

Département de Mathématiques
Faculté de Sciences Exactes
Université Frères Mentouri, Constantine 1

Polycopié de MESC



Ce cours est présenté par

Lescheb Ines

aux étudiants de **Master I** spécialité : **M**athématiques **A**ppiquées à l'**E**conomie et à la **F**inance
MAEF

Module intitulé : Modèles Econométriques et Séries Chronologiques

Année universitaire **2020-2021**



Lescheb Ines

Maitre de Conférences A

Table des matières

Avant-propos	v
Introduction	vii
1 Séries chronologiques : définitions et exemples	1
1.1 Objectifs principaux	3
1.2 Modélisation	3
1.3 Tendance et saisonnalités	4
1.4 Elimination de la tendance en l'absence de la saisonnalité	5
1.4.1 Estimation paramétrique	5
1.4.2 Estimation non paramétrique	7
1.4.3 Méthode des différences itérés	9
1.5 Elimination de la tendance et de la saisonnalité	11
1.5.1 Méthode des différences	13
2 Processus Stationnaire	14
2.1 Fonctions Caractéristiques	15
2.1.1 Fonction d'autocovariance	15
2.1.2 Fonction d'autocorrélation (ACF)	15
2.1.3 Fonction d'autocorrélation partielle (PACF)	16
2.2 Stationnarité	17
2.2.1 Définitions et exemples	17
2.2.2 Bruit blanc	18

2.2.3	Exemple d'un processus non stationnaire : Marche aléatoire	18
3	Processus linéaires	21
3.1	Processus linéaires	21
3.2	Processus Moyennes Mobiles	24
3.3	Fonctions caractéristiques d'un processus MA(q)	25
3.4	Stationnarité et causalité	33
3.5	Inversibilité	33
3.6	Processus Autorégressifs	36
3.7	Fonctions caractéristiques d'un AR(p)	36
3.7.1	Fonction d'autocovariance	37
3.7.2	Fonctions d'autocorrélation et d'autocorrélation partielle	37
3.8	Stationnarité et causalité	46
3.9	Inversibilité	50
4	Processus mixtes : ARMA	52
4.1	Fonctions caractéristiques	54
4.1.1	Fonction d'autocovariance	54
4.1.2	Densité spectrale d'un processus ARMA	55
4.1.3	Fonction d'autocorrélation	55
4.1.4	Fonction d'autocorrélation partielle	56
4.1.5	Fonctions caractéristiques d'un ARMA(1,1)	57
4.2	Stationnarité et causalité	60
4.2.1	Stationnarité d'un processus ARMA(p,q)	60
4.2.2	Causalité	61
4.2.3	Inversibilité	62
5	Estimation des paramètres des processus MA, AR et ARMA	64
5.1	Estimation des paramètres d'un processus MA	

5.1.1	Estimation par la méthode du maximum de vraisemblance	64
5.1.2	Comportement asymptotique	65
5.1.3	Estimation par la méthode des moindres carrés	66
5.1.4	Comportement asymptotique	67
5.1.5	Normalité asymptotique des différentes méthodes	68
5.2	Simulation avec le modèle $MA(2)$	69
5.2.1	Résultats et commentaires	77
5.3	Estimation des paramètres d'un processus autorégressifs	78
5.3.1	Méthode de Yule-Walker	78
5.3.2	Méthode du maximum de vraisemblance	81
5.3.3	Méthode des moindres carrés	82
5.3.4	Normalité asymptotique des différentes méthodes	82
5.4	Simulation avec le modèle $AR(2)$	83
5.4.1	Résultats et commentaires	87
5.5	Estimation des paramètres d'un ARMA	88
5.5.1	Estimation par la méthode du maximum de vraisemblance	88
5.5.2	Comportement asymptotique	90
5.5.3	Estimation par la méthode des moindres carrés ordinaires	92
5.5.4	Comportement asymptotique	93
5.5.5	Comparaison des différentes méthodes d'inférence	93
5.6	Simulation avec le modèle $ARMA(1,1)$	94
5.6.1	Résultats et commentaires	103
6	Modèles (G)ARCH	104
6.1	Généralité sur les modèles GARCH(p,q)	105
6.2	Etude de Stationnarité	107
6.2.1	Cas du modèle GARCH(1,1)	107
6.2.2	Stationnarité stricte	108

6.2.3	Stationnarité au second ordre	111
6.3	Stationnarité du modèle GARCH(p,q)	113
6.3.1	Stationnarité stricte	113
6.3.2	Stationnarité au second ordre	116
6.3.3	Région de stationnarité	118
6.4	Estimation des paramètres du modèle GARCH	119
6.4.1	Méthode de maximum de vraisemblance	120
6.5	Simulation	122
6.5.1	Simulation d'un GARCH(1,1)	122
6.5.2	Script R	122
6.5.3	Script R	125
6.5.4	Conclusion	128

Avant-propos

Dans le présent cours, j'étudie quelques éléments de séries chronologiques et quelques modèles linéaires, conformément aux programmes de master I de mathématiques spécialité " Mathématiques Appliquée à l'Économie et à la Finance". Je présente la modélisation de séries chronologiques, je discute comment modéliser et estimer les composantes non stationnaires d'un modèle classique de décomposition d'une série de temps. Dans le deuxième chapitre, j'intéresse à modéliser la composante stationnaire, c'est-à-dire, je fournis une série de définitions de base qui caractérisent les processus stationnaires. J'étudie également dans le troisième et le quatrième chapitres, les processus linéaires ainsi que trois cas particuliers de tels processus : les processus AR , les processus MA et les processus mixtes $ARMA$. Je donne d'une façon bien détaillée les définitions et les fonctions caractéristiques de chaque processus ainsi que une étude probabiliste est faite pour chaque modèles, en donnant les conditions nécessaires et suffisantes assurant la stationnarité, la causalité et l'inversibilité de ces trois processus. Le chapitre cinq présente les différentes méthodes d'estimation, en particulier, la méthode du maximum de vraisemblance, méthode de Yule-Walker et la méthode des moindres carrés. Ces méthodes d'estimation sont utilisées pour estimer les paramètres des trois modèles étudiés. Quant au dernier chapitre, il est consacré aux modèles non linéaires $ARCH$ et de sa généralisation naturelle : la classe des processus $GARCH$. Alors dans ce chapitre j'ai présenté d'abord la définition et la représentation des modèles $GARCH$. En suite, j'établis la condition de sta-

tionnarité strict et au second-ordre, à partir du modèle *GARCH* de premier ordre, car les preuves sont plus facile et les résultats sont plus explicites. Ensuite, j'étudie le problème de l'estimation et de l'ajustement d'un modèle de type *(G)ARCH*, où l'estimation par maximum de vraisemblance jouera un rôle prépondérant.

Introduction

Dans ce premier chapitre introductif, nous allons discuter dans un premier temps, sur un niveau tout d'abord descriptif, du cadre général de la modélisation de séries chronologiques. Ceci se fait surtout à l'aide de plusieurs exemples typiques. Après avoir donné la base pour la suite de notre analyse de séries temporelles, nous discuterons comment modéliser et estimer les composantes non stationnaires d'un modèle classique de décomposition d'une série de temps. Ici, l'idée est de séparer la tendance générale, ainsi qu'un comportement saisonnier éventuel des données, d'une partie stationnaire qui inclurera toutes les petites fluctuations aléatoires autour de ces deux composantes dépendant explicitement du temps. La stationnarité, et le concept de série stationnaire, seront définis et discutés dans le deuxième chapitre et signifient que, en moyenne, les caractéristiques de la série (sa tendance et sa variation) ne changent pas avec le temps. Toutefois, le but de ce chapitre sera de comprendre pourquoi c'est surtout cette dernière partie stationnaire qui nous intéresse le plus quant à la modélisation d'une structure de notre série observée une structure qui est susceptible de trouver un bon modèle ajusté pour effectuer le but ultime de la prévision des séries temporelles.

Chapitre 1

Séries chronologiques : définitions et exemples

La théorie des séries temporelles (ou chronologiques) abordée dans ce cours d'introduction est appliquée de nos jours dans des domaines aussi divers que l'économétrie, la médecine ou la démographie, pour n'en citer qu'une petite partie.

Définition 1 Une série chronologique est une suite formée d'observations au cours du temps. On peut songer par exemple à l'évolution du nombre de voyageurs utilisant le train, à l'accroissement relatif mensuel de l'indice des prix ou encore à l'occurrence d'un phénomène naturel (comme le nombre de taches solaires).

La définition mathématique adéquate pour l'étude et la prévision de telles suites consiste à les considérer comme une réalisation particulière d'une famille de variables aléatoires $(X_t, t \in I)$ définies sur un espace de probabilité.

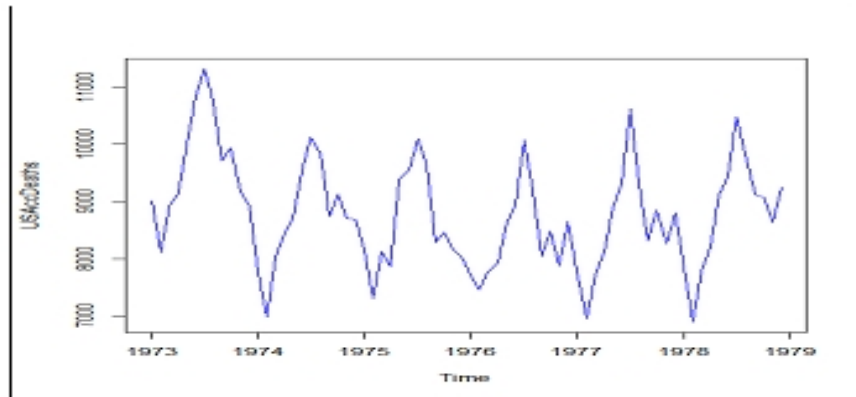
Dans la suite de ce cours, nous les noterons :

$$(X_t, t \in I) \text{ ou } (X_t)_{t \in I}$$

où I est un intervalle de temps qui peut être discret ou continu.

- **Script R**

```
data()  
USAccDeaths  
plot(USAccDeaths)
```



Nombre de morts accidentels aux Etas-Unis

- **Script R**

```
pib <- read.table("http://hamrita.e-monsite.com/medias/files/tunpib.txt",
skip = 2)
```

```
export <- read.table("http://hamrita.e-monsite.com/medias/files
/exportation.txt",skip = 2)
```

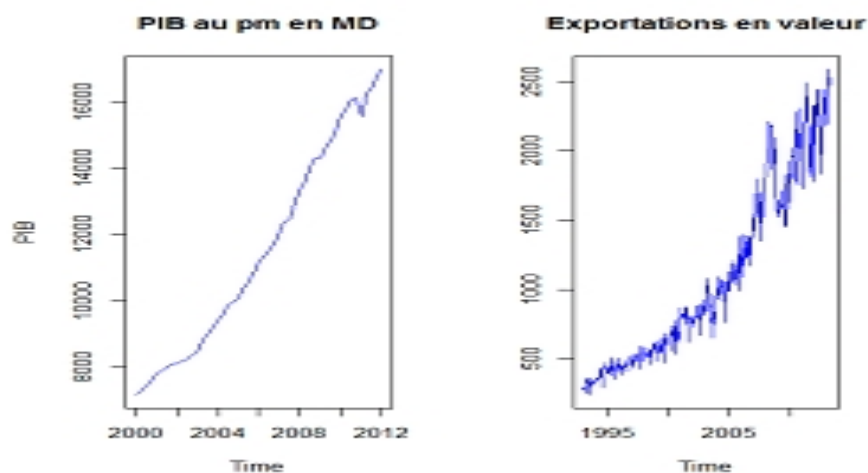
```
pib <- ts(pib, start = 2000, freq = 4)
```

```
Export <- ts(export, star = c(1993, 1), freq = 12)
```

```
par(mfrow = c(1, 2))
```

```
plot(pib, ylab = "PIB",col="blue", main = "PIB au pm en MD")
```

```
plot(Export, ylab = "",col="blue", main = "Exportations en valeur")
```



1.1 Objectifs principaux

Nous pouvons déjà poser quelques problèmes à résoudre lorsqu'on observe une série chronologique.

