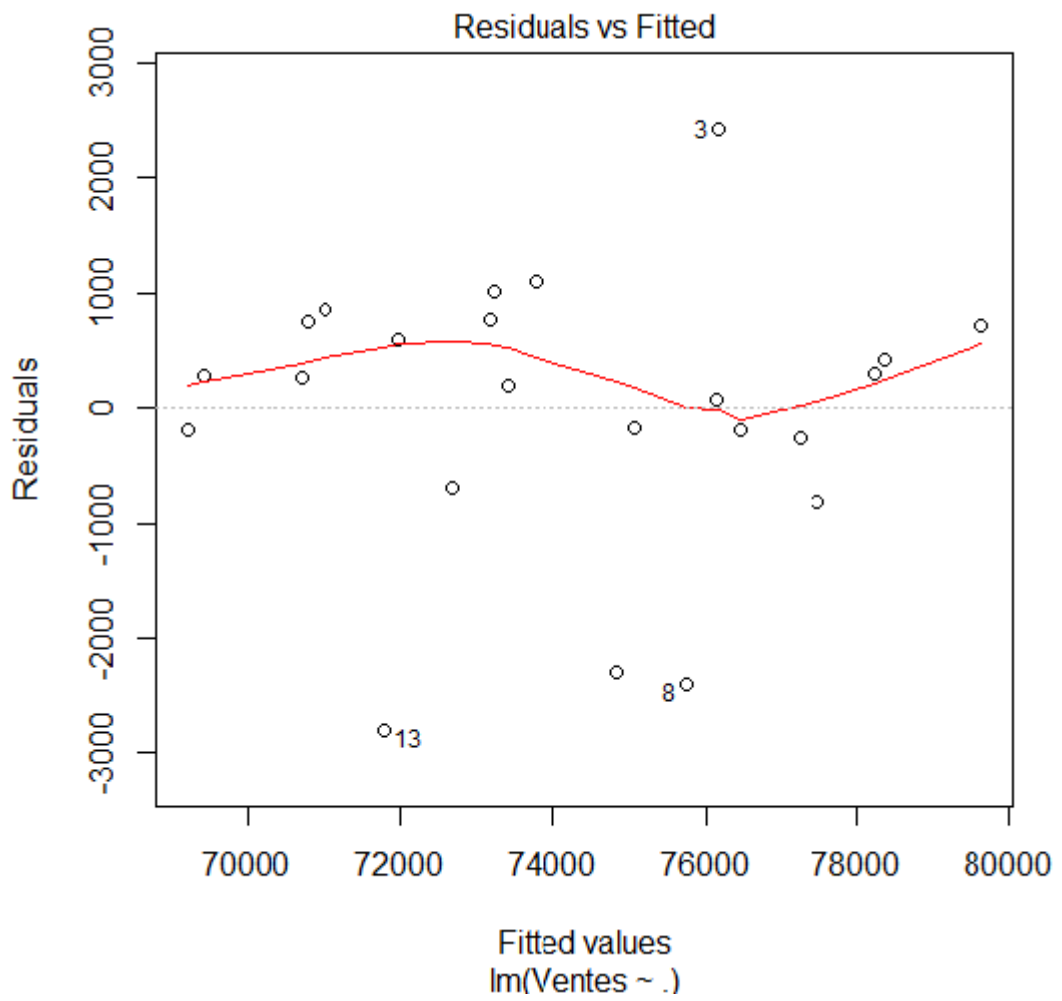
1	2	3	4	5	6	7	8
73189.66	70795.41	76161.97	79638.98	78354.81	72680.67	73784.68	75742.65
9	10	11	12	13	14	15	16
77471.43	71016.74	73402.34	75068.03	71799.48	78237.65	74832.70	73227.34
17	18	19	20	21	22	23	
76441.84	71976.61	69217.38	76133.91	69409.89	77258.45	70713.35	

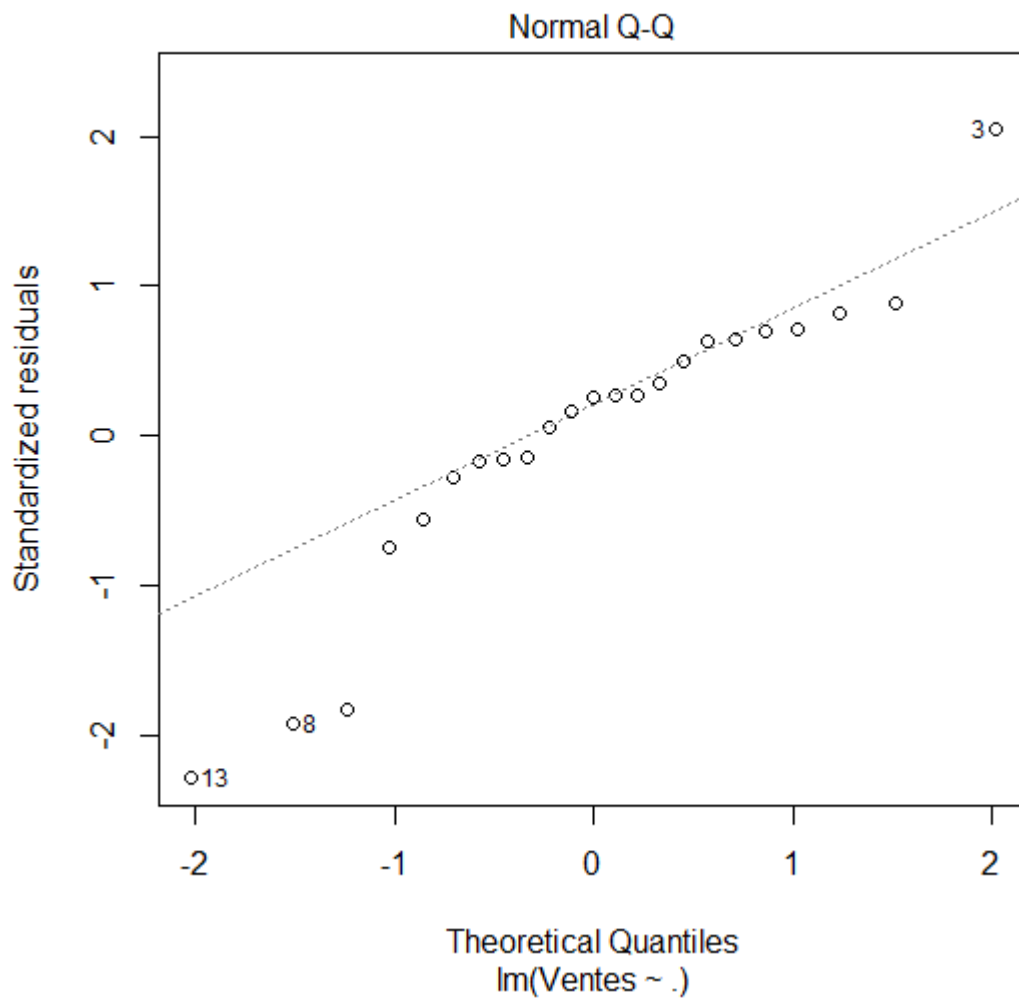


Shapiro-Wilk normality test

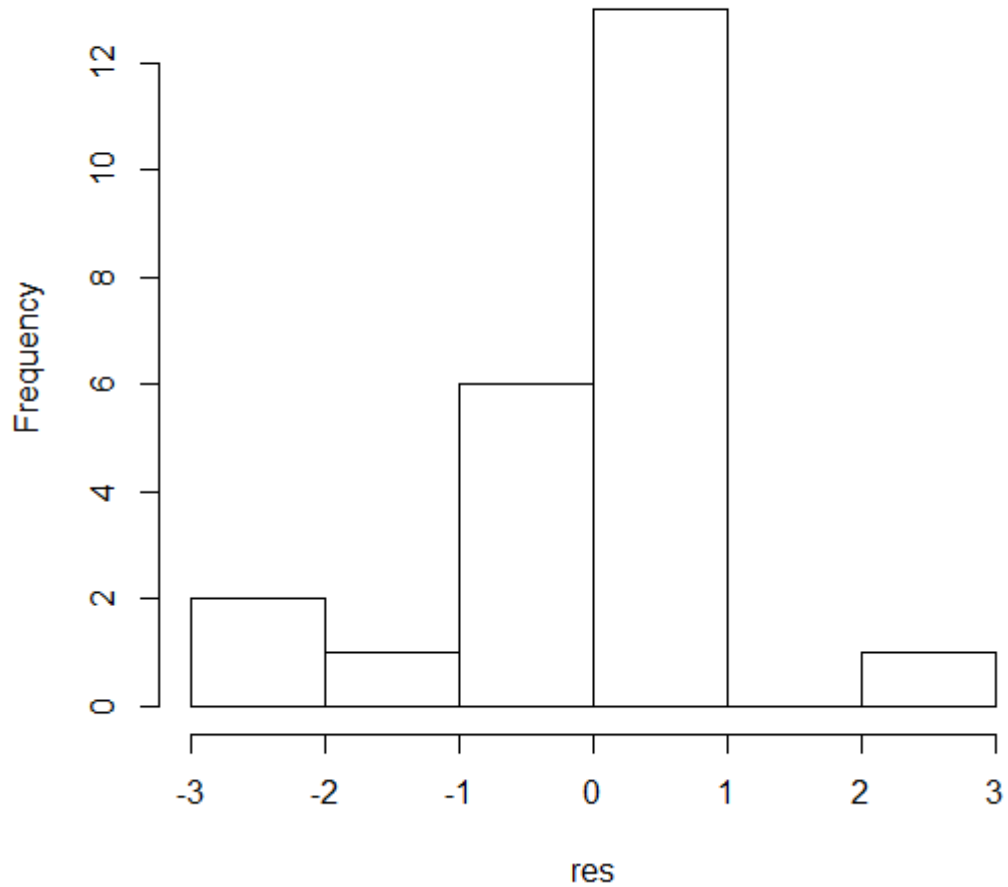
data : res

$W = 0.89354$, $p - value = 0.01865$





Histogram of res



Interprétation des résultats de R

Les coefficients de corrélation sont assez proches de 1 , alors le commerçant avait raison et les ventes sont corrélées avec les 3 autres variables. Toutes les variables explicatives sont statistiquement significatives. Le budget alloué pour l'internet est plus rentable que celui alloué pour la télévision. Le test de shapiro donne une p-value égale à 0,02369 inférieure à 5⁰/₀ donc les résidus sont gaussiens.

Remarque

On déduit alors que l'estimateur $\alpha_{*,n}$ de α est le vecteur :

$$\alpha_{*,n} = \begin{pmatrix} \alpha_{*,0} \\ \alpha_{*,1} \\ \alpha_{*,2} \\ \alpha_{*,3} \end{pmatrix} = \begin{pmatrix} -21721.9 \\ -5069.6 \\ 656 \\ 1741.2 \end{pmatrix}$$

Bien que cet estimateur soit BLUE , il est non robuste car il est sensible aux déviations vis à vis du modèle pour lequel il a été optimisé , d'où le concept d'**estimation robuste**.