Tout d'abord, la prévision consiste à prévoir les valeurs futures X_{T+h} , où $h = 1, 2, 3, \dots$, de la série chronologique à partir de ses valeurs observées jusqu'au temps T , c'est-à-dire, X_1, \dots, X_T . La prédiction de la série chronologique au temps $T + h$ est notée $\hat{X}_T(h)$ et, en général, est différente de la valeur réelle X_{T+h} que prend la série en $T + h$. Pour mesurer cette différence, on définira l'erreur de prédiction par la différence $\hat{X}_{T+h} - X_{T+h}$ "en moyenne" avec l'idée que plus h est grand, plus que l'erreur est grande. L'intervalle de précision, défini par les deux valeurs $\hat{X}_T^{(1)}(h)$ et $\hat{X}_T^{(2)}(h)$, est susceptible de contenir la valeur inconnue de X_{T+h} . La qualité de la prédiction pourra être mesurée en se basant sur 80% des observations, puis en simulant une prédiction sur les 20% d'observations restantes. Cette technique est aussi utile pour :

- Les séries qui contiennent des "trous".
- Mesurer l'effet d'un phénomène accidentel (erreur, ...).

Un autre problème intéressant est la détection de ruptures résultantes, par exemple, d'un changement de politique (économique). Ces ruptures peuvent être de deux ordres :

- Une rupture de niveau.
- Une rupture de pente.

La prévision de ces dates de rupture est bien évidemment très importante.

Il faudra également éliminer la tendance (ou trend) représentant l'évolution à moyen terme du phénomène étudié. Cette tendance agit comme une forte corrélation entre les variables X_t qui n'exprime aucune liaison à caractère explicatif. Nous verrons comment enlever cette tendance pour voir si de telles liaisons existent afin de n'étudier que les corrélations sans tendance qui sont les quantités qui nous intéressent pour expliquer un phénomène observé.

1.2 Modélisation

Un modèle est une image simplifiée de la réalité. Un modèle peut être meilleur qu'un autre pour décrire la réalité. Plusieurs questions se posent tout de suite : Comment mesurer cette qualité ? Comment diagnostiquer un modèle ?

Nous allons traiter, bien sûr, ces questions au fur et à mesure du développement de plusieurs modèles particuliers. mais à présent nous voulons simplement donner une petite liste qui sert à résumer et classier les modèles à développer dans notre traitement.

Dans le traitement des modèles d'ajustement, il faut d'abord parler du "modèle classique de décomposition". En effet

$$X_t = m_t + s_t + \epsilon_t \quad (1.1)$$

où les observations X_t sont modélisées comme superposition additive d'une tendance déterministe m_t , une saisonnalité (déterministe) s_t et une perturbation aléatoire ϵ_t qui représente les composantes ("erreurs") non systématiques, de moyenne nulle, mais qui possède une structure de corrélation non nulle. (Ce sera justement cette structure qui nous intéressera le plus par rapport au développement d'un bon modèle à ajuster pour réaliser notre but ultime, la prévision).

1.3 Tendances et saisonnalités

Lorsqu'on souhaite analyser une série chronologique présentée sous la forme d'un tableau de chiffres, il faut tout d'abord représenter la série sous la forme d'un graphe. Dans les représentations des figures précédentes, on peut observer dans certains cas une composante déterministe qui se présente sous la forme d'une tendance ou d'un cycle saisonnier. Une série chronologique peut également présenter ces deux comportements en même temps.

Bien entendu, cette tendance ou cette saisonnalité ne sont pas toujours faciles à déceler. Il peut même arriver que des séries ne contiennent aucune de ces composantes déterministes. Il se peut aussi que la tendance n'apparaisse qu'après avoir transformé les données par une fonction, par exemple, logarithmique.

Nous supposons que toute série chronologique $(X_t, t \in T)$ puisse être mise sous la forme de la décomposition vue déjà dans (1.1)

$$X_t = m_t + s_t + \epsilon_t$$

où :

- m_t est appelé tendance (ou trend) et est soit une fonction décrite par un nombre fini de paramètres, par exemple, une fonction linéaire du temps $a + bt$ ou un polynôme en t , soit, plus généralement, une fonction lisse en temps t .

• s_t est appelé saisonnalité et est une fonction périodique. Soit d la période de s_t , on a alors : $s_t = s_{t+d}$ pour tout t . On supposera également que s_t ne contient pas de tendance, ce qu'on écrira :

$$\sum_{i=1}^d s_{t+i} = 0.$$

En d'autres termes, toute la tendance est modélisée dans m_t .

• $m_t + s_t$ est la composante déterministe du modèle.

• ϵ_t est, par opposition, la composante aléatoire, supposée de moyenne nulle (parce qu'elle représente les composantes (erreurs) non systématiques), mais possédant en général une structure de corrélation non nulle. C'est bien évidemment la partie la plus intéressante à modéliser. Nous verrons comment, sous certaines hypothèses, nous pourrions ajuster un modèle pour ϵ_t qui permettra de prévoir les valeurs futures de la série chronologique. Notons que, si ϵ_t n'est pas de moyenne nulle, on pourra remplacer m_t , par $m_t + E\{\epsilon_t\}$, et ϵ_t sera remplacé par $W_t = \epsilon_t - E\{\epsilon_t\}$. On a alors $E\{W_t\} = 0$.

Dans un premier temps, nous allons estimer la composante déterministe du modèle. De cette façon, nous n'aurons plus qu'à nous intéresser à la partie aléatoire qui constitue le véritable enjeu de ce cours.

1.4 Elimination de la tendance en l'absence de la saisonnalité

Supposons, dans un premier temps, que la partie déterministe du modèle soit uniquement composée d'une tendance m_t

$$X_t = m_t + \epsilon_t \tag{1.2}$$

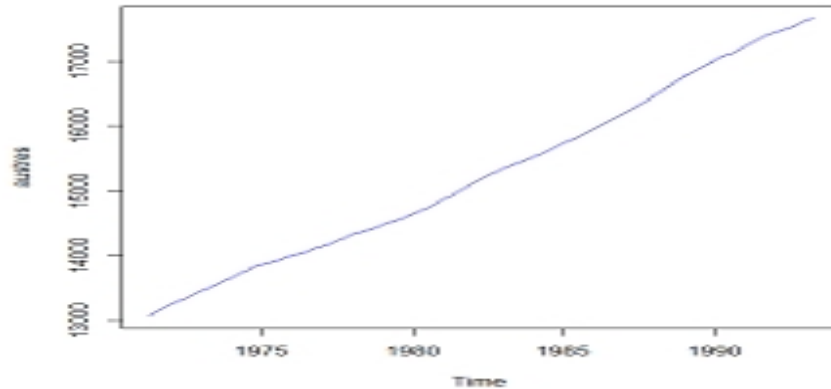
Dans cette section, nous donnons trois méthodes pour estimer m_t .

1.4.1 Estimation paramétrique

Cette méthode consiste à estimer la tendance par la méthode des moindres carrés ordinaires. Supposons que l'on observe la série de la figure suivante :

- **Script R**

```
plot(austres,col="blue")
```



Time Series of the Number of Ausralian Residents

Il semble naturel d'estimer la composante m_t de cette série par une fonction linéaire

$$\hat{m}_t = \hat{a} + \hat{b}t$$

où \hat{a} et \hat{b} sont des estimateurs des coefficients de la fonction linéaire estimant m_t . Pour trouver ces estimateurs par la méthode des moindres carrés, il faut minimiser l'erreur quadratique commise en remplaçant m_t par \hat{m}_t dans (1.2). En conséquence, si on observe la série X_1, \dots, X_T , il faut trouver les coefficients \hat{a} et \hat{b} minimisant

$$\sum_{t=1}^T (X_t - a - bt)^2. \quad (1.3)$$

La solution de ce problème de minimisation est, en posant $\bar{X} = \frac{1}{T} \sum_{t=1}^T X_t$,

$$\begin{cases} \hat{a} = \frac{4T+2}{T-1} \bar{X} - \frac{6}{T(T-1)} \sum_{t=1}^T tX_t \\ \hat{b} = \frac{6}{T(T-1)} \left(\frac{2}{T(T+1)} \sum_{t=1}^T tX_t - \bar{X} \right) \end{cases}$$

Preuve. Pour minimiser en a et b l'équation (1.3), il faut dériver cette expression selon a et b , ce qui donne, après quelques simplifications et en égalant

à zéro, le système

$$(i) \quad \sum_{t=1}^T X_t - aT - b \sum_{t=1}^T t = 0$$

$$(ii) \quad \sum_{t=1}^T tX_t - a \sum_{t=1}^T t - b \sum_{t=1}^T t^2 = 0$$

On peut à présent utiliser les formules usuelles suivantes

$$\sum_{t=1}^T t = \frac{T(T+1)}{2}$$

$$\sum_{t=1}^T t^2 = \frac{T(T+1)(2T+1)}{6}$$

pour obtenir, en divisant (i) et (ii) par T

$$(i') \quad \bar{X} - a - \frac{(T+1)}{2}b = 0$$

$$(ii') \quad \frac{1}{T} \sum_{t=1}^T tX_t - \frac{(T+1)}{2}a - \frac{(T+1)(2T+1)}{6}b = 0$$

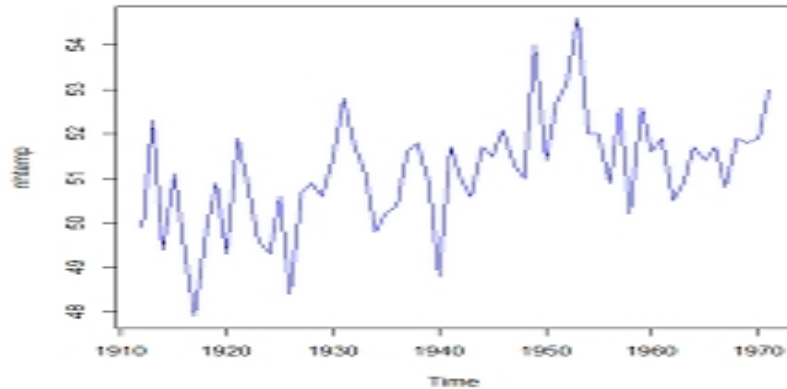
où \bar{X} est la moyenne empirique de $\{X_t\}_{t=1}^T$. La solution découle directement de ce système, en vérifiant que \hat{a} et \hat{b} sont bien des minima, et non pas des maxima (en considérant les dérivées partielles du second ordre). ■

1.4.2 Estimation non paramétrique

Dans certaines situations, il n'est pas facile de trouver le degré du polynôme d'ajustement pour m_t . Par exemple, dans la figure suivante,