Comment construire le meilleur modèle pour la prévision ?

Les variables explicatives qui apportent beaucoup d'information sur le modèle sont appelées "variables exhaustives".

En effet , il est indispensable de passer par cette étape lorsque la méthode n'est pas robuste , ainsi pour construire un tel modèle , plusieurs stratégies sont proposées parmi tant d'autres :

Méthode descendante : On construit le modèle complet , on reconstruit un modèle sans la variable explicative la moins intéressante ie celle ayant la plus grande probabilité d'être dans la région critique ; puis on itère jusqu'à ce que toutes les variables explicatives soient intéressantes.

Méthode ascendante : On part du modèle avec la variable la plus intéressante , on ajoute la variable qui connaissant les autres variables apportent le plus d'information ; puis on itère jusqu'à ce qu'aucune variable ne soit exhaustive.

Méthode du R^2 : On construit tous les sous modèles possibles et on retient celui pour lequel la probabilité critique du test du R^2 est la plus petite (on rejette le plus souvent l'hypothèse $H_0 : \alpha_0 = \alpha_1 = \dots \alpha_p = 0$)

Remarque

En effet , dans notre exemple il est possible d'estimer les ventes du commerçant en adoptant une des trois options proposées ci dessus , nous allons choisir la dernière procédure de **sélection de variables** pour construire le modèle final afin de prévoir les valeurs futures :

```

$summary
                R2      Pvalue
Model with 1 variable 0.6333637 5.609729e-06
Model with 2 variables 0.7919749 1.517598e-07
Model with 3 variables 0.8619756 2.282903e-08

$best

Call:
lm(formula = as.formula(as.character(formul)), data = don)

Residuals:
    Min       1Q   Median       3Q      Max
-2796.5  -224.4   273.6   736.8  2425.0

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -21721.9     25485.6  -0.852  0.404651
Prix         -5069.6       532.0  -9.528 1.14e-08 ***
PubTV          656.0       211.3   3.104 0.005840 **
PubInt         1741.2       404.6   4.303 0.000384 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1288 on 19 degrees of freedom
Multiple R-squared:  0.862,    Adjusted R-squared:  0.8402
F-statistic: 39.55 on 3 and 19 DF,  p-value: 2.283e-08

```

Ainsi le modèle final à retenir pour estimer les ventes est :

$$\text{Ventes} = 1741.2 \times \text{PubInt} - 5069.6 \times \text{Prix} + 656 \times \text{PubTV} - 21721.9 + \epsilon$$

Remarque

La valeur des coefficients estimés du modèle final peut être retrouvée en utilisant la commande **confint** qui donne les intervalles de confiance de chaque paramètre estimé.

```

                2.5 %    97.5 %
(Intercept) -75063.8471 31620.001
Prix         -6183.1611 -3955.989
PubTV          213.6986  1098.364
PubInt         894.3333  2588.165

```

5 Estimation Robuste et Régression linéaire

En statistiques , la robustesse d'un estimateur est sa qualité à ne pas être perturbé par une petite modification dans les données ou dans les paramètres du modèle choisi pour l'estimation.

Les techniques de régression robuste sont les plus recommandées pour la recherche de tels modèles car elles donnent une estimation plus solide des paramètres du modèle même après modification de l'échantillon de données.

Point de rupture

Le point de rupture est une notion très importante quand on parle de robustesse des estimateurs , il est défini comme étant la plus petite portion de contamination dans un échantillon qui déstabilise complètement un estimateur. **Donoho** et **Huber** , en 1983 ont donné une définition plus simple de cette notion et cette définition permet de travailler avec des échantillons finis.

Définition 3 (Donoho et Huber)

Soit $E = \{(x_i, y_i)\}_{i=1}^n$ un échantillon de n points.
Soit le vecteur $\alpha_{*,n}$ l'estimateur de α et $\lambda(m, \alpha_{*,n}, E)$ tq:

$$\lambda(m, \alpha_{*,n}, E) = \sup \| \alpha_{*,n}(X') - \alpha_{*,n}(X) \|_2$$

pour tous les échantillons corrompus X' dans lequel m points de l'échantillon original X ont été remplacées par des valeurs arbitraires. Le point de rupture de $\alpha_{*,n}$ est :

$$\varphi_n = \min_m \left\{ \frac{m}{n}; \lambda(m, \alpha_{*,n}, E) \right\}$$

En effet , si on remplace un seul point de l'échantillon original par un point arbitrairement loin de cet échantillon on obtient:

$$\lambda(1, \alpha_{*,n}, E) = \infty$$

Le point de rupture de l'estimateur des (MCO) $\alpha_{*,n}$ est :

$$\varphi_n = \min \left\{ \frac{1}{n}; \infty \right\} = \frac{1}{n}$$

Faisons un test sur les données de ventes précédentes et analysons les résultats obtenus :

Le but de ce test est de mettre en évidence les nuances entre cette régression et celle obtenue avec l'échantillon original afin d'avoir une idée sur la présence

d'outliers (points aberrants) du modèle obtenu avec la méthode des (MCO).

Call :
 $lm(formula = Ventes ., data = ventes)$

Residuals :

<i>Min</i>	<i>1Q</i>	<i>Median</i>	<i>3Q</i>	<i>Max</i>
-2524.4	-462.0	-109.0	575.8	3956.7

Coefficients :

	<i>Estimate</i>	<i>Std. Error</i>	<i>tvalue</i>	<i>Pr(> t)</i>
<i>(Intercept)</i>	-17144.3	29368.1	-0.584	0.56624
<i>Prix</i>	-4378.9	583.4	-7.505	4.27e-07 ***
<i>PubTV</i>	648.9	240.9	2.693	0.01440*
<i>PubInt</i>	1582.3	469.6	3.369	0.000322 **

Signif. codes : 0 '***' 0.001 '***' 0.01 '***' 0.05 '.' 0.1 '1'

Residual standard error : 1468 on 19 degrees of freedom

Multiple R-squared : 0.7959 , *Adjusted R-squared :* 0.7637

F-statistic : 24.69 on 3 and 19 DF , *p-value :* 9.065e-07

Remarque

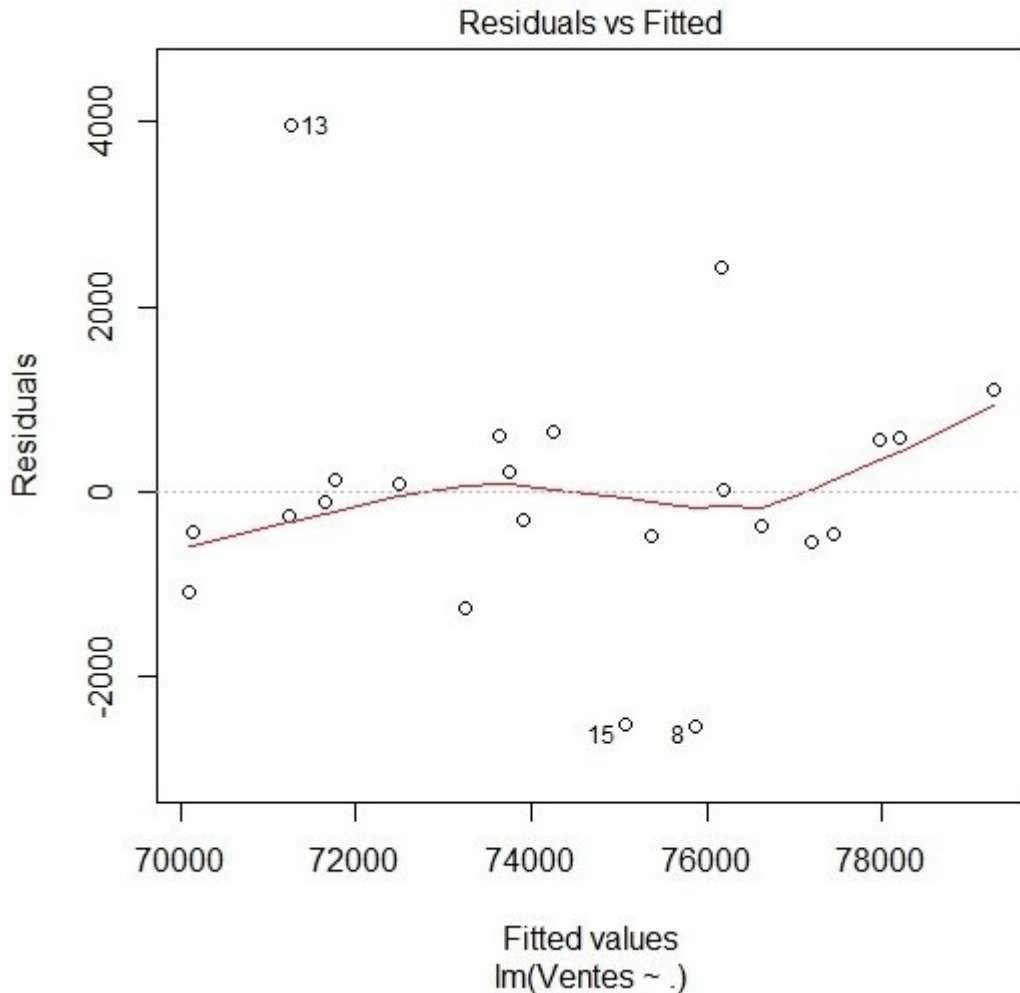
Le taux de significativité des deux dernières variables explicatives a diminué, ce qui nous donne déjà une idée sur la qualité de la régression par rapport à l'échantillon original.

Les coefficients de détermination R^2 et R^2 ajustés ont diminué de 10⁰/0 mais par contre on constate une augmentation de la statistique de Fisher.

Globalement , le modèle reste toujours significatif mais nous conseillons au commerçant de ne pas choisir ce modèle pour estimer ses ventes car ceci a été réalisé dans le seul but de montrer l'utilité de l'estimation robuste.

Malgré sa non robustesse , le meilleur estimateur linéaire possible valable pour estimer les ventes du commerçant reste toujours celui obtenu avec l'échantillon initial :

$$\alpha_{*,n} = \begin{pmatrix} \alpha_{*,0} \\ \alpha_{*,1} \\ \alpha_{*,2} \\ \alpha_{*,3} \end{pmatrix} = \begin{pmatrix} -21721.9 \\ -5069.6 \\ 656 \\ 1741.2 \end{pmatrix}$$



Lorsqu'on remplace un point de l'échantillon original des données de ventes par des valeurs arbitraires , on remarque que les observations 8 et 13 semblent toujours être des points aberrants tandis que l'observation 3 a été remplacée par l'observation 15 et par conséquent ne semble plus aberrante par rapport à l'échantillon original ; on peut donc en déduire que les observations 8 et 13 sont bien des **outliers** de l'échantillon initial. On observe également une grande dispersion de certains points autour de la courbe de régression , ce qui signifie que plus on change de valeurs , moins l'estimation est bonne ; d'où la **sensibilité** de la méthode des (MCO).