- **Script R**

```
plot(nhtemp,col="blue")
```



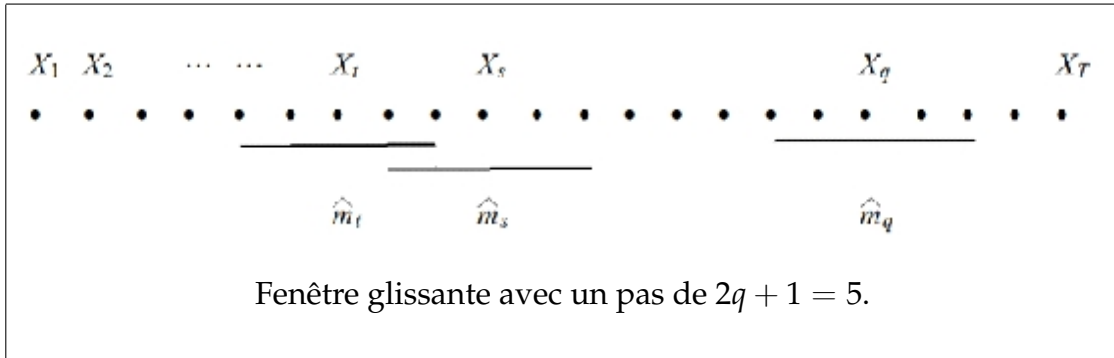
A Verage Yearly Temperature in New Haven

il n'est pas possible d'utiliser la méthode des moindres carrés car le polynôme utilisé au départ pour \hat{m}_t n'est ni linéaire, ni quadratique. On pourrait utiliser un polynôme avec un degré élevé, mais le nombre de paramètres à estimer serait important et rendrait les calculs fastidieux. Dans cette situation, on a recourt à la théorie non paramétrique de l'estimation de la tendance, qui ne suppose pas que la tendance soit polynomiale à priori. Pour comprendre cette technique, supposons que m_t soit linéaire dans un intervalle $[t - q, t + q]$. Dans ce cas, un bon estimateur de la tendance est donné par

$$\hat{m}_t = \frac{1}{2q + 1} \sum_{k=-q}^q X_{t+k}.$$

On peut à présent calculer \hat{m}_t pour chaque valeur de t en calculant cette moyenne sur les $2q$ observations autour de t . On peut aussi voir l'intervalle $[t - q, t + q]$ comme une "fenêtre" sur les observations, que l'on déplace lorsque t varie. A chaque valeur de t , l'estimateur \hat{m}_t calcule la moyenne des observations tombant dans cette fenêtre glissante (on dit alors qu'on effectue une

estimation par moyenne mobile).



Pour éviter que des problèmes se posent aux bords de la série (lorsque $t \leq q$ ou $t > T - q$), on pose

$$\begin{cases} X_t = X_1 \text{ si } t < 1 \\ X_t = X_T \text{ si } t > T \end{cases}$$

1.4.3 Méthode des différences itérés

On peut également éliminer la tendance sans l'estimer. Définissons tout d'abord l'opérateur de retard comme étant la fonction linéaire B qui à toute donnée X_t fait correspondre la donnée précédente X_{t-1}

$$BX_t = X_{t-1}.$$

Cet opérateur peut être utilisé récursivement, pour obtenir l'avant-dernière observation avant t

$$B^2X_t = BBX_t = BX_{t-1} = X_{t-2}$$

et on peut écrire par induction

$$B^kX_t = X_{t-k}.$$

On introduit également l'opérateur de différence ∇ dont l'application à X_t fournit la différence entre X_t et la valeur précédente X_{t-1} de la série

$$\nabla X_t = X_t - X_{t-1}.$$

En vertu de la définition précédente, on a directement

$$\begin{aligned} \nabla X_t &= X_t - X_{t-1} \\ &= X_t - BX_t \\ &= (1 - B) X_t. \end{aligned}$$

Tout comme avec l'opérateur B , on peut itérer ∇ pour obtenir

$$\begin{aligned}\nabla^2 X_t &= \nabla \nabla X_t \\ &= (1 - B)(1 - B) X_t \\ &= (1 - 2B + B^2) X_t \\ &= X_t - 2X_{t-1} + X_{t-2}.\end{aligned}$$

Plus généralement, l'opérateur de différence d'ordre d envoie la valeur X_t de la série observée sur la différence de X_t avec la valeur prise par la série au temps $t - d$

$$\nabla_d X_t = X_t - X_{t-d}$$

d'où, on a

$$\nabla_d X_t = (1 - B^d) X_t.$$

Supposons maintenant que, dans (1.2), la tendance soit de la forme d'un polynôme d'ordre k

$$m_t = \sum_{j=0}^k a_j t^j.$$

Si on applique à m_t l'opérateur de différence (d'ordre 1), on obtient

$$\begin{aligned}\nabla m_t &= m_t - m_{t-1} \\ &= \sum_{j=0}^k a_j t^j - \sum_{j=0}^k a_j (t-1)^j\end{aligned}$$

et, en développant $(t-1)^j$, on obtient, dans le membre de droite

$$\sum_{j=0}^k a_j (t-1)^j = \sum_{j=0}^k a_j t^j + \sum_{j=1}^k b_{j-1} t^{j-1}$$

pour certains coefficients b_{j-1} . Donc, ∇m_t devient

$$\begin{aligned}\nabla m_t &= \sum_{j=1}^k b_{j-1} t^{j-1} \\ &= \sum_{i=0}^{k-1} b_i t^i \text{ avec } i = j - 1\end{aligned}$$

et on voit que l'opérateur ∇ a diminué l'ordre du polynôme m_t , m_t est d'ordre k , tandis que ∇m_t est d'ordre $k - 1$. Alors, si on applique k fois l'opérateur ∇ au polynôme m_t , on obtient

$$\nabla^k m_t = k! a_k = \text{constante indépendante de } t$$

et donc, on applique à (1.2), l'opérateur ∇^k on obtient

$$\nabla^k X_t = \text{constante} + \nabla^k \epsilon_t.$$

Bien entendu, en pratique, on ignore le degré du polynôme m_t . C'est pourquoi on applique l'opérateur ∇ à la série observée autant de fois qu'il est nécessaire pour obtenir une série $\nabla^k X_t$ dont la moyenne est nulle (et donc seule la composante aléatoire est présente). En pratique k est assez petit (de l'ordre 1, 2 et 3).

1.5 Élimination de la tendance et de la saisonnalité

Supposons à présent que, en plus d'une tendance m_t , la série comporte une composante saisonnière s_t . On considère donc le modèle

$$X_t = m_t + s_t + \epsilon_t$$

où, comme nous l'avons spécifié plus haut,

- s_t est une fonction périodique en t , c'est-à-dire, $s_t = s_{t+d}$ pour un certain $d > 0$.

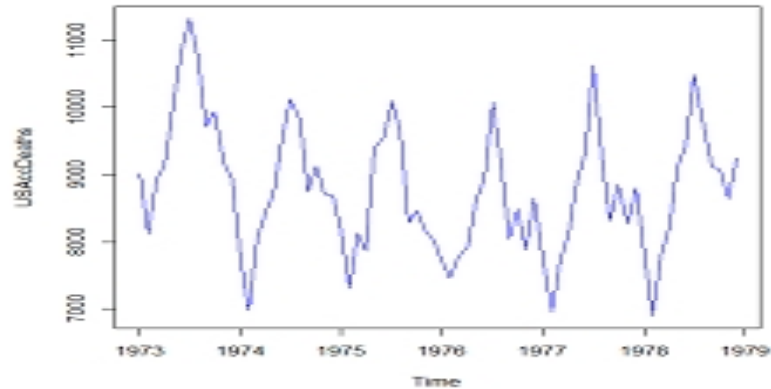
- s_t ne comporte pas de composante "tendance" et donc

$$\sum_{j=1}^d s_{t+j} = 0.$$

Voyons comment éliminer la composante saisonnière sur l'exemple de la figure suivante :

- **Script R**

```
plot(USAccDeaths,col="blue")
```



Nombre de morts accidentels aux Etas-Unis

dans lequel nous observons des données mensuelles dont la période est visiblement $d = 12$. Dans cet exemple, il est utile de réorganiser les données $\{X_t\}_{t=1}^T$ dans un tableau dont les entrées sont les années (1973 à 1978) et les mois (1 à 12)

$$\{X_t\}_{t=1}^T = \{X_{jk}\}_{\substack{k=1,\dots,12 \text{ (mois)} \\ j=1,\dots,6 \text{ (ans)}}$$

avec la relation : $X_{jk} = X_{k+12(j-1)}$.

Dans la figure précédente, on observe que la tendance ne varie pas beaucoup et il semble naturel de la supposer constante sur une année. Un estimateur naturel de la tendance pour l'année j est donc

$$\hat{m}_j = \frac{1}{12} \sum_{k=1}^{12} X_{jk}, \text{ où } j = 1, \dots, 6.$$

Les résidus $X_j - \hat{m}_j$ comportent toujours la composante saisonnière. Remarquons que cette composante saisonnière dépend par définition que du mois et non de l'année. En conséquence, un estimateur de cette composante s_j est donné par la moyenne empirique sur le même mois de chaque année des résidus

$$\hat{s}_k = \frac{1}{6} \sum_{j=1}^6 (X_{jk} - \hat{m}_{jk}), \text{ où } k = 1, \dots, 12.$$

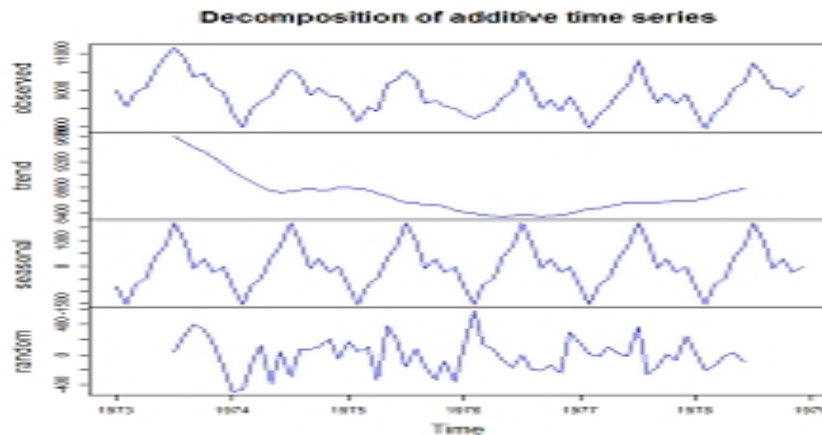
Après élimination de la tendance et de la saisonnalité

$$\hat{\epsilon}_t = X_{jk} - \hat{m}_j - \hat{s}_k.$$

On obtient donc la composante aléatoire de la série que nous analyserons dans les prochains chapitres.

- **Script R**

```
plot(decompose(USAccDeaths),col="blue")
```



1.5.1 Méthode des différences

On peut également appliquer la méthode des différences pour éliminer à la fois la tendance et la saisonnalité. Etant donné le modèle (1.1), appliquons l'opérateur de différence d'ordre d , ∇^d , où d est la période de s_t , à la série observée X_t

$$\begin{aligned}\nabla_d X_t &= X_t - X_{t-d} \\ &= m_t + s_t + \epsilon_t - m_{t-d} - s_{t-d} - \epsilon_{t-d} \\ &= (m_t - m_{t-d}) + (s_t - s_{t-d}) + (\epsilon_t - \epsilon_{t-d})\end{aligned}$$

où, par définition, le terme $s_t - s_{t-d}$ est nul. ∇_d est donc une série chronologique désaisonnalisée, mais contenant une tendance $\nabla_d m_t$. Les méthodes d'élimination de la tendance en l'absence de saisonnalité s'appliquent donc.

Chapitre 2

Processus Stationnaire

Dans le chapitre précédent, nous nous sommes occupés de la modélisation des composantes déterministes (dépendantes explicitement du temps t et par conséquent non stationnaires) d'un modèle additif de décomposition de type (1.1).

Le but de ce chapitre est maintenant d'introduire le cadre pour modéliser la composante stationnaire ϵ_t . C'est-à-dire, dans ce chapitre nous fournissons une série de définitions de base de quantités qui caractérisent des processus stationnaires. Sans perte de généralité, nous utilisons X_t pour représenter nos observations, à présent supposées stationnaires.

Lorsqu'on parle du modèle pour une série temporelle X_1, \dots, X_T on utilise certaines notations et définitions spécifiques. La suite de variables aléatoires $(X_t, t \in \mathbb{Z})$ est appelée processus stochastique. Les observations (données) empiriques X_1, \dots, X_T sont interprétées comme des réalisations particulières de ce processus, tandis que $(X_t, t \in \mathbb{Z})$, lorsqu'on ajoute des conditions sur sa structure probabiliste (ses moyennes, ses variances et covariances, ...) sert à dénoter le modèle (paramétrique) théorique. Celui-ci aidera à mieux comprendre le comportement "en moyenne" et des propriétés statistiques des estimateurs des paramètres du modèle.

Ayant introduit la notion d'un processus stochastique servant à modéliser des séries chronologiques, on se concentre maintenant sur la classe importante des processus stationnaires. Considérons une suite de variables aléatoires $(X_t, t \in \mathbb{Z})$. On dit que cette suite est stationnaire en moyenne lorsque la moyenne de chacune des variables de la suite est identique

$$E \{X_t\} = E \{X_0\} \quad \forall t \in \mathbb{Z}.$$

De même, cette suite est stationnaire en variance lorsque

$$\text{Var} \{X_t\} = \text{Var} \{X_0\} \quad \forall t \in \mathbb{Z}.$$

2.1 Fonctions Caractéristiques

Maintenant, nous donnons quelques définitions et notations utiles pour la suite de ce cours.

2.1.1 Fonction d'autocovariance

Définition 2 Soit $(X_t, t \in \mathbb{Z})$ un processus de moyenne $E(X_t) = \mu$, on appelle la fonction d'autocovariance la fonction γ définie par :

$$\gamma(h) = \text{Cov}(X_t, X_{t+h}) = E[(X_{t+h} - \mu)(X_t - \mu)], \quad \forall h \in \mathbb{Z}$$

Remarque 3 La fonction d'autocovariance a les propriétés suivantes :

1. La fonction $\gamma(h)$ est symétrique, i.e $\forall h \in \mathbb{Z}, \gamma(h) = \gamma(-h)$.
2. $\gamma(0) \geq 0$
3. $|\gamma(h)| \leq \gamma(0)$, pour tout $h \in \mathbb{Z}$.

2.1.2 Fonction d'autocorrélation (ACF)

Définition 4 La fonction d'autocorrélation (ACF) du processus $(X_t, t \in \mathbb{Z})$ notée ρ est définie par

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}, \quad \forall h \in \mathbb{Z}$$

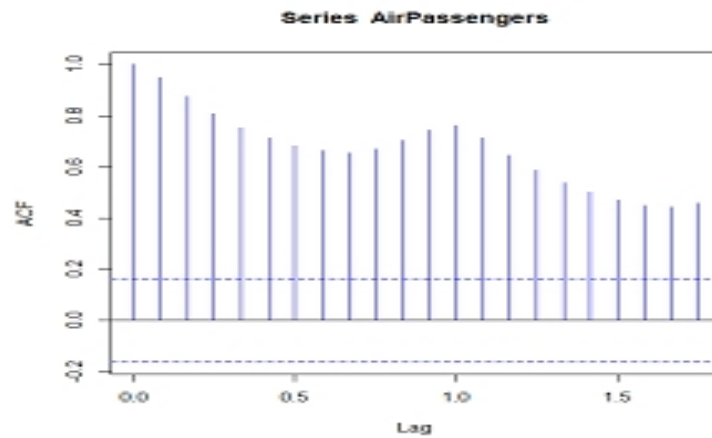
avec $\rho(h) \in [-1, 1]$.

Remarque 5

1. $\rho(0) = 1$.
2. La fonction $\rho(h)$ est symétrique, i.e $\forall h \in \mathbb{Z}, \rho(h) = \rho(-h)$.

- **Script R**

```
acf(AirPassengers,col="blue")
```



2.1.3 Fonction d'autocorrélation partielle (PACF)

Définition 6 Le coefficient de corrélation partielle entre les deux variables X_1 et X_n d'un processus stochastique $(X_t, t \in \mathbb{Z})$ est le coefficient de corrélation entre les deux variables auxquelles on a retranché leurs meilleures explications en termes de X_1, X_2, \dots, X_{n-1} :

$$r_{X_2, \dots, X_{n-1}}(X_1, X_n) = \text{corr}(X_1 - P_{X_2, \dots, X_{n-1}}(X_1), X_n - P_{X_2, \dots, X_{n-1}}(X_n)),$$

où corr est le coefficient de corrélation classique, et $P_{X_2, \dots, X_{n-1}}(X_1)$ est la projection de la variable X_1 dans l'espace vectoriel engendré par les variables X_2, \dots, X_{n-1} .

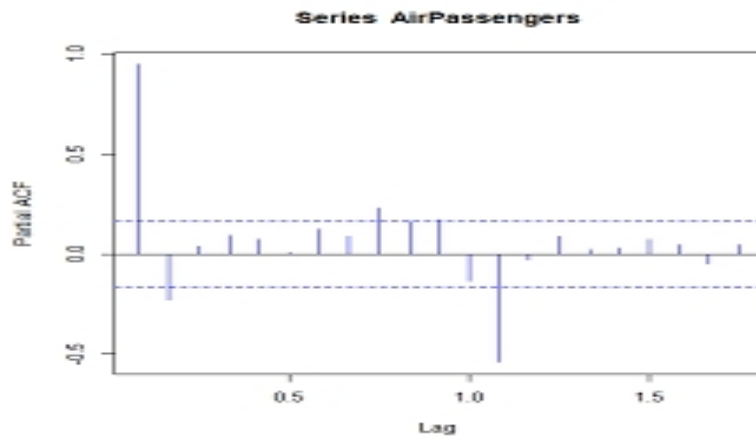
Définition 7 On appelle une fonction d'autocorrélation partielle notée (PACF) la fonction r définie par :

$$\begin{aligned} r(h) &= r_{X_2, \dots, X_h}(X_1, X_{h+1}) \quad \forall h \geq 2 \\ r(h) &= r(-h) \quad \forall h \neq 0 \\ r(1) &= \rho(1) \end{aligned}$$

Remarque 8 La fonction d'autocorrélation partielle est constituée par l'ensemble des autocorrélations partielles, le coefficient d'autocorrélation partielle mesurant la corrélation entre les variables X_t et X_{t-h} , l'influence de la variable X_{t-h-i} étant contrôlée pour $i < h$.

- **Script R**

```
pacf(AirPassengers,col="blue")
```

2.2 Stationnarité

La stationnarité d'une série temporelle est une propriété fondamentale. Elle indique si les caractéristiques de celui-ci changent avec le temps ou non. Si le processus n'est pas stationnaire alors les informations recueillies dans les données peuvent être mal interprétées. Par exemple, si on observe dans les données deux phases avec des moyennes clairement distinctes, alors les informations telle que la loi marginale ou la structure de corrélation, calculées sur l'ensemble des données n'ont plus un sens. Il faudra dans ce cas séparer les phases et utiliser un modèle différent pour chacune d'entre elles. Formellement, il existe deux types de stationnarité pour un processus $(X_t, t \in \mathbb{Z})$.

2.2.1 Définitions et exemples

Définition 9 (Stationnarité au sens strict) Un processus $(X_t, t \in \mathbb{Z})$ est strictement stationnaire si, pour des entiers arbitraires $m_1 < m_2 < \dots < m_n$, la distribution jointe de $X_{t-m_1}, X_{t-m_2}, \dots, X_{t-m_n}$ ne dépend pas du temps t .

Définition 10 (Stationnarité au second-ordre) Un processus $(X_t, t \in \mathbb{Z})$ est dit stationnaire au second-ordre si $\forall m, h \in \mathbb{Z}$, on a

- (i) $E(X_t^2) < \infty$.
- (ii) $E(X_t) = m$, i. e., $E(X_t)$ ne dépend pas de t .
- (iii) $Cov(X_t, X_{t+h}) = \gamma(h)$, i. e., $E(X_t)$ ne dépend pas de t .

2.2.2 Bruit blanc

L'exemple le plus simple de processus stationnaire au second-ordre est celui du bruit blanc. Ce processus est particulièrement important car il permet de construire des processus stationnaires plus complexes.

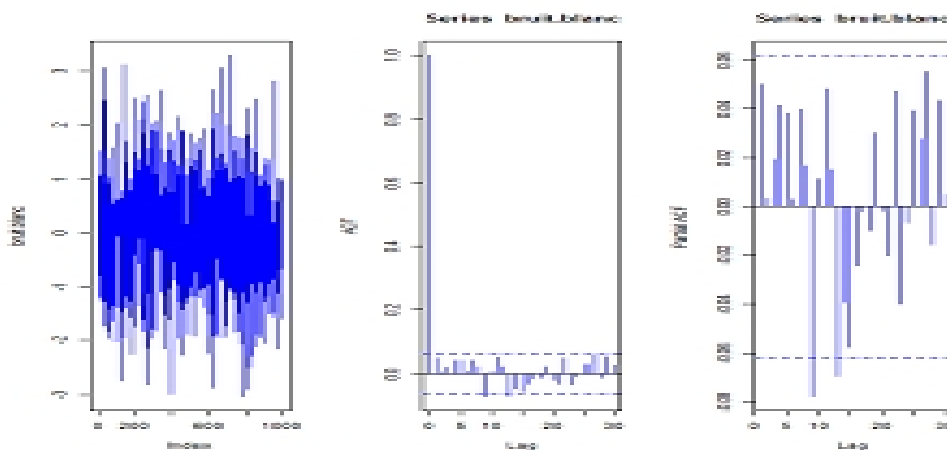
Définition 11 *Le processus $(\epsilon_t, t \in \mathbb{Z})$ est appelé bruit blanc faible s'il vérifie,*

$$\begin{cases} E(\epsilon_t) = 0, \forall t \in \mathbb{Z} \\ \gamma(h) = E(\epsilon_t \epsilon_{t+h}) = \begin{cases} \sigma_\epsilon^2 & \text{si } h=0 \\ 0 & \text{si } h \neq 0 \end{cases}, \forall t, h \in \mathbb{Z} \end{cases}$$

On dit qu'il est fort lorsque de plus les variables ϵ_t sont indépendantes.