Remarque

Par ailleurs , il est possible d'utiliser un ajustement robuste sur ces données pour voir le comportement du modèle par rapport aux outliers et faire une comparaison avec le modèle final proposé par les (MCO).

5.1 Quelques méthodes d'estimation robuste

Il existe plusieurs techniques d'estimation plus robuste que celle des (MCO) mais seulement quelques unes d'entre elles en particulier nous intéressent et seront présentées dans la suite de ce mémoire , à savoir la **L-estimation** et la **M-estimation** des paramètres du modèle

5.1.1 L-estimation

L-estimateurs

Définition 4

Soient $X_{(1)} \leq X_{(2)} \leq \dots X_{(n)}$ les statistiques d'ordre d'un échantillon de taille n .

Un L-estimateur T est une combinaison linéaire particulière des $X_{(t)}$, ie :

$$T = \sum_{t=1}^n \alpha_t X_{(t)}$$

Avec $\alpha_t, t = \overline{1, n}$ des réels vérifiant $0 \leq \alpha_t \leq 1$ et $\sum_{t=1}^n \alpha_t = 1$ appelés poids de T

Quelques exemples de L-estimateurs :

Moyenne

La moyenne d'un échantillon de taille n est le L-estimateur dont tous les poids sont égaux à $\frac{1}{n}$. Pour la moyenne , toutes les statistiques d'ordre ont la même importance.

Médiane

La médiane d'un échantillon est le L-estimateur qui ne fait intervenir que la statistique d'ordre centrale si n est impair et qui donne la moyenne des deux statistiques d'ordre centrales si n est pair.

$$mediane = \begin{cases} X_{(k)} & Si \quad n=2k+1 \\ \frac{1}{2}X_{(k)} + \frac{1}{2}X_{(k+1)} & Si \quad n=2k \end{cases}$$

Remarque

La médiane est donc plus robuste que la moyenne empirique d'une population statistique.

Preuve

Le poids d'une population statistique est une v.a de moyenne μ inconnue. Pour estimer μ , on dispose d'un échantillon de 9 individus dont les poids en kg sont : 23 ; 38 ; 50 ; 25 ; 47 ; 29 ; 55 ; 52 ; et 23

La moyenne des observations est 38 et est égale à la médiane ; ensuite on choisit par erreur le poids de l'individu 7 égal à 155, la moyenne devient alors ≈ 49 tandis que la médiane reste invariante.

En conclusion, la médiane est meilleure que la moyenne en terme de robustesse.

Moyenne β -censurée

Soit $\beta \in R$ vérifiant $0 \leq \beta \leq \frac{1}{2}$

La moyenne β -censurée notée $T(\beta)$ est un L-estimateur ayant comme poids respectifs :

$$w_t = 0 \quad \text{Si } t \leq s \quad \text{ou } t \geq n - s + 1$$

$$w_t = \frac{1}{n(1 - 2\beta)} \quad \text{Si } t \in [s + 2 ; n - s - 1]$$

$$w_t = \frac{1 - r}{n(1 - 2\beta)} \quad \text{Si } t = s + 1 \quad \text{ou } t = n - s$$

Avec $s = [\beta n]$ et $r = \beta n - s$

Le principe général de cet estimateur repose sur l'élimination d'une proportion β des plus petites valeurs et une proportion β des plus grandes valeurs, puis de calculer la moyenne de l'ensemble des valeurs restantes. Par exemple si on choisit $n = 18$ et $\beta = 0.4$, on obtient $\beta n = 7.2$ et $s = 7$.

Après avoir ranger les observations, on élimine les 7 premières et 7 dernières valeurs, les valeurs restantes seront utilisées dans le calcul de la moyenne comme suit :

$$T(0.4) = \frac{0.8}{3.6}X_{(8)} + \frac{1}{3.6} \sum_{t=9}^{10} X_{(t)} + \frac{0.8}{3.6}X_{(11)}$$

Cette moyenne est appelée la **moyenne 0.4-censurée**.

Si on prend $n = 20$ et $\beta = 0.25$ on obtient la **mimoyenne** car seule la moitié centrale des observations triées est utilisée :

$$T(0.25) = \frac{1}{20} \sum_{t=6}^{10} X_{(t)}$$

5.1.2 M-estimation

L'étude de cette méthode est intéressante car c'est une forme généralisée de la méthode du (MV) vue plus haut. Elle a pour principe de minimiser une fonction coût qui permettra de pénaliser les résidus ϵ_t les plus élevés. La fonction coût à minimiser est :

$$C(\epsilon) = \sum_{t=1}^n \rho(\epsilon_t)$$

La fonction ρ est une fonction continue et symétrique appelée **fonction objective**, pour bien évaluer cette fonction il est nécessaire d'introduire la notion de ψ - fonction.

Définition 4

Une ψ - fonction est une fonction continue par morceaux définie de R dans R telle que :

- ψ est symétrique par rapport à l'origine $\{\psi(-x) = -\psi(x)\}$ pour tout x dans R
- $\psi(x) \leq 0$ pour $x \leq 0$ et $\psi(x) > 0$ pour $0 < x < x_r$
 $x_r = \sup\{x : \psi(x) > 0\}$ ($x_r > 0$ *possiblement* $x_r = \infty$)
- $\psi'(0) = 1$
- $\psi(0) = 0$ pour tous les cas où ψ est discontinue en $x = 0$

Définition 5

Une ρ - **fonction** au point x est représentée par l'intégrale d'une ψ - **fonction** sur $[0, x]$

$$\rho(\mathbf{x}) = \int_{\mathbf{0}}^{\mathbf{x}} \psi(\mathbf{u}) \, d\mathbf{u}$$

- 1) $\rho(0) = 0$
- 2) $\rho(-x) = \rho(x)$
- 3) $\rho'(x) = \psi(x)$

Remarque

Les ρ - *fonctions* dépendent de certaines constantes qui permettent d'augmenter la robustesse des estimateurs lorsqu'il y a présence de données aberrantes mais au détriment de leur efficacité, les deux types de ρ - *fonctions* les plus utilisées sont celles proposées par **Huber** et **Tuckey**.

La fonction de Huber

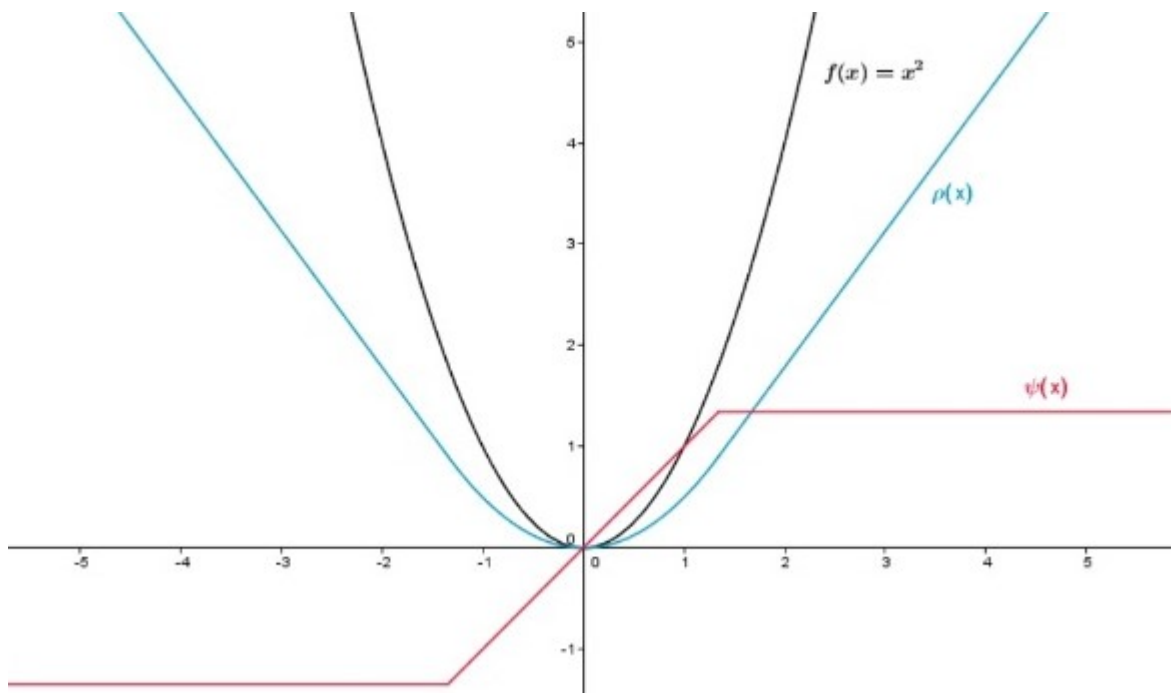
La fonction ρ proposée par **Huber** est définie pour une constante $c > 0$ par :

$$\rho(x) = \begin{cases} \frac{1}{2}x^2 & \text{Si } |x| \leq c \\ c(|x| - \frac{c}{2}) & \text{Si } |x| > c \end{cases}$$

La dérivée associée à cette fonction est donnée comme suit :

$$\psi(x) = \begin{cases} x & \text{Si } |x| \leq c \\ c \times \text{sign}(x) & \text{Si } |x| > c \end{cases}$$

Voyons ensemble le graphe de la fonction de **Huber** avec comme paramètre $c=1.345$:



La courbe en bleu est celle de la ρ – fonction de Huber , en rouge celle de la ψ – fonction et en noire celle de la fonction $y = x^2$, on remarque aussi que toutes les courbes ont l’origine comme point commun , ie qu’elles passent toutes par le point $P(0,0)$

La fonction de Tukey

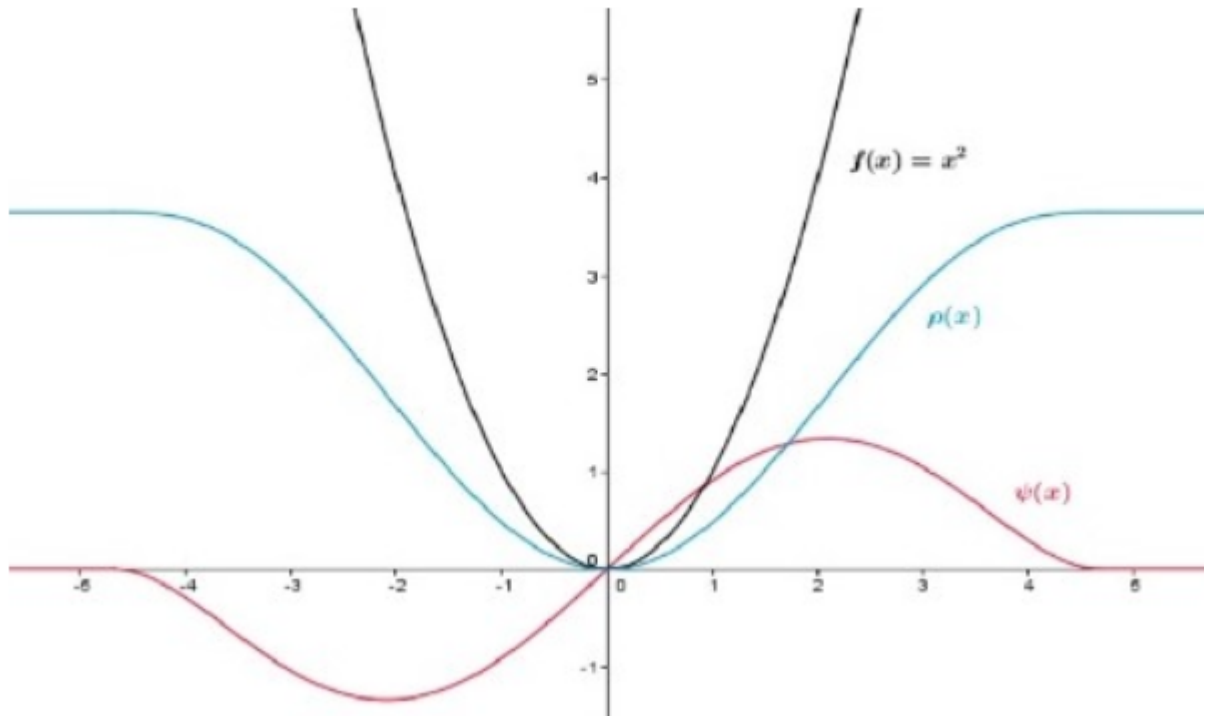
La fonction ρ proposée par **Tukey** est définie pour une constante $c > 0$ par :

$$\rho(x) = \begin{cases} \frac{c^2}{6} \left[1 - \left(1 - \left(\frac{x}{c} \right)^2 \right)^3 \right] & \text{Si } |x| \leq c \\ \frac{c^2}{6} & \text{Si } |x| > c \end{cases}$$

La dérivée associée à cette fonction est :

$$\psi(x) = \begin{cases} x \left[1 - \left(\frac{x}{c} \right)^2 \right]^2 & \text{Si } |x| \leq c \\ 0 & \text{Si } |x| > c \end{cases}$$

Regardons aussi la figure de la fonction de **Tukey** avec $c=4.685$:



Les deux dernières figures observées représentent les fonctions de **Huber** et **Tukey** avec des constantes $c=1.345$ et $c=4.685$ respectivement , nous verrons plutôt qu'utiliser ces valeurs de c permet d'obtenir une efficacité relative de 95%

On peut voir aussi que ces deux fonctions pénalisent les plus grandes erreurs en leur donnant moins de poids comparativement à la fonction objective $f(x) = x^2$ utilisée par la méthode des (MCO)

A partir du point $x = c$ les deux fonctions sont croissantes , mais elles croient moins vite que la fonction $f(x) = x^2$

Remarque

La médiane d'une population statistique est un M-estimateur pour la fonction ρ définie par :

$$\rho(x, t) = |x - t|$$

Dans ce cas , la valeur absolue n'a pas de dérivée en 0 , alors il faut choisir ψ tel que :

$$\psi(x, t) = \text{sgn}(x, t)$$

$$\text{sgn}(u) = \begin{cases} 1 & \text{Si } u > 0 \\ -1 & \text{Si } u < 0 \end{cases} \quad \text{et} \quad \text{sgn}(u) = 0 \quad \text{Si } u = 0$$

Efficacité relative

L'efficacité relative d'un estimateur $\alpha_{*,2}$ est le rapport entre son erreur quadratique moyenne (MSE) et celle de l'estimateur BLUE $\alpha_{*,1}$ connu pour le paramètre qu'on cherche à estimer ; dans le cas unidimensionnel elle s'écrit :

$$E_r = \frac{E[(\alpha_{*,1} - \alpha)^2]}{E[(\alpha_{*,2} - \alpha)^2]}$$

Considérons le même modèle de régression multiple :

$$\mathbf{Y} = \mathbf{X}\alpha + \epsilon$$

La M-estimation des paramètres du modèle consiste à trouver :

$$\alpha_*^M = \underset{\alpha}{\operatorname{argmin}} \sum_{t=1}^n \rho(\mathbf{y}_t - \alpha_0 - \alpha_1 \mathbf{x}_{t-1} - \dots - \alpha_p \mathbf{x}_{t-p})$$

Posons $\epsilon_{*,t} = y_t - \alpha_{*,0}^M - \alpha_{*,1}^M x_{t-1} - \dots - \alpha_{*,p}^M x_{t-p}$, $t = 1, \dots, n$

Pour avoir un estimateur qui ne dépend pas de l'unité utilisée , on doit introduire un estimateur de la dispersion des résidus :

Un estimateur robuste et très populaire pour la dispersion des résidus est le **re-scales MAD** définie par :

$$\sigma_* = 1.4826 \quad MAD$$

où **MAD** signifie **Median Absolute Deviation** et est obtenue par :

$$MAD = \operatorname{median} | \epsilon_{*,t} |$$

Cet estimateur est peu sensible aux données aberrantes et possède un point de rupture de $\frac{50}{100}$ car il utilise la médiane au lieu de la moyenne.

La multiplication par le coefficient 1.4826 permet quand l'échantillon est grand et que $\epsilon_t \sim N(0, \sigma^2)$ d'avoir un estimateur robuste pour l'écart-type.

En effet :

$$P(|\epsilon_t| < MAD) \approx 0.5$$

$$P\left(\left|\frac{\epsilon_t - 0}{\sigma}\right| < \frac{MAD}{\sigma}\right) \approx 0.5$$

$$P\left(|Z| < \frac{MAD}{\sigma}\right) \approx 0.5$$

Ainsi , $\frac{MAD}{\sigma}$ est approximativement égale au quantile de la loi normale au point 0.75 noté $q(0.75)$, alors :

$$\begin{aligned}\sigma &\approx \frac{MAD}{q(0.75)} \\ &\approx 1.4826 \quad MAD\end{aligned}$$

Alors , on définit le M-estimateur de façon suivante :

$$\alpha_*^M = \underset{\alpha}{\operatorname{argmin}} \sum_{t=1}^n \rho\left(\frac{\epsilon_t}{\sigma_*}\right)$$

$$\alpha_*^M = \underset{\alpha}{\operatorname{argmin}} \sum_{t=1}^n \rho\left[\frac{1}{\sigma_*} (y_t - \alpha_0 - \alpha_1 x_{t-1} - \dots - \alpha_p x_{t-p})\right]$$

Ce qui nous ramène à la résolution du problème d'optimisation suivant :

$$\mathbf{L}(\alpha) = \underset{\alpha}{\operatorname{argmin}} \sum_{t=1}^n \rho\left[\frac{1}{\sigma_*} (\mathbf{y}_t - \alpha_0 - \alpha_1 \mathbf{x}_{t-1} - \dots - \alpha_p \mathbf{x}_{t-p})\right]$$

Comme dans la méthode des (MCO) , le calcul différentiel nous permettra encore d'avoir les estimateurs du vecteur α qui sont $\{\alpha_{*,j}^M\}_{j=0}^p$

En calculant les dérivées partielles de la fonction ρ par rapport à chaque paramètres α_t du modèle recherché et en posant :

$$\psi(u) = \frac{d\rho(u)}{du}$$

On aura un système à $p + 1$ équations :

$$\frac{\partial L(\alpha)}{\partial \alpha_0} = \sum_t \psi\left[\frac{y_t - \alpha_{*,0}^M - \alpha_{*,1}^M x_{t-1} - \dots - \alpha_{*,p}^M x_{t-p}}{\sigma_*}\right] \left(\frac{-1}{\sigma_*}\right) = 0$$

$$\frac{\partial L(\alpha)}{\partial \alpha_1} = \sum_t \psi\left[\frac{y_t - \alpha_{*,0}^M - \alpha_{*,1}^M x_{t-1} - \dots - \alpha_{*,p}^M x_{t-p}}{\sigma_*}\right] \left(\frac{-x_{t-1}}{\sigma_*}\right) = 0$$

·
·
·

$$\frac{\partial L(\alpha)}{\partial \alpha_p} = \sum_t \psi \left[\frac{y_t - \alpha_{*,0}^M - \alpha_{*,1}^M x_{t-1} - \dots - \alpha_{*,p}^M x_{t-p}}{\sigma_*} \right] \left(\frac{-x_{t-p}}{\sigma_*} \right) = 0$$

Pour résoudre ce système d'équations , Draper et Smith ont défini la fonction de poids suivante :

$$w(u) = \frac{\psi(u)}{u}$$

Soit $w_t = w \left(\frac{\epsilon_{*,t}}{\sigma_*} \right)$, $t = 1, \dots, n$ avec $w_t = 1$ si $\epsilon_{*,t} = 0$

On peut alors déduire de l'égalité suivante :

$$w_t = \frac{\psi \left(\frac{\epsilon_{*,t}}{\sigma_*} \right)}{\frac{\epsilon_{*,t}}{\sigma_*}} \implies \psi \left(\frac{\epsilon_{*,t}}{\sigma_*} \right) = \frac{w_t \epsilon_{*,t}}{\sigma_*}$$

En substituant $\psi \left(\frac{\epsilon_{*,t}}{\sigma_*} \right) = \frac{w_t \epsilon_{*,t}}{\sigma_*}$ dans le système d'équations ci-dessus , ont peut le réécrire de façon suivante :

$$\frac{\partial L(\alpha)}{\partial \alpha_0} = \sum_t \psi \left(\frac{\epsilon_{*,t}}{\sigma_*} \right) = \frac{1}{\sigma_*} \sum_t w_t \epsilon_{*,t} = 0$$

$$\frac{\partial L(\alpha)}{\partial \alpha_1} = \sum_t \psi \left(\frac{\epsilon_{*,t}}{\sigma_*} \right) x_{t-1} = \frac{1}{\sigma_*} \sum_t w_t \epsilon_{*,t} x_{t-1} = 0$$

·
·
·

$$\frac{\partial L(\alpha)}{\partial \alpha_p} = \sum_t \psi \left(\frac{\epsilon_{*,t}}{\sigma_*} \right) x_{t-p} = \frac{1}{\sigma_*} \sum_t w_t \epsilon_{*,t} x_{t-p} = 0$$

En remplaçant $\epsilon_{*,t}$ par son expression , on obtient le système suivant :

$$\begin{aligned}
& \sum_t w_t (y_t - \alpha_{*,0}^M - \alpha_{*,1}^M x_{t-1} - \dots - \alpha_{*,p}^M x_{t-p}) = 0 \\
& \sum_t x_{t-1} w_t (y_t - \alpha_{*,0}^M - \alpha_{*,1}^M x_{t-1} - \dots - \alpha_{*,p}^M x_{t-p}) = 0 \\
& \quad \cdot \\
& \quad \cdot \\
& \quad \cdot \\
& \sum_t x_{t-p} w_t (y_t - \alpha_{*,0}^M - \alpha_{*,1}^M x_{t-1} - \dots - \alpha_{*,p}^M x_{t-p}) = 0
\end{aligned}$$