- **Script R**

```
bruit.blanc<-rnorm(1000)
par(mfrow=c(1,3))
plot(bruit.blanc,type="l",col="blue")
acf(bruit.blanc,col="blue")
pacf(bruit.blanc,col="blue")
abline(h=mean(bruit.blanc),col="red",lwd=2)
```



2.2.3 Exemple d'un processus non stationnaire : Marche aléatoire

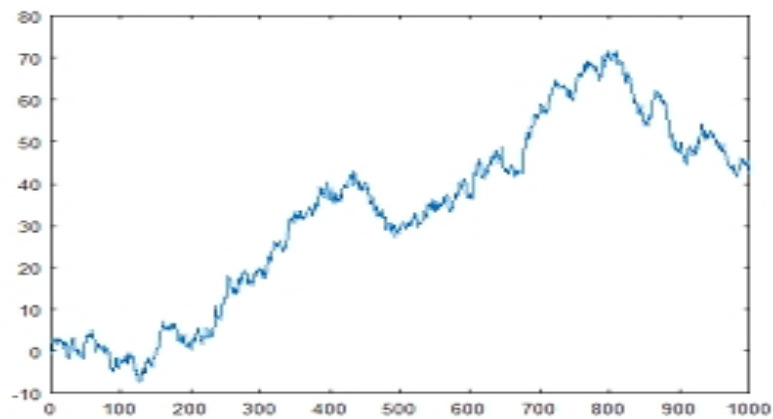
La suite $(X_t, t \in \mathbb{Z})$ est une marche aléatoire (random walk) lorsqu'elle s'écrit sous la forme suivante

$$\begin{aligned} X_t &= X_{t-1} + \epsilon_t \\ X_t &= X_0 + \sum_{i=1}^t \epsilon_i, \end{aligned}$$

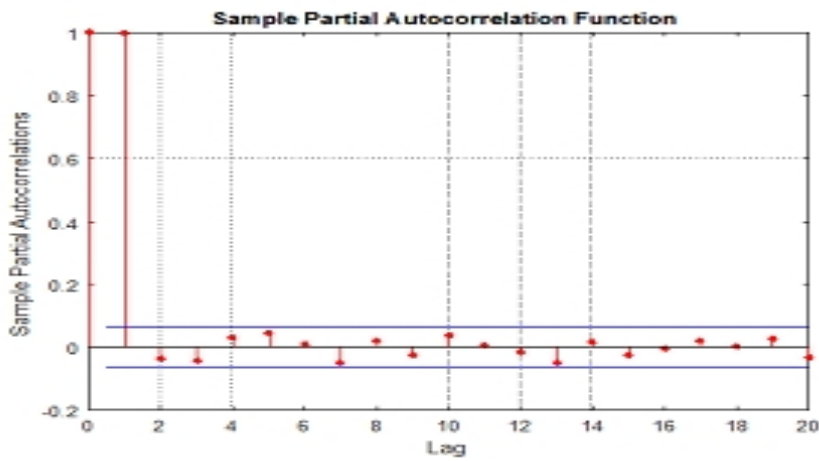
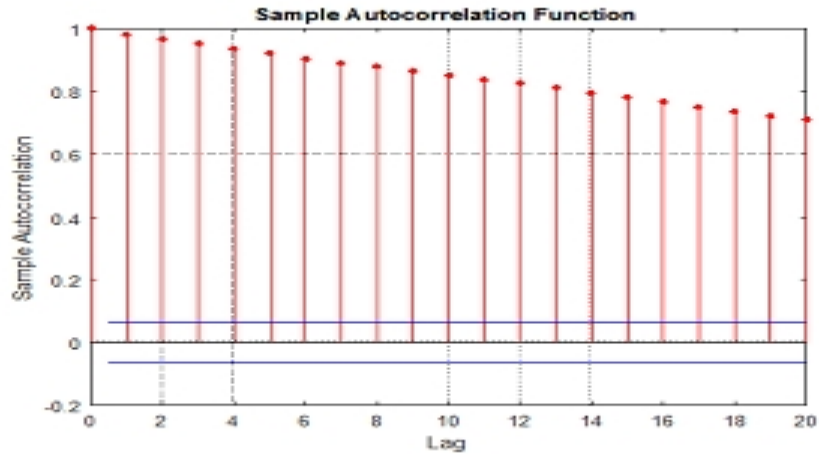
où $\{\epsilon_i\}$ est un bruit blanc de moyenne nulle et de variance σ_ϵ^2 et X_0 dénote une valeur initiale quelconque. On note que son espérance $E\{X_t\}$ est constante et égale à $E\{X_0\}$ pour tout t . De plus,

$$\begin{aligned} \text{Cov}(X_{t+h}, X_t) &= \text{Cov}\left(\sum_{i=1}^{t+h} \epsilon_i, \sum_{j=1}^t \epsilon_j\right) \\ &= \sum_{i=1}^{t+h} \sum_{j=1}^t \epsilon_i \epsilon_j \\ &= \sum_{i=1}^t \sigma_\epsilon^2 = t \cdot \sigma_\epsilon^2. \end{aligned}$$

Donc une marche aléatoire n'est pas stationnaire car la suite des covariances dépend de t . En fait, on observe en plus que sa variance n'est pas constante mais augmente en temps t . L'évolution de X_t est aléatoire, et on ne peut pas faire de prévision car il manque une structure adéquate.



Trajectoire d'une marche alatoire



Chapitre 3

Processus linéaires

Dans ce chapitre, nous entamons l'étude détaillée de la partie aléatoire d'une série chronologique. Les processus linéaires constituent le modèle le plus simple pour décrire cette composante. À la fin de ce chapitre, on verra trois exemples importants, les processus autorégressifs, les processus moyennes mobiles (moving average) et les processus mixtes. Ces processus sont des combinaisons linéaires de processus de bruit blanc (fort ou faible) dont l'intérêt tient au comportement de leur fonction d'autocovariance, qui tend vers zéro lorsque l'on compare deux variables éloignées dans le temps :

$$\gamma(h) = \text{Cov}(X_{t+h}, X_t) \rightarrow 0 \text{ lorsque } |h| \rightarrow +\infty.$$

Intuitivement, cela signifie que la "mémoire" du processus est contrôlée : la liaison du second ordre est plus importante pour deux variables proches dans le temps que pour des variables éloignées. Ce modèle est donc à mémoire appelée "courte" même si elle est "infinie" pour un processus autorégressif.

Nous étudions dans ce chapitre les processus linéaires ainsi que deux cas particuliers de tels processus : les processus autorégressifs, les processus moyennes mobiles.

3.1 Processus linéaires

Un processus linéaire est un processus stochastique $(X_t, t \in \mathbb{Z})$ formé par une combinaison linéaire (non nécessairement finie) de bruits blancs forts. On définit également la classe des processus linéaires généraux, qui sont constitués de combinaisons linéaires de bruits blancs faibles. Introduisons formellement ces deux types de processus.

Définition 12 Le processus $(X_t, t \in \mathbb{Z})$ est un processus linéaire (resp. linéaire général) de moyenne μ s'il peut être écrit sous la forme

$$X_t = \mu + \sum_{i=-\infty}^{+\infty} b_i \epsilon_{t-i}$$

où $(\epsilon_t, t \in \mathbb{Z})$ est un bruit blanc fort (resp. faible), de variance σ_ϵ^2 et où la suite des coefficients b_i est supposée telle que

$$\sum_{i=-\infty}^{+\infty} b_i^2 < +\infty.$$

Le résultat suivant montre que les processus linéaires que nous avons définis sont des processus stationnaires.

Proposition 13 Si le processus $(X_t, t \in \mathbb{Z})$ est un processus linéaire général alors $(X_t, t \in \mathbb{Z})$ est stationnaire et on a pour tout $h \in \mathbb{Z}$

$$\begin{cases} \gamma(h) = \text{Cov}(X_{t+h}, X_t) = \sigma_\epsilon^2 \sum_{i=-\infty}^{+\infty} b_{i+h} b_i \\ \gamma(0) = \text{Var}(X_t) = \sigma_\epsilon^2 \sum_{i=-\infty}^{+\infty} b_i^2 \end{cases}$$

Preuve. On a déjà remarqué ci-dessus que $E\{X_t\} = \mu$ (pour tout $t \in \mathbb{Z}$). Pour vérifier les deux autres conditions, on calcule la covariance entre X_t et X_{t+h}

$$\begin{aligned} \text{Cov}(X_{t+h}, X_t) &= E\{X_{t+h} \cdot X_t\} - E\{X_{t+h}\} E\{X_t\} \\ &= E \left\{ \mu^2 + \mu \sum_{i=-\infty}^{+\infty} b_i \epsilon_{t+h-i} + \mu \sum_{j=-\infty}^{+\infty} b_j \epsilon_{t-j} + \sum_{i,j=-\infty}^{+\infty} b_j b_i \epsilon_{t-j} \epsilon_{t+h-i} \right\} \\ &\quad - \mu^2 \\ &= \mu \sum_{i=-\infty}^{+\infty} b_i E\{\epsilon_{t+h-i}\} + \mu \sum_{j=-\infty}^{+\infty} b_j E\{\epsilon_{t-j}\} \\ &\quad + \sum_{i,j=-\infty}^{+\infty} b_j b_i E\{\epsilon_{t-j} \epsilon_{t+h-i}\} \end{aligned}$$

où le passage à l'espérance dans la somme infinie est justifiée par le théorème de la convergence monotone et le fait que $E\{X_t\} = \mu < 1$ pour tout t . Puisque $(\epsilon_t, t \in \mathbb{Z})$ est un bruit blanc de variance σ_ϵ^2 , on a donc

$$\begin{aligned} \text{Cov}(X_{t+h}, X_t) &= \sum_{i,j=-\infty}^{+\infty} b_j b_i \text{Cov}(\epsilon_{t-j}, \epsilon_{t+h-i}) \\ &= \sigma_\epsilon^2 \sum_{i=-\infty}^{+\infty} b_{i+h} b_i. \end{aligned}$$

Donc $Cov(X_{t+h}, X_t)$ ne dépend pas de t et on a les résultats annoncés. ■

Nous présentons maintenant deux résultats portant sur l'image de processus stationnaires par certaines transformations.

Proposition 14 Soit $(X_t, t \in \mathbb{Z})$ un processus stationnaire et soit $(a_i)_{i \in \mathbb{Z}}$ une suite de réels absolument sommable (i.e. telle que $\sum_{i=-\infty}^{+\infty} |a_i| < \infty$). Alors le processus $(Y_t, t \in \mathbb{Z})$ défini par

$$Y_t = \sum_{i=-\infty}^{+\infty} a_i X_{t-i}$$

est stationnaire.

Preuve. A faire en exercice. ■

Exemple 15 Soit $(\epsilon_t, t \in \mathbb{Z})$ un bruit blanc et soit $(a_i)_{i \in \mathbb{Z}}$ une suite de réels absolument sommable (i.e. $\sum_{i=-\infty}^{+\infty} |a_i| < \infty$). Alors le processus $(Y_t, t \in \mathbb{Z})$ défini par

$$Y_t = \sum_{i=-\infty}^{+\infty} a_i \epsilon_{t-i}$$

est un processus stationnaire. Il est naturellement appelé *moyenne mobile infinie*. On a alors

$$E \{Y_t\} = 0 \text{ et } Var \{Y_t\} = \sigma_\epsilon^2 \sum_{i=-\infty}^{+\infty} a_i^2$$

et plus généralement

$$\gamma(h) = \sigma_\epsilon^2 \sum_{i=-\infty}^{+\infty} a_i a_{i-h}.$$

Terminons cette section par quelques définitions. Soit $(X_t, t \in \mathbb{Z})$ un processus stationnaire centré. On note $H(X)$ et on appelle *histoire* de X l'espace engendré par les $(X_t, t \in \mathbb{Z})$ (i.e. l'espace des combinaisons linéaires des $(X_t, t \in \mathbb{Z})$). Le passé linéaire de X jusqu'à l'instant t , noté $H_{\leq t}(X)$ est l'espace engendré par les $(X_u)_{u \leq t}$. Les $H_{\leq t}(X)$ forment une chaîne croissante de sous-espaces de $H(X)$. Leur réunion est dense dans $H(X)$ et leur intersection $H_{-\infty}(X) = \bigcap_t H_{\leq t}(X)$ est appelée *passé lointain* de X .