Après réorganisation de ce nouveau système , on trouve les égalités :

$$\begin{aligned}
& \sum_t w_t y_t = \sum_t w_t \alpha_{*,0}^M + \sum_t w_t \alpha_{*,1}^M x_{t-1} + \dots + \sum_t w_t \alpha_{*,p}^M x_{t-p} \\
& \sum_t w_t x_{t-1} y_t = \sum_t x_{t-1} w_t \alpha_{*,0}^M + \sum_t w_t \alpha_{*,1}^M x_{t-1}^2 + \dots + \sum_t x_{t-1} w_t \alpha_{*,p}^M x_{t-p} \\
& \quad \cdot \\
& \quad \cdot \\
& \quad \cdot \\
& \sum_t w_t x_{t-p} y_t = \sum_t x_{t-p} w_t \alpha_{*,0}^M + \sum_t x_{t-p} w_t \alpha_{*,1}^M x_{t-1} + \dots + \sum_t w_t \alpha_{*,p}^M x_{t-p}^2
\end{aligned}$$

En ramenant ce système sous forme matricielle , on obtient :

$$\begin{pmatrix} \sum_t w_t y_t \\ \sum_t w_t y_t x_{t-1} \\ \cdot \\ \cdot \\ \sum_t w_t y_t x_{t-p} \end{pmatrix} = \begin{pmatrix} \sum_t w_t & \sum_t w_t x_{t-1} & \sum_t w_t x_{t-2} \dots \sum_t w_t x_{t-p} \\ \sum_t w_t x_{t-1} & \sum_t w_t x_{t-1}^2 & \sum_t w_t x_{t-2} x_{t-1} \dots \sum_t w_t x_{t-p} x_{t-1} \\ & & \cdot \\ & & \cdot \\ \sum_t w_t x_{t-p} & \sum_t w_t x_{t-1} x_{t-p} & \sum_t w_t x_{t-2} x_{t-p} \dots \sum_t w_t x_{t-p}^2 \end{pmatrix} \begin{pmatrix} \alpha_{*,0}^M \\ \alpha_{*,1}^M \\ \cdot \\ \cdot \\ \alpha_{*,p}^M \end{pmatrix}$$

Ce qui est équivalent à :

$$\mathbf{X}^t \mathbf{W} \mathbf{Y} = \mathbf{X}^t \mathbf{W} \mathbf{X} \alpha_{*,n}^M$$

Où la matrice W est une matrice diagonale $n \times n$ contenant tous les poids w_t :

$$W = \begin{pmatrix} w_1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & w_2 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & w_3 & \cdot & \cdot & 0 \\ & & & \cdot & & \\ & & & \cdot & & \\ & & & \cdot & & \\ 0 & 0 & \cdot & \cdot & \cdot & w_n \end{pmatrix}$$

Si la matrice $X^t W X$ est inversible , on en déduit que :

$$\alpha_{*,n}^M = (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W} \mathbf{Y}$$

Remarque

Le **M-estimateur** des paramètres du modèle obtenu est très similaire à celui retrouvé par la méthode des (MCO) , la seule différence est la présence d'une **matrice poids dans l'équation**.

Cette **matrice poids** permet de réduire l'influence des données aberrantes. À retenir que cette matrice dépend des résidus qui à leur tour dépendent de l'estimation des paramètres de la régression.

Algorithme : M-estimateur

Susanti , Pratiwi , Sulistijowati et Liana ont proposé l'algorithme suivant pour trouver le M-estimateur des paramètres du modèle :

itération 0

- Calcul de $\alpha_{*,n}^{M(0)}$ en utilisant la méthodes (MCO)
- Calcul des résidus $\{\epsilon_{*,t}^{(0)}\}_{t=1}^n$ en utilisant les paramètres estimés $\{\alpha_{*,j}^{M(0)}\}_{j=0}^p$
- On trouve $\sigma_*^{(0)} = 1.4826 MAD$ où MAD est obtenu avec $\{\epsilon_{*,t}^{(0)}\}_{t=1}^n$
- On choisit une **fonction poids** $w(u)$ associée à une fonction objective $\rho(u)$

- Calcul des poids $w_t^{(0)} = w\left(\frac{\epsilon_{*,t}^{(0)}}{\sigma_*^{(0)}}\right)$
- On obtient alors la matrice :

$$W^{(0)} = \begin{pmatrix} w_1^{(0)} & 0 & . & . & . & 0 \\ 0 & w_2^{(0)} & . & . & . & 0 \\ 0 & 0 & w_3^{(0)} & . & . & 0 \\ & & . & & & \\ & & . & & & \\ & & . & & & \\ 0 & 0 & . & . & . & w_n^{(0)} \end{pmatrix}$$

itération 1

- Calcul de $\alpha_{*,n}^{M^{(1)}} = (X^t W^{(0)} X)^{-1} X^t W^{(0)} Y$
- Dédution des résidus $\{\epsilon_{*,t}^{(1)}\}_{t=1}^n$ en utilisant les paramètres estimés $\{\alpha_{*,j}^{M^{(1)}}\}_{j=0}^p$
- On trouve $\sigma_*^{(1)} = 1.4826MAD$ où MAD est obtenu avec $\{\epsilon_{*,t}^{(1)}\}_{t=1}^n$
- Calcul des poids $w_t^{(1)} = w\left(\frac{\epsilon_{*,t}^{(1)}}{\sigma_*^{(1)}}\right)$
- On trouve alors la matrice :

$$W^{(1)} = \begin{pmatrix} w_1^{(1)} & 0 & . & . & . & 0 \\ 0 & w_2^{(1)} & . & . & . & 0 \\ 0 & 0 & w_3^{(1)} & . & . & 0 \\ & & . & & & \\ & & . & & & \\ & & . & & & \\ 0 & 0 & . & . & . & w_n^{(1)} \end{pmatrix}$$

itération L

- Calcul de $\alpha_{*,n}^{M^{(L)}} = (X^t W^{(L-1)} X)^{-1} X^t W^{(L-1)} Y$

- Dédution des résidus $\{\epsilon_{*,t}^{(L)}\}_{t=1}^n$ en utilisant les paramètres estimés $\{\alpha_{*,j}^{M^{(L)}}\}_{j=0}^p$
- On trouve $\sigma_*^{(L)} = 1.4826MAD$ où MAD est obtenu avec $\{\epsilon_{*,t}^{(L-1)}\}_{t=1}^n$
- Calcul des poids $w_t^{(L)} = w\left(\frac{\epsilon_{*,t}^{(L)}}{\sigma_*^{(L)}}\right)$
- On trouve alors la matrice :

$$W^{(L)} = \begin{pmatrix} w_1^{(L)} & 0 & . & . & . & 0 \\ 0 & w_2^{(L)} & . & . & . & 0 \\ 0 & 0 & w_3^{(L)} & . & . & 0 \\ & & . & & & \\ & & . & & & \\ & & . & & & \\ 0 & 0 & . & . & . & w_n^{(L)} \end{pmatrix}$$

Critère d'arrêt

Le processus s'arrête à la L^{ime} itération si :

$$\frac{\|\alpha_{*,\mathbf{n}}^{\mathbf{M}^{(L)}} - \alpha_{*,\mathbf{n}}^{\mathbf{M}^{(L-1)}}\|_2}{\|\alpha_{*,\mathbf{n}}^{\mathbf{M}^{(L)}}\|_2} < \varepsilon$$

Où ε est un très petit réel positif fixé au départ , soit $\varepsilon = 0.0001$.
 Dans le logiciel R , la fonction **rlm** utilise le pourcentage de changement entre les résidus de l'itération L et ceux de l'itération $(L - 1)$ pour un ε fixé:

$$\frac{\|\epsilon_{*,\mathbf{n}}^{\mathbf{M}^{(L)}} - \epsilon_{*,\mathbf{n}}^{\mathbf{M}^{(L-1)}}\|_2}{\|\epsilon_{*,\mathbf{n}}^{\mathbf{M}^{(L)}}\|_2} < \varepsilon$$

Convergence

Jureckova , Sen et Picek ont montré que le **M-estimateur** des paramètres du modèle obtenu converge en distribution :

$$\sqrt{\mathbf{n}}(\alpha_{*,\mathbf{n}}^{\mathbf{M}} - \alpha) \xrightarrow{d} \mathbf{N}\left(\mathbf{0}, \sigma^2 \frac{\mathbf{Q}^{-1}}{e}\right)$$

$$e = \frac{[\int \psi'(x)\phi(x)dx]^2}{\int \psi^2(x)\phi(x)dx}$$

$$Q = \lim \left(\frac{1}{n} X^t X \right), n \rightarrow \infty$$

Pour plus de précision , voir [JSP13] pages [215 à 224]

5.2 Application sur les données de ventes

Le programme suivant résume le résultat de l'algorithme de la **M-estimation** en utilisant les données de ventes précédentes :

```
> library(MASS)
Warning message:
le package 'MASS' a été compilé avec la version R 3.6.3
> ar=rlm(Ventes~.,data=drv,psi=psi.huber)
> weights(ar)
[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
> summary(ar)

Call: rlm(formula = Ventes ~ ., data = drv, psi = psi.huber)
Residuals:
    Min       1Q   Median       3Q      Max
-2982.80  -418.99   80.32   479.12  2374.79

Coefficients:
            Value      Std. Error t value
(Intercept) -23542.8948  17399.0355  -1.3531
Prix         -4921.4280   363.2288  -13.5491
PubTV         648.8124   144.2798    4.4969
PubInt       1760.2156   276.2465    6.3719

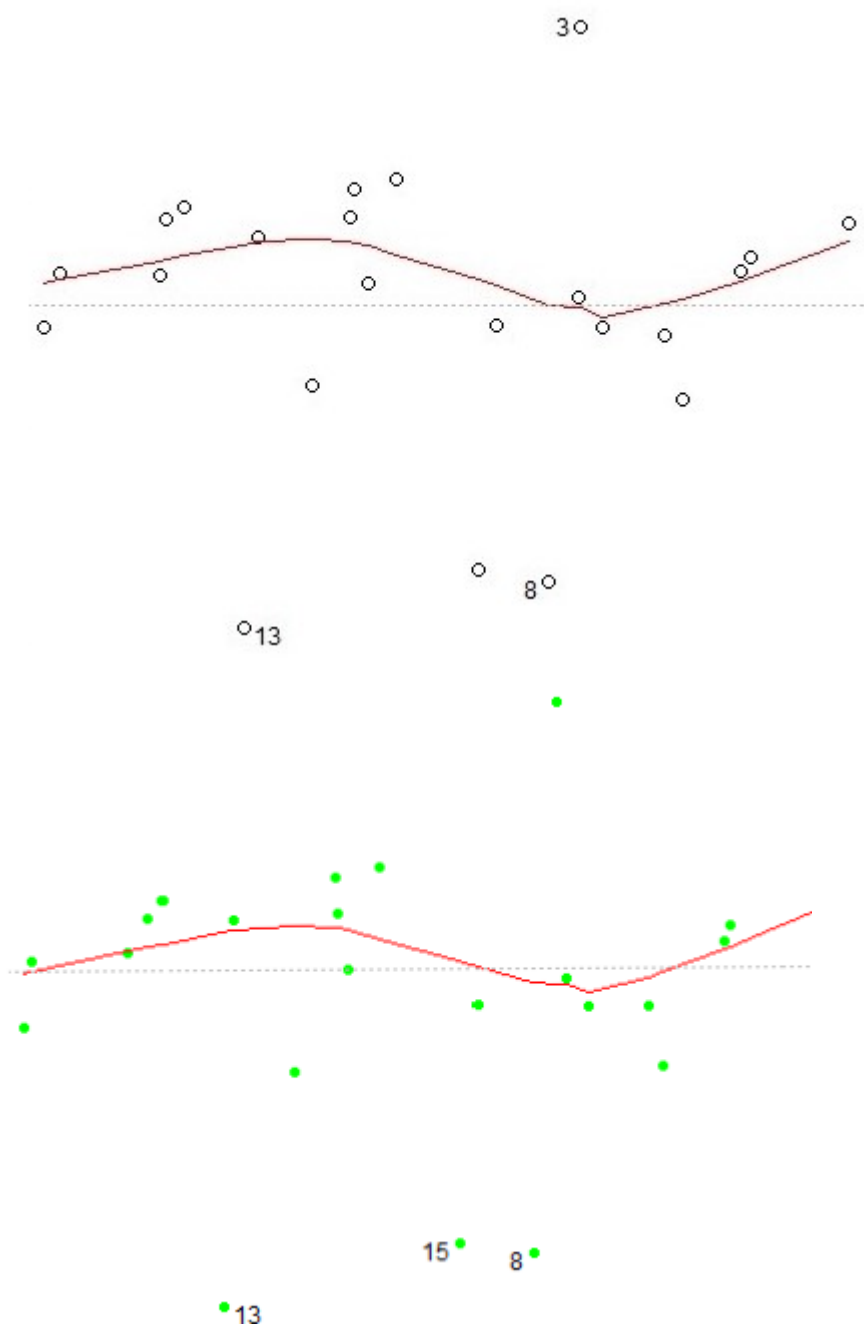
Residual standard error: 733.8 on 19 degrees of freedom
```