Définissons maintenant pour tout t , P_t la projection orthogonale sur $H_{\leq t}(X)$. Notons alors

$$\epsilon_t = X_t - P_{t-1}(X_t).$$

Définition 16 La série temporelle $(\epsilon_t, t \in \mathbb{Z})$ est un bruit blanc appelé bruit blanc d'innovation et les ϵ_t sont les innovations fondamentales du processus X .

La relation

$$X_t = P_{t-1}(X_t) + \epsilon_t$$

montre que ϵ_t est ce qui arrive de "nouveau" à X à l'instant t . La variance de ϵ est nécessairement inférieure ou égale à celle de X .

Définition 17

1. On dit que X est purement déterministe si son bruit blanc d'innovation est nul. On peut alors (en théorie) prévoir exactement les valeurs futures de X en fonction de la connaissance de toutes ses valeurs du passé. Ce qui est équivalent à $H_{\leq t}(X) = H_{\leq t+1}(X)$ pour au moins un t ; ce qui est encore équivalent à $H_{-\infty}(X) = H(X)$.
2. L'extrême opposé est le suivant : on dit que X est purement innovant si X est non nul et si $H_{-\infty}(X) = \{0\}$.

Définition 18

1. Un processus $(X_t, t \in \mathbb{Z})$ admet une représentation inversible s'il peut s'écrire comme combinaison linéaire des valeurs d'un autre processus, c'est-à-dire, qu'il existe une suite $(\psi_i, i \in \mathbb{Z})$ et un processus $(Y_t, t \in \mathbb{Z})$ tels que

$$\forall t \in \mathbb{Z} : Y_t = \sum_{i \in \mathbb{Z}} \psi_i X_{t-i}.$$

2. Un processus $(X_t, t \in \mathbb{Z})$ admet une représentation causale s'il peut s'écrire comme combinaison linéaire des valeurs passées d'un autre processus, c'est-à-dire, qu'il existe une suite $(\psi_i, i \in \mathbb{Z})$ et un processus $(Y_t, t \in \mathbb{Z})$ tels que

$$\forall t \in \mathbb{Z} : X_t = \sum_{i \in \mathbb{N}} \psi_i Y_{t-i}.$$

3.2 Processus Moyennes Mobiles

Dans cette section nous considérons les processus à moyenne mobile. Ces processus forment une classe flexible de modèles pour nombreux phénomènes observés. Ils sont construits à partir de l'idée que l'observation au temps t s'explique linéairement par les observations d'un bruit blanc. Pour étudier ces processus, nous allons commencer par les fonctions caractéristiques en donnant les définitions avec la représentation de leurs trajectoires.

On commence par la définition suivante qui présente ces processus.

Définition 19 Un processus $(X_t, t \in \mathbb{Z})$, est dit un processus moyenne mobile d'ordre q (Moving Average), noté $MA(q)$ s'il est défini par :

$$X_t = \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i}$$

où les θ_i sont des réels et $\epsilon_t \rightsquigarrow bb(0, \sigma_\epsilon^2)$.

On introduit l'opérateur de retard B (Celui-ci se définit simplement par : $B^i X_t = X_{t-i}$) pour simplifier la formule d'un $MA(q)$:

$$\begin{aligned} X_t &= \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} \\ X_t &= (1 + \theta_1 B + \dots + \theta_q B^q) \epsilon_t \\ X_t &= \Theta(B) \epsilon_t \end{aligned}$$

où $\Theta(B)$ est un polynôme de retard d'ordre q associé au processus X_t .

3.3 Fonctions caractéristiques d'un processus MA(q)

Nous allons présenter dans cette section les fonctions caractéristiques d'un processus $MA(q)$. En commençant par l'espérance et la variance. En effet, par un calcul direct, nous avons

$$E(X_t) = E\left[\sum_{i=0}^q \theta_i \epsilon_{t-i}\right] = \sum_{i=0}^q \theta_i E(\epsilon_{t-i}) = 0,$$

et

$$var(X_t) = \gamma(0) = E(X_t^2) - [E(X_t)]^2 = E(X_t^2),$$

donc

$$\begin{aligned} var(X_t) &= E(X_t^2) = E\left[\left(\sum_{i=0}^q \theta_i \epsilon_{t-i}\right)^2\right] \\ &= E\left[\sum_{i=0}^q \sum_{j=0}^q \theta_i \theta_j \epsilon_{t-i} \epsilon_{t-j}\right] = \sum_{i=0}^q \sum_{j=0}^q \theta_i \theta_j E(\epsilon_{t-i} \epsilon_{t-j}) \end{aligned}$$

on a

$$E(\epsilon_{t-i} \epsilon_{t-j}) = \begin{cases} 0 & \text{si } j \neq i \\ \sigma_\epsilon^2 & \text{si } j = i \end{cases}.$$

D'où

$$var(X_t) = \sigma_\epsilon^2 \sum_{i=0}^q \theta_i^2$$

Fonction d'autocovariance

La fonction d'autocovariance est donnée par :

$$\gamma(h) = E(X_t X_{t+h}) = \begin{cases} \sigma^2 \sum_{i=0}^q \theta_{|h|+i} \theta_i & \text{si } |h| \leq q, h \in \mathbb{Z} \\ 0 & \text{sinon} \end{cases}.$$

En effet, on a

$$\begin{aligned} \gamma(h) &= E(X_t X_{t+h}) = E \left[\left(\sum_{i=0}^q \theta_i \epsilon_{t-i} \right) \left(\sum_{j=0}^q \theta_j \epsilon_{t+h-j} \right) \right] \\ &= \sum_{i=0}^q \sum_{j=0}^q \theta_i \theta_j E(\epsilon_{t-i} \epsilon_{t+h-j}). \end{aligned}$$

Sachant que : $E(\epsilon_{t-i} \epsilon_{t+h-j}) = \sigma_\epsilon^2$ quand $t+h-j = t-i \Rightarrow i = j-h$

$$\begin{cases} 0 \leq i \leq q \\ 0 \leq j \leq q \end{cases} \Rightarrow i \leq q-h \Rightarrow i \leq \min(q, q-h)$$

$$\begin{aligned} E(X_t X_{t+h}) &= \sigma_\epsilon^2 \sum_{i=0}^{q-|h|} \theta_{|h|+i} \theta_i \text{ si } |h| \leq q \\ \gamma(h) &= \begin{cases} \sigma_\epsilon^2 \sum_{i=0}^{q-|h|} \theta_{|h|+i} \theta_i & \text{si } |h| \leq q, h \in \mathbb{Z} \\ 0 & \text{sinon} \end{cases} \end{aligned}$$

Fonction d'autocorrélation

La fonction d'autocorrélation est donnée par :

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \begin{cases} \frac{\sum_{i=0}^{q-|h|} \theta_{|h|+i} \theta_i}{\sum_{i=0}^q \theta_i^2} & \text{si } |h| \leq q \\ 0 & \text{sinon} \end{cases}$$

Fonction d'autocorrélation partielle

La fonction d'autocorrélation partielle d'un processus $MA(1)$, qui est s'écrit sous la forme,

$$X_t = \epsilon_t + \theta \epsilon_{t-1}$$

est donnée par

$$\partial(h) = -\frac{(-\theta)^h (1 - \theta^2)}{1 - \theta^{2(h+1)}}.$$

Cette fonction est décroît exponentiellement vers 0 quand h augmente.

Proposition 20 *Un processus à moyenne mobile est automatiquement centré et stationnaire.*

Preuve. Soit $(X_t, t \in \mathbb{Z})$ un processus moyenne mobile. Nous avons

$$\begin{aligned} E(X_t) &= E\left(\sum_{i=0}^q \theta_i \epsilon_{t-i}\right) \\ &= \sum_{i=0}^q \theta_i E(\epsilon_{t-i}) = 0 \end{aligned}$$

d'où $E(X_t)$ ne dépend pas de t . De plus, on a

$$\text{var}(X_t) = \sigma_\epsilon^2 \sum_{i=0}^q \theta_i^2$$

qui ne dépend pas de t . Pour $\gamma(h)$, nous avons

$$\gamma(h) = \begin{cases} \sigma_\epsilon^2 \sum_{i=0}^{q-|h|} \theta_{|h|+i} \theta_i & \text{si } |h| \leq q, h \in \mathbb{Z} \\ 0 & \text{sinon} \end{cases}$$

il est clair que $\gamma(h)$ aussi ne dépend pas de t . Donc le processus $(X_t, t \in \mathbb{Z})$ est centré ($E(X_t) = 0$) et il est stationnaire. ■

Exemple 21 *Fonctions caractéristiques d'un processus MA(1)*

Soit $(X_t, t \in \mathbb{Z})$ un processus MA(1), X_t prend la forme suivante :

$$X_t = \epsilon_t + \theta \epsilon_{t-1}$$

pour tel processus, on a

$$\begin{aligned} E(X_t) &= E(\epsilon_t + \theta \epsilon_{t-1}) \\ &= E(\epsilon_t) + \theta E(\epsilon_{t-1}) = 0 \end{aligned}$$

et

$$\begin{aligned} \text{var}(X_t) &= \text{var}(\epsilon_t + \theta \epsilon_{t-1}) \\ &= \text{var}(\epsilon_t) + \theta^2 \text{var}(\epsilon_{t-1}) \\ &= (1 + \theta^2) \sigma_\epsilon^2 \end{aligned}$$

- La fonction d'autocovariance est donnée par :

$$\begin{aligned}\gamma(h) &= E(X_t X_{t+h}) = \text{cov}(\epsilon_t + \theta\epsilon_{t-1}, \epsilon_{t-h} + \theta\epsilon_{t+h-1}) \\ \gamma(1) &= E(X_t X_{t+1}) = E[(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t+1} + \theta\epsilon_t)] \\ &= E(\epsilon_t \epsilon_{t+1} + \theta\epsilon_t^2 + \theta\epsilon_{t-1}\epsilon_{t+1} + \theta^2\epsilon_{t-1}\epsilon_t) \\ &= E(\epsilon_t \epsilon_{t+1}) + \theta E(\epsilon_t^2) + \theta E(\epsilon_{t-1}\epsilon_{t+1}) + \theta^2 E(\epsilon_{t-1}\epsilon_t) \\ &= 0 + \theta\sigma_\epsilon^2 + 0 + 0 = \theta\sigma_\epsilon^2 \\ \gamma(h) &= \begin{cases} \theta\sigma_\epsilon^2 & \text{si } |h| = 1 \\ 0 & \text{si } |h| \geq 2 \end{cases}\end{aligned}$$

- Les coefficients d'autocorrélation :

$$\rho(h) = \begin{cases} \frac{\theta}{1+\theta^2} & \text{si } |h| = 1 \\ 0 & \text{si } |h| \geq 2 \end{cases}$$

- La fonction d'autocorrélation partielle :

$$\begin{aligned}\partial(1) &= -\frac{(-\theta)(1-\theta^2)}{1-\theta^4} = \frac{\theta}{1+\theta^2} \\ \partial(2) &= -\frac{(-\theta)^2(1-\theta^2)}{1-\theta^{2(2+1)}} = -\frac{\theta^2+\theta^4}{1-\theta^6} = \frac{-\theta^2}{\theta^4+\theta^2+1}\end{aligned}$$

• Les coefficients d'autocorrélation partielle d'un processus $MA(1)$ décroissent exponentiellement en fonction de h et seront tous négatifs si $\theta < 0$, soit alterneront en signe si $\theta > 0$.