Le **M-estimateur** de **Huber** obtenu est donc :

$$\alpha_{*,n}^M = \begin{pmatrix} -23542.894 \\ -4921.4280 \\ 648.8124 \\ 1760.2156 \end{pmatrix}$$

Cet estimateur est plus robuste que celui des (MCO) car il pénalisent les grandes valeurs résiduelles en leur donnant le moins de poids possible , ce qui permettra de réduire leur influence sur le modèle.

Les graphiques ci dessous représentent les courbes de sensibilité respectives de la méthode des (MCO) et celle de la M-estimation :

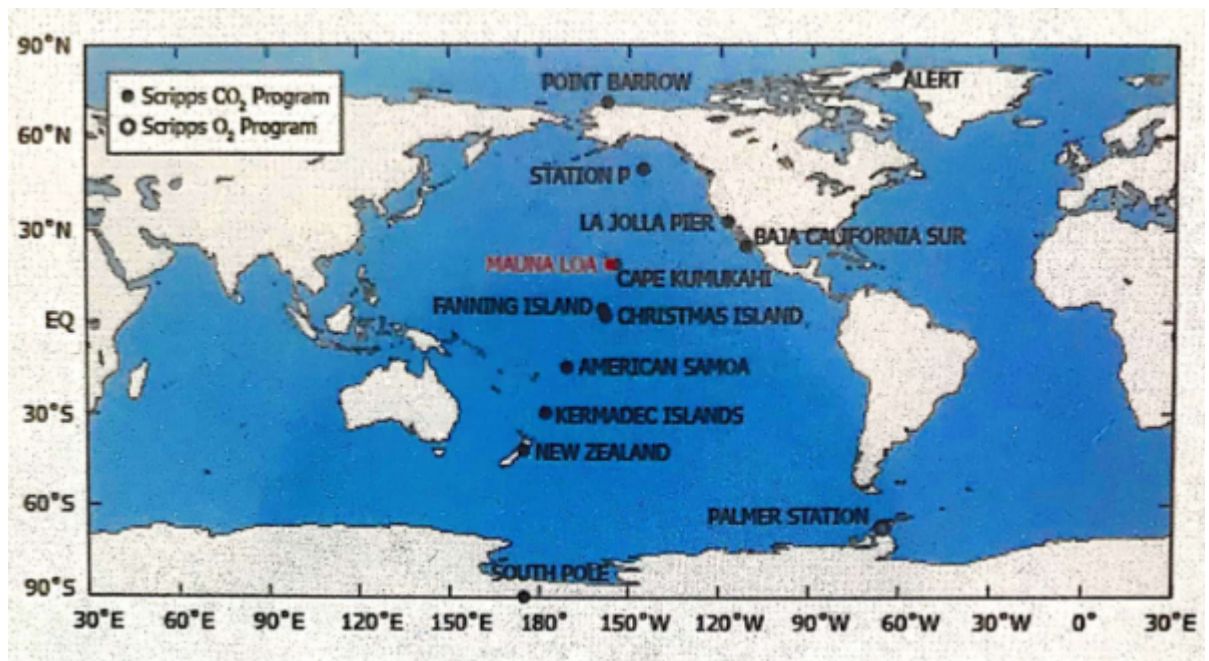


6 Mise en oeuvre pratique avec R

6.1 Étude de la concentration de CO_2

Le programme **Scripps CO_2** mesure les changements dans les niveaux de dioxyde de carbone atmosphérique dans certaines zones géographiques à partir d'échantillons d'air prélevés dans des stations du monde entier ; ce réseau d'échantillonnage donne une perspective sur la variabilité du CO_2 contenu dans l'air.

Dans notre étude , nous utiliserons les données provenant d'une seule station nommée Mauna Loa situé aux États Unis , cette station réalise des prélèvements depuis 1958 et les dernières mesures datent de Mai dernier.



Indication

Le fichier de données est téléchargeable via le lien :

<http://scripps.co2.ucsd.edu>

Les données contiennent des valeurs manquantes notées par -99.99

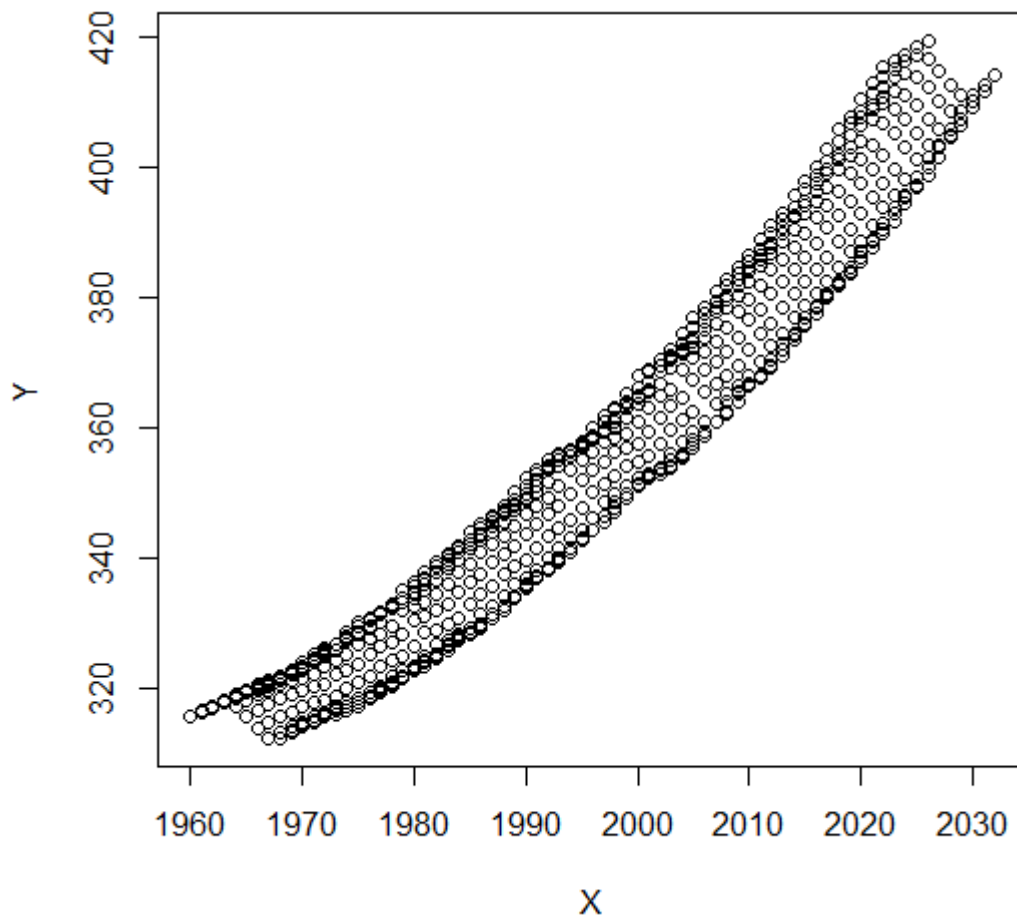
6.2 Programme R

Chargement des données

```
x = read.table("CO2.txt", header = TRUE)
attach(x)
names(x)
```

Visualisation des données

```
X = Mn + Yr
Y = CO2
plot(X, Y)
```

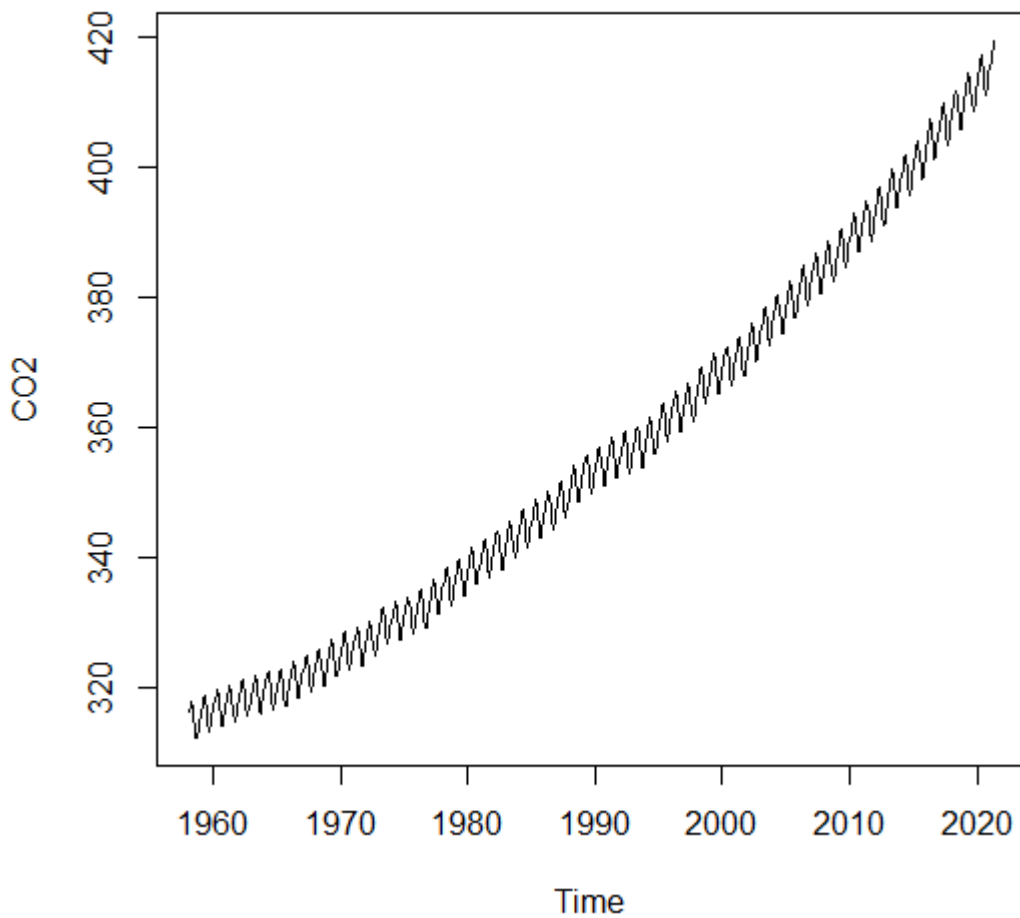


On constate une forte suspicion d'hétéroscédasticité , ie que la variance de Y dépend de X , ce qui implique que la variance des résidus n'est pas finie comme le suppose la méthode des (MCO).