- Soit $(X_t, t \in \mathbb{Z})$ un processus $MA(1)$ s'écrit sous la forme suivante

$$X_t = \epsilon_t + 0.8\epsilon_{t-1}$$

cherchons les trois premières autocorrélations partielles, on sait que la ACF d'un $MA(1)$ est définie par

$$\rho(h) = \begin{cases} \frac{\theta}{1+\theta^2} & \text{si } |h| = 1 \\ 0 & \text{si } |h| \geq 2 \end{cases}$$

donc pour notre exemple

$$\rho(h) = \begin{cases} 1 & \text{si } h = 0 \\ \frac{\theta}{1+\theta^2} = 0.4878 & \text{si } h = 1 \\ 0 & \text{si } h \geq 2 \end{cases}$$

de plus, nous avons

$$\begin{aligned}
 \partial(1) &= \rho(1) = 0.4878 \\
 \partial(2) &= \frac{|R^*(2)|}{|R(2)|} \\
 &= \frac{\begin{vmatrix} \rho(0) & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}}{\begin{vmatrix} \rho(0) & \rho(1) \\ \rho(1) & \rho(0) \end{vmatrix}} \\
 &= \frac{\rho(0)\rho(2) - \rho^2(1)}{\rho^2(0) - \rho^2(1)} = \frac{0 - (0.4878)^2}{1 - (0.4878)^2} = -0.3122
 \end{aligned}$$

maintenant, on a

$$\begin{aligned}
 \partial(3) &= \frac{|R^*(3)|}{|R(3)|} \\
 &= \frac{\begin{vmatrix} \rho(0) & \rho(1) & \rho(1) \\ \rho(1) & \rho(0) & \rho(2) \\ \rho(2) & \rho(1) & \rho(3) \end{vmatrix}}{\begin{vmatrix} \rho(0) & \rho(1) & \rho(2) \\ \rho(1) & \rho(0) & \rho(1) \\ \rho(2) & \rho(1) & \rho(0) \end{vmatrix}} = 0.22147.
 \end{aligned}$$

On peut vérifier que la relation

$$\partial(h) = -\frac{(-\theta)^h (1 - \theta^2)}{1 - \theta^{2(h+1)}}$$

donne les mêmes résultats :

$$\begin{aligned}
 \partial(1) &= \frac{\theta}{1 + \theta^2} = \frac{0.8}{1 + (0.8)^2} = 0.4878 \\
 \partial(2) &= \frac{-(-\theta^2)}{(1 + \theta^2 + \theta^4)} = \frac{-(-0.8)^2}{(1 + (0.8)^2 + (0.8)^4)} = -0.3122 \\
 \partial(3) &= \frac{-(-\theta^3)}{(1 + \theta^2 + \theta^4 + \theta^6)} = \frac{-(-0.8)^3}{(1 + (0.8)^2 + (0.8)^4 + (0.8)^6)} = 0.22147
 \end{aligned}$$

Exercice 22 Soit $(X_t, t \in \mathbb{Z})$ un processus MA(2) défini par

$$(1 + 2.4B + 0.8B^2) \epsilon_t = X_t$$

où $\epsilon_t \sim \text{bb}(0,1)$.

- Calculer leurs autocovariances.

Solution 23 Puisque le processus est par définition stationnaire, on a donc

$$\gamma(0) = [1 + \theta_1^2 + \theta_2^2] \sigma_\epsilon^2$$

$$\gamma(1) = [\theta_1 + \theta_1\theta_2] \sigma_\epsilon^2$$

$$\gamma(2) = \theta_2\sigma_\epsilon^2$$

$$\gamma(3) = \gamma(4) = \dots = 0.$$

Maintenant, il suffit de remplacer θ_1 par 2.4 et θ_2 par 0.8, ce qui nous donne

$$\gamma(0) = 7.4, \gamma(\pm 1) = 4.32, \gamma(\pm 2) = 0.8$$

et

$$\gamma(j) = 0 \text{ pour } |j| > 2.$$

1. Trajectoire, corrélogramme et corrélogramme partiel du processus $MA(1)$:

$$X_t = \epsilon_t + 0.8\epsilon_{t-1}$$

• **Script R**

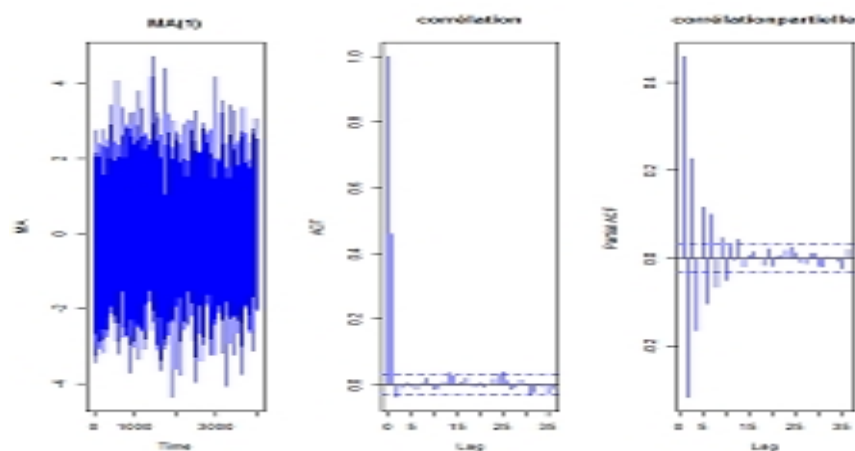
```
MA<-arima.sim(list(ma=0.8),n=4000)
```

```
par(mfrow=c(1,3))
```

```
plot(MA,col="blue",main="MA(1)")
```

```
acf(MA,col="blue",main="corrélation")
```

```
pacf(MA,col="blue",main="corrélationspartielle")
```



On peut calculer ces coefficients à l'aide du logiciel **R**, tout en utilisant la commande `ARMAacf` avec l'option (`pacf=TRUE`)

- **Script R**

```
ARMAacf(ma=0.8,lag.ma=6,pacf=T)
```

```
≠[1]0.4878049-0.31225600.2214778-0.16519350.1266695-0.0987133
```

2. Trajectoire, corrélogramme et corrélogramme partiel du processus $MA(1)$:

$$X_t = \epsilon_t - 0.6\epsilon_{t-1}$$

- **Script R**

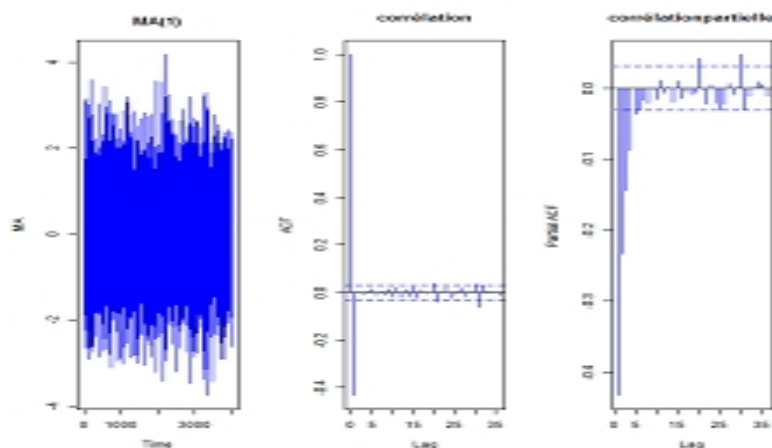
```
MA<-arima.sim(list(ma=-0.6),n=4000)
```

```
par(mfrow=c(1,3))
```

```
plot(MA,col="blue",main="MA(1)")
```

```
acf(MA,col="blue",main="corrélation")
```

```
pacf(MA,col="blue",main="corrélationpartielle")
```



On calcule les 5 premiers coefficients partiels à l'aide du logiciel R,

- **Script R**

```
>ARMAacf(ma=-0.6,lag.ma=5,pacf=T)
```

```
[1] -0.44117647-0.24167562-0.14060157-0.08344858-0.04987497
```

On remarque que les coefficients d'autocorrélation partielle dans cet exemple, décroissent exponentiellement en fonction de h et sont tous négatifs car $\theta < 0$, par contre dans l'exemple de

$$X_t = \epsilon_t + 0.8\epsilon_{t-1},$$

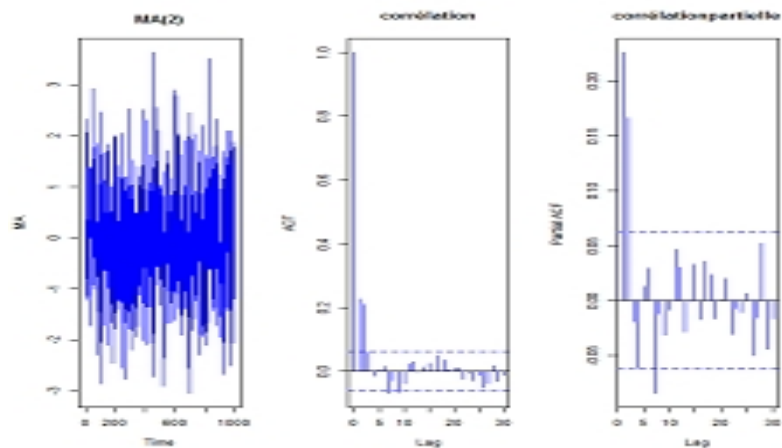
ils décroissent exponentiellement et sont alternés en signe ($\theta > 0$).

3. Trajectoire, corrélogramme et corrélogramme partiel du processus $MA(2)$:

$$X_t = \epsilon_t + 0.15\epsilon_{t-1} + 0.25\epsilon_{t-2}$$

- **Script R**

```
MA<-arima.sim(list(order=c(0,0,2),ma=c(0.15,0.25)),n=1000)
par(mfrow=c(1,3))
plot(MA,col="blue",main="MA(2)")
acf(MA,col="blue",main="corrélation")
pacf(MA,col="blue",main="corrélationpartielle")
```

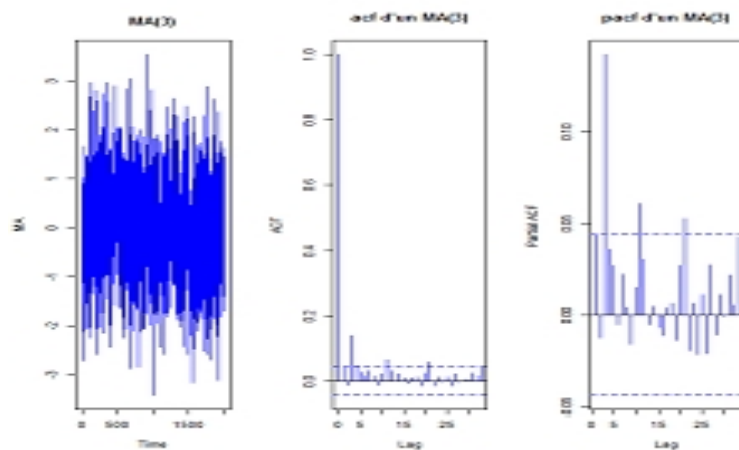


4. Trajectoire, corrélogramme et corrélogramme partiel du processus $MA(3)$:

$$X_t = \epsilon_t + 0.12\epsilon_{t-3}$$

- **Script R**

```
MA<-arima.sim(list(order=c(0,0,3),ma=c(0,0,0.12)),n=2000)
par(mfrow=c(1,3))
plot(MA,col="blue",main="MA(3)")
acf(MA,col="blue",main="acf d'un MA(3)")
pacf(MA,col="blue",main="pacf d'un MA(3)")
```



3.4 Stationnarité et causalité

La stationnarité d'une série temporelle est une propriétés fondamentale. Elle indique si les caractéristiques de celui-ci changent avec le temps ou non. Dans cette section, nous passons à l'étude probabiliste des processus $MA(q)$. En particulier, la stationnarité, la causalité et l'inversibilité.