Création de la série chronologique

```
y = ts(CO2, start = c(1958, 3), end = c(2021, 5), freq = 12)  
plot(y, xlab = "Time", ylab = "CO2", main = "évolution du CO2 en ppm")
```

évolution du CO2 en ppm



La concentration de CO_2 contenu dans les échantillons d'air augmente au fil du temps, on peut voir que la série chronologique est affectée d'une tendance linéaire et de fortes variations saisonnières plus un terme d'erreurs aléatoires. Le modèle initial s'écrit alors sous la forme :

$$Y_t = T_t + S_t + \epsilon_t$$

Test de Dickey-Fuller

```
> library(tseries)
Registered S3 method overwritten by 'quantmod':
  method      from
as.zoo.data.frame zoo

'tseries' version: 0.10-48

'tseries' is a package for time series analysis and computational
finance.

See 'library(help="tseries")' for details.

Warning message:
le package 'tseries' a été compilé avec la version R 3.6.3

> adf.test(y)

      Augmented Dickey-Fuller Test

data:  y
Dickey-Fuller = 0.92858, Lag order = 9, p-value = 0.99
alternative hypothesis: stationary

Warning message:
In adf.test(y) : p-value greater than printed p-value
```

Le test de Dickey-Fuller donne une p-value supérieur à $5^0/0$, ce qui signifie que l'hypothèse alternative de stationnarité a été rejetée , alors on conclut que la chronique n'est pas stationnaire.

On sait que les processus autorégressifs sont des modèles de séries chronologiques stationnaires , alors avant toute modélisation , il faudra stationnariser la série.

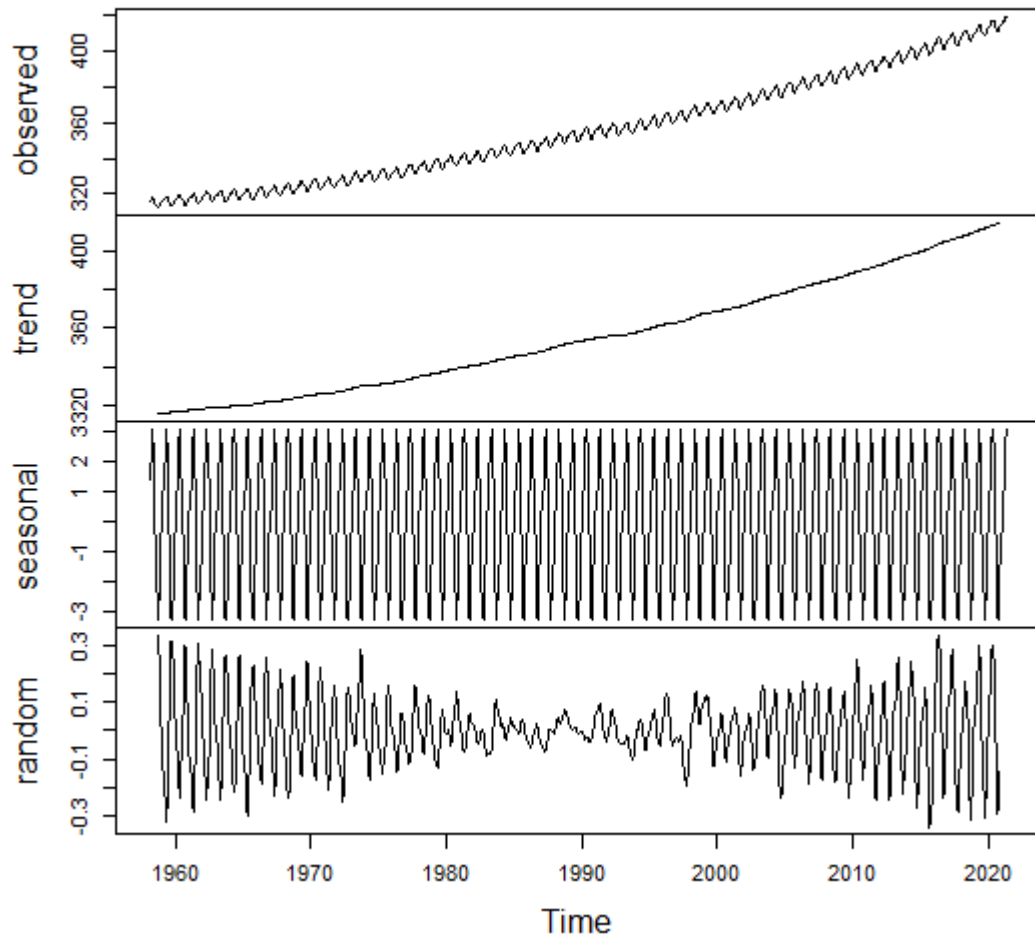
Remarque

Avant de passer à la stationnarisation , nous allons commencer par une représentation graphique des différentes composantes de la série.

Décomposition de la série

`plot(decompose(y))`

Decomposition of additive time series



La série temporelle débarrassée de toute tendance nous donne un **processus bruit blanc** stationnaire avec mouvement saisonnier , ainsi le modèle devient :

$$\epsilon_t + S_t = Y_t - T_t$$


```

> summary(ar)

Call: rlm(formula = CO2 ~ ., data = x, psi = psi.huber)
Residuals:
    Min       1Q   Median       3Q      Max
-7.0618 -2.9890 -0.7755  3.0505 12.8753

Coefficients:
            Value      Std. Error t value
(Intercept) -2793.3116     17.4417  -160.1517
Yr           1.5839       0.0088   180.7370
Mn          -0.2637       0.0464   -5.6832

Residual standard error: 4.477 on 756 degrees of freedom

```

Une estimation robuste de la tendance globale de la chronique est de la forme :

$$\mathbf{T}_{*,t} = 1.5839\mathbf{Yr} - 0.2637\mathbf{Mn} - 2793.3116$$

Après avoir estimé la tendance, l'idée est de la retirer du modèle pour obtenir un modèle de séries chronologiques stationnaire.

Stationnarisation

Nous pouvons à présent enlever cette tendance en effectuant une opération de différenciation sur la série temporelle ; a priori, la seconde différence de la série brute avec un décalage égal à sa période supprime l'effet de toute tendance et de mouvements saisonniers dans la chronique, théoriquement on a :

$$\epsilon_t = \mathbf{Y}_t - \mathbf{T}_t - \mathbf{S}_t$$

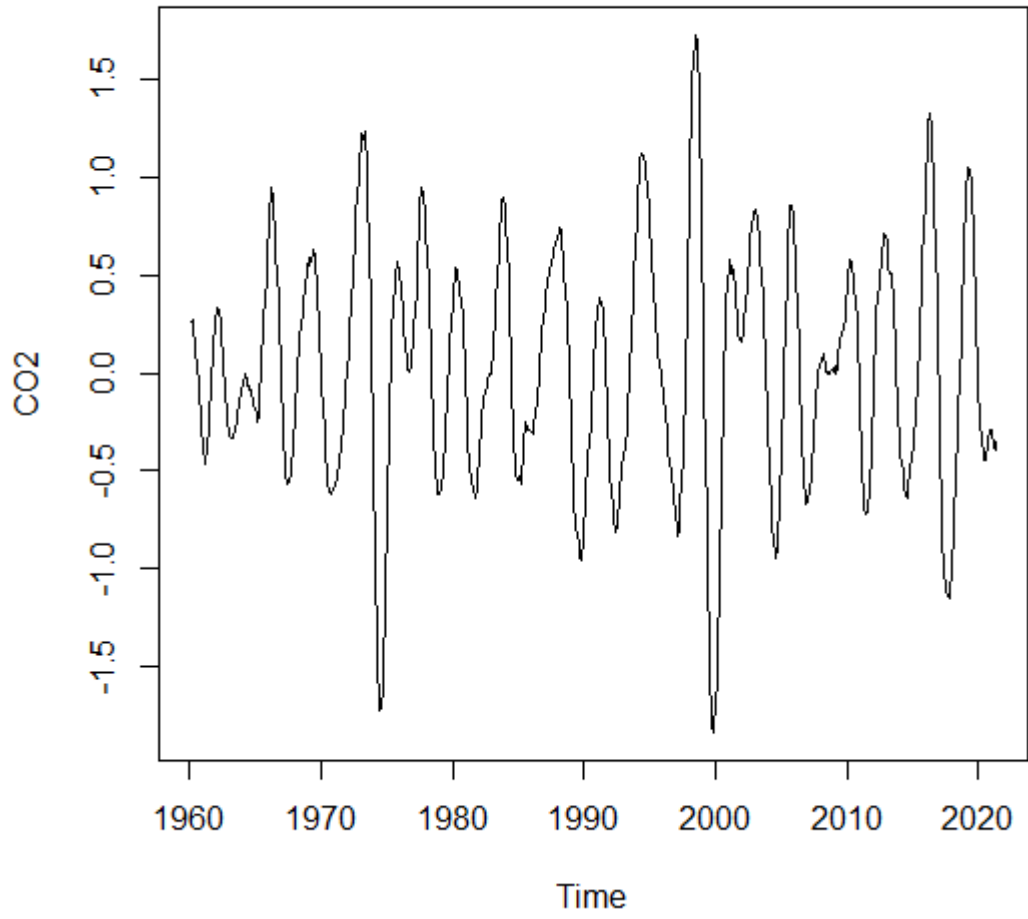
En pratique, cela donne les résultats suivants :

```

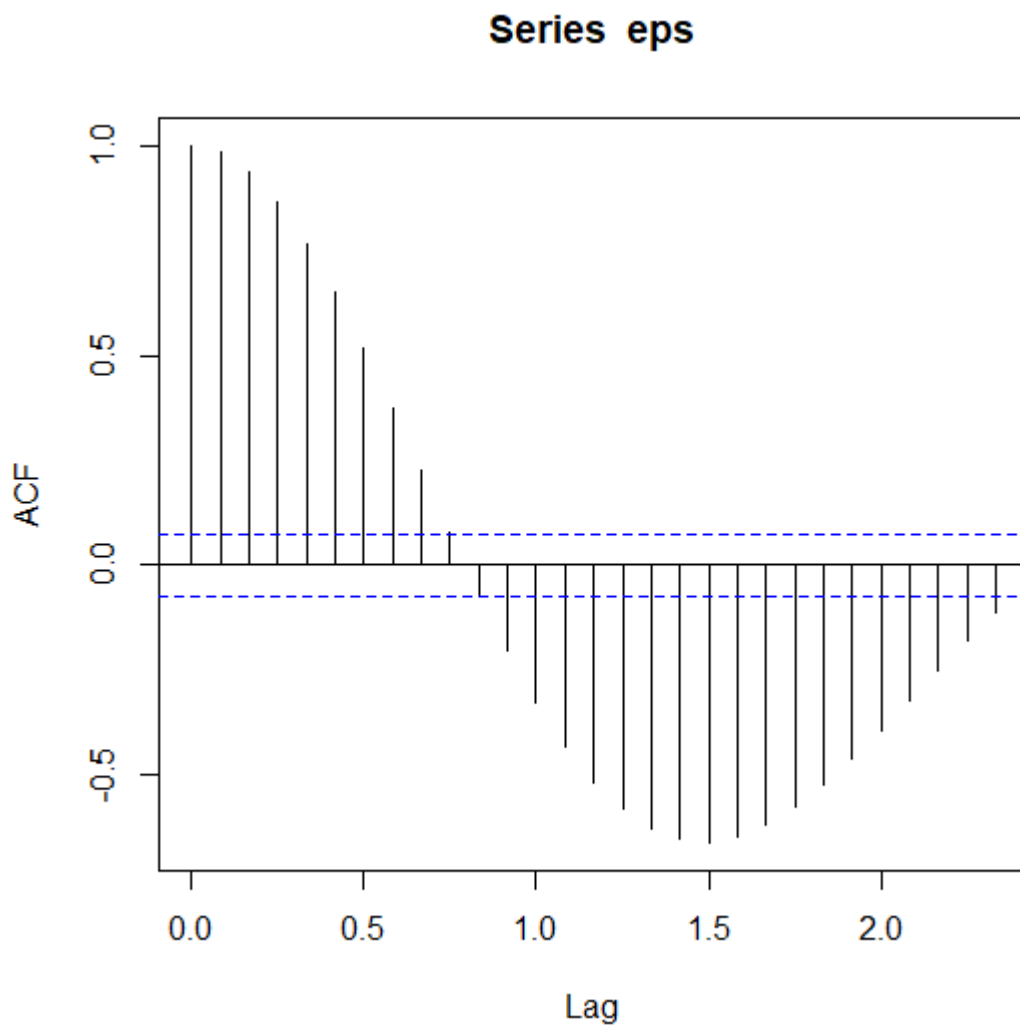
eps=diff(y,differences=2,lag=12)
plot(eps,ylab="CO2",main="série différenciée")

```

série différenciée

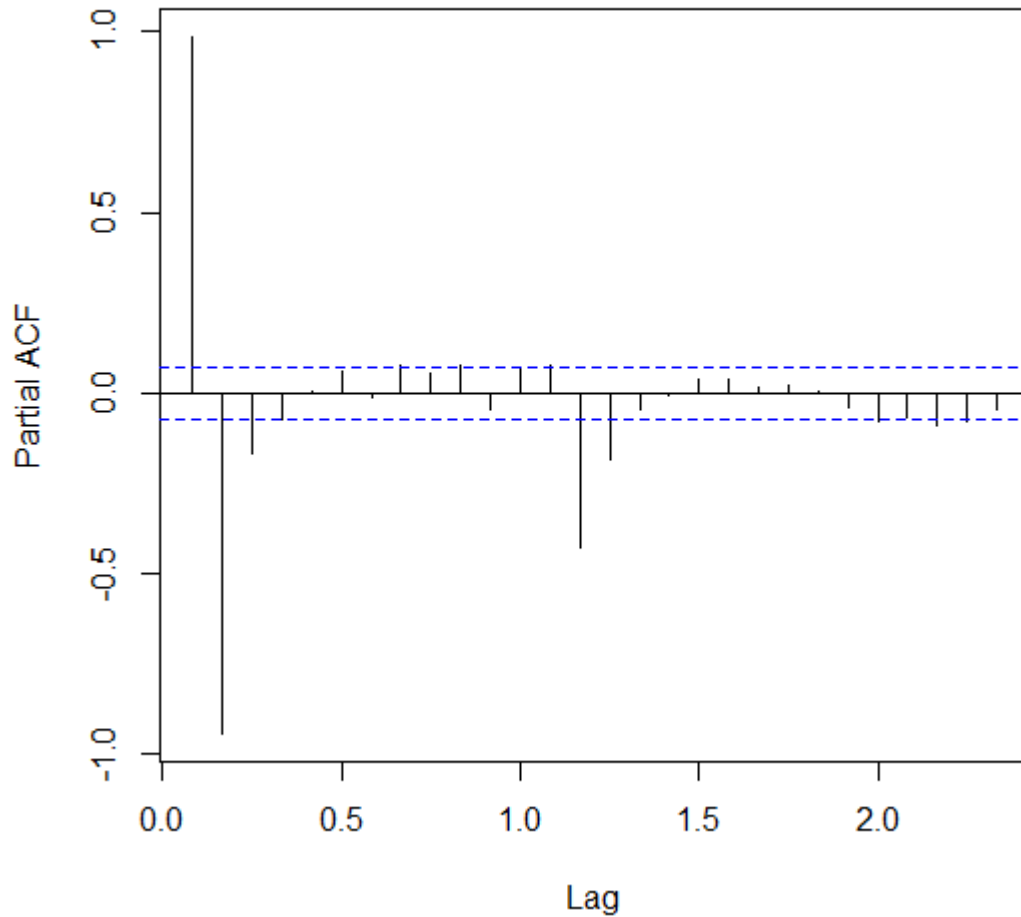


Après stationnarisation de la série , nous pouvons maintenant estimer le modèle autorégressif qui s'ajuste du mieux possible avec la nouvelle série stationnaire. Pour cela , nous allons utiliser les fonctions d'autocorrélation et d'autocorrélation partielle :



Le graphe de l'ACF (**Auto Correlation Function**) décroît , ce qui montre que la série est corrélée avec ses observations passées , alors on doit estimer un processus bruit blanc autorégressif. Pour déterminer l'ordre du processus , il suffit de passer à la PACF (**Partial Auto Correlation Function**) :

Series eps



La PACF montre que les 2 premières observations passées du bruit blanc sont fortement corrélées avec le processus, ie que la série différenciée à l'instant t est corrélée avec la série aux instants $t - 1$ et $t - 2$.

Alors on doit estimer un processus bruit blanc autorégressif d'ordre 2, pour estimer ce modèle, nous allons utiliser la commande **arima** du logiciel R qui donne une approximation des paramètres autorégressifs du modèle.

```

> arima(eps,order=c(2,0,0))

Call:
arima(x = eps, order = c(2, 0, 0))

Coefficients:
      ar1      ar2  intercept
  1.9277 -0.9585   0.0254
s.e.  0.0100  0.0100   0.0352

sigma^2 estimated as 0.0008698:  log likelihood = 1542.72,  aic = -3077.44

> res=residuals(arima(eps,order=c(2,0,0)))
> Box.test(res,type="Ljung-Box",lag=12)

```

Box-Ljung test

```

data:  res
X-squared = 327.09, df = 12, p-value < 2.2e-16

```

```

> shapiro.test(res)

```

Shapiro-Wilk normality test

```

data:  res
W = 0.9949, p-value = 0.0149

```

Le test de non autocorrélation montre une p-value inférieure à 5⁰/0 , ce qui signifie que les résidus sont corrélés.

Le test de Shapiro donne une p-value égale à 0.01 qui est inférieure à 5⁰/0

Alors les nouveaux résidus du processus ne sont pas gaussiens , ce qui est logique car ces nouveaux résidus sont issus d'un échantillon de résidus hétéroscédastiques.

```

> t.test(res)

```

One Sample t-test

```

data:  res
t = 0.0052999, df = 734, p-value = 0.9958
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.002131283  0.002142821
sample estimates:
 mean of x
5.769211e-06

```

Les résidus sont de moyenne nulle.

Ainsi , le modèle retenu est un **processus bruit blanc autorégressif d'ordre 2** ayant pour expression :

$$\epsilon_t = 0.0254 + 1.9277 \times \epsilon_{t-1} - 0.9585 \times \epsilon_{t-2} + \varepsilon_t$$

Avec $\sigma_{*,\epsilon}^2 = 0.0008698$ et ε_t un bruit blanc quelconque **autocorrélé**.

CONCLUSION GÉNÉRALE

Lorsque nous disposons d'un échantillon de données d'un modèle autorégressif , l'estimateur des (MCO) est le meilleur estimateur linéaire possible pour ce modèle ; ensuite nous avons vu que si une modification intervient dans ces données , ce qui est souvent le cas , cet estimateur est beaucoup plus affecté par cette perturbation.

Pour avoir un bon estimateur qui prend en compte cette modification dans l'échantillon , le M-estimateur de Huber est un **estimateur robuste** qui est beaucoup moins sensible face à cette perturbation.

Pour une analyse plus générale , nous avons estimé le modèle autorégressif provenant des données réelles sur la concentration de dioxyde de carbone présent dans l'air , qui est un modèle de séries temporelles stationnaires.

Ce modèle peut éventuellement être soumis à d'autres **perspectives** de travail comme par exemple l'**estimation robuste des paramètres obtenus** ou l'**évolution future du taux de CO_2**

Références

- 1) Mourid Tahar , sur l'identification d'un processus autorégressif
- 2) Jérémie BIGOT , Université de Bordeaux , L3 MIASHS
- 3) Nita HADJER , Statistiques robustes
- 4) Simard JOANIE , Méthodes de régression robuste
- 5) Frédéric Proia , Autocorrélation et stationnarité dans le processus autorégressif
- 6) Agnès LAGNOUX , Séries chronologiques
- 7) V.Monbet , Modélisation de séries temporelles
- 8) Gilbert Colletaz , les processus AR
- 9) Clément Rau , ajustement linéaire par les moindres carrés
- 10) Ndèye Niang et Gilbert Saporta , régression robuste
- 11) Donoho D , Huber P.J(1983) The notion of breakdown point
- 12) J.Jureckova , P.K.Sen , and J.Picek Methodology in Robust and Non parametric Statistics , CRC Press,2013