Soit $(X_t, t \in \mathbb{Z})$ un processus $MA(q)$

$$X_t = \sum_{i=0}^q \theta_i \epsilon_{t-i}$$

$(X_t, t \in \mathbb{Z})$ est stationnaire par définition car :

1. $E(X_t) = E\left[\sum_{i=0}^q \theta_i \epsilon_{t-i}\right] = \sum_{i=0}^q \theta_i E(\epsilon_{t-i}) = 0$ ne dépend pas de t .
2. $var(X_t) = \sigma_\epsilon^2 \sum_{i=0}^q \theta_i^2$ ne dépend pas de t .
3. $\gamma(h) = \begin{cases} \sigma_\epsilon^2 \sum_{i=0}^{q-|h|} \theta_{|h|+i} \theta_i & \text{si } |h| \leq q, h \in \mathbb{Z} \\ 0 & \text{sinon} \end{cases}$ ne dépend de t .

Définition 24 Un processus est dit **causal** s'il existe une suite de constantes réelle (a_k) telle que

$$\sum_{k=0}^{\infty} |a_k| < \infty \text{ et } X_t = \sum_{k=0}^{\infty} a_k \epsilon_{t-k}.$$

Proposition 25 Parfois lorsque l'on parle d'un processus causal, on dit que ceci a une représentation $MA(\infty)$, c'est-à-dire,

$$X_t = \sum_{k=0}^{\infty} a_k \epsilon_{t-k}.$$

Alors, $MA(q)$ est un $MA(\infty)$ où

$$a_k = \begin{cases} \theta_k & \text{si } 0 \leq k \leq q \\ 0 & \text{sinon} \end{cases}$$

donc un processus moyenne mobile d'ordre q est toujours causal.

3.5 Inversibilité

Définition 26 Un processus est dit **inversible**, s'il existe une suite de constantes réelle (b_k) telle que

$$\sum_{k=0}^{\infty} |b_k| < \infty \text{ et } \epsilon_t = \sum_{k=0}^{\infty} b_k X_{t-k}.$$

Proposition 27 *On dit qu'un processus est inversible, s'il admet une représentation AR (∞), c'est-à-dire,*

$$\epsilon_t = \sum_{k=0}^{\infty} b_k X_{t-k}.$$

Maintenant, on détermine les conditions d'inversibilité d'un processus $(X_t, t \in \mathbb{Z})$ moyenne mobile d'ordre q défini par

$$X_t = \Theta(B)\epsilon_t.$$

En effet, soient $\omega_1, \dots, \omega_q$ les racines de l'équation

$$1 - \theta_1 B - \dots - \theta_q B^q = 0.$$

On suppose l'absence de la racine double, d'où

$$X_t = (1 - \omega_1^{-1}B) (1 - \omega_2^{-1}B) \dots (1 - \omega_q^{-1}B) \epsilon_t$$

avec

$$\begin{aligned} \epsilon_t &= (1 - \omega_1^{-1}B)^{-1} (1 - \omega_2^{-1}B)^{-1} \dots (1 - \omega_q^{-1}B)^{-1} X_t \\ \epsilon_t &= \left(\sum_{i=0}^{\infty} \omega_1^{-i} B^i \right) \dots \left(\sum_{i=0}^{\infty} \omega_q^{-i} B^i \right) X_t \\ \epsilon_t &= \sum_{s=0}^{\infty} \left(\sum_{s=i_1+\dots+i_q} \omega_1^{-i_1} \dots \omega_q^{-i_q} \right) X_{t-s}. \end{aligned}$$

Pour que cette dernière écriture ait un sens, il faut que les sommes convergent ce qui implique que les racines $\omega_1, \dots, \omega_q$, éventuellement complexes, aient un module supérieur à l'unité ou, en d'autres termes, qu'elles soient situées à l'extérieur du cercle unitaire : $\|\omega_i\| > 1, \forall i \in \overline{0, q}$.

Il est également possible d'exprimer les conditions d'inversibilité non pas sur les racines du polynôme $1 - \theta_1 B - \dots - \theta_q B^q = 0$ mais directement sur les coefficients $\theta_1, \dots, \theta_q$ du processus.

Exemple 28 *Soit le processus MA (1) qui est défini par*

$$X_t = \epsilon_t + \theta \epsilon_{t-1}$$

ou

$$X_t = (1 - \theta B) \epsilon_t$$

la racine de $1 - \theta\omega = 0$ est $\omega = \frac{1}{\theta}$ (ω est nécessairement un réel). D'après la condition d'inversibilité nous avons :

$$|\omega| > 1 \iff |\theta| < 1.$$

D'où

$$\epsilon_t = (1 - \theta B)^{-1} X_t = \sum_{i=0}^{\infty} \theta^i B^i X_t = \sum_{i=0}^{\infty} \theta^i X_{t-i}$$

la représentation est alors inversible.

Exemple 29 Soit un processus MA (2) donné par l'équation

$$X_t = \epsilon_t + 0.5\epsilon_{t-1} + 0.3\epsilon_{t-2},$$

il est stationnaire toujours par définition mais il est inversible ssi les racines de l'équation caractéristique

$$1 - \theta_1 B - \theta_2 B^2 = 0$$

sont en dehors du cercle unité. En effet,

$$\begin{aligned} X_t &= (1 - \omega_1^{-1} B) (1 - \omega_2^{-1} B) \\ &= \left(1 - (\omega_1^{-1} + \omega_2^{-1}) B + \omega_1^{-1} \omega_2^{-1} B^2\right) \epsilon_t \\ &= (1 - \theta_1 B - \theta_2 B^2) \epsilon_t. \end{aligned}$$

D'où l'on tire :

$$\begin{aligned} \omega_1^{-1} &= \frac{\theta_1 - \sqrt{\theta_1^2 + 4\theta_2}}{-2\theta_2} \text{ et } \omega_2^{-1} = \frac{\theta_1 + \sqrt{\theta_1^2 + 4\theta_2}}{-2\theta_2} \\ \omega_1^{-1} &= \frac{\theta_1 + \sqrt{\theta_1^2 + 4\theta_2}}{-2\theta_2} \text{ et } \omega_2^{-1} = \frac{\theta_1 - \sqrt{\theta_1^2 + 4\theta_2}}{-2\theta_2} \end{aligned}$$

et on peut montrer que :

$$\begin{cases} \omega_1^{-1} > 1 \\ \omega_2^{-1} > 1 \end{cases} \Rightarrow \begin{cases} \theta_1 + \theta_2 < 1 \\ \theta_2 - \theta_1 < 1 \\ -1 < \theta_2 < 1 \end{cases}$$

On remarquera encore que les solutions ω_1 et ω_2 seront complexes si $\theta_1^2 + 4\theta_2 < 0$, c'est-à-dire, si le déterminant du polynôme caractéristique est négatif. Or

$$\begin{cases} \theta_1 + \theta_2 = 0.5 + 0.3 = 0.8 < 1 \\ \theta_2 - \theta_1 = 0.3 - 0.5 = -0.2 < 1 \\ -1 < \theta_2 = 0.3 < 1 \end{cases}$$

Donc, ce processus est bien inversible.

Remarque 30 On peut tester si un processus est inversible ou non par logiciel R en utilisant la fonction `causal.inversible` qui est téléchargeable depuis l'adresse suivante : <http://hamrita.e-monsite.com/medias/files/causal-inversible.txt>.

- **Script**

```
>source("http://hamrita.e-monsite.com/medias/files/causal-inversible
.txt")
```

```
>causal.inversible(ma=c(0.5,0.3))
```

Résultat : Le processus est causal et inversible.

3.6 Processus Autorégressifs

Dans cette section, nous considérons une classe de processus linéaires qui est très importante dans la modélisation des séries chronologiques : les processus *AR*. Ces processus sont linéaires et, sous certaines conditions, stationnaires. Leur linéarité fournira une théorie simple de la prévision. Pour étudier cette classe importante de processus, nous allons tout d'abord commencer par les fonctions caractéristiques de ce processus en donnant leurs définitions avec une représentation de leurs différents trajectoires.

Introduisons les processus autorégressifs qui sont au coeur de ce cours. La définition générale d'un processus *AR* est la suivante :

Définition 31 On dit que $(X_t, t \in \mathbb{Z})$ est un processus autorégressif d'ordre p noté $AR(p)$ s'il s'écrit sous la forme

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \epsilon_t \quad \text{pour tout } t \in \mathbb{Z} \quad (3.1)$$

où les ϕ_i sont des réels et $(\epsilon_t, t \in \mathbb{Z})$ est un bruit blanc de variance σ_ϵ^2 . L'équation (3.1) est équivalente à l'écriture

$$\Phi(B) X_t = \epsilon_t \quad \text{où } \Phi(B) = 1 - \phi_1 B - \dots - \phi_p^p B$$

où B est l'opérateur de retard.

3.7 Fonctions caractéristiques d'un $AR(p)$

On considère un processus $(X_t, t \in \mathbb{Z})$ représenté par un $AR(p)$ tel que :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t.$$

3.7.1 Fonction d'autocovariance

On cherche tout d'abord à déterminer la fonction d'autocovariance

$$\gamma(h) = \text{Cov}(X_t, X_{t-h}) = E(X_t X_{t-h}), \forall t \in \mathbb{Z} \quad (3.2)$$

Proposition 32 La fonction d'autocovariance $\gamma(h)$ d'un processus $AR(p)$ satisfait une relation de récurrence de la forme

$$\gamma(h) = \begin{cases} \phi_1 \gamma(1) + \phi_2 \gamma(2) + \dots + \phi_p \gamma(p) + \sigma_\epsilon^2 & \text{si } h = 0 \\ \phi_1 \gamma(h-1) + \phi_2 \gamma(h-2) + \dots + \phi_p \gamma(h-p) & \text{si } h > 0 \end{cases} \quad (3.3)$$

où $h \in \mathbb{Z}$.

Preuve. On considère la définition suivante

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t.$$

D'après la définition (3.2), on a $\forall h > 0$

$$\begin{aligned} \gamma(h) &= \text{Cov}(X_t, X_{t-h}) = E(X_t X_{t-h}) \\ &= \phi_1 E(X_{t-1} X_{t-h}) + \phi_2 E(X_{t-2} X_{t-h}) + \dots + \phi_p E(X_{t-p} X_{t-h}) + E(\epsilon_t X_{t-h}) \\ &= \phi_1 \gamma(h-1) + \phi_2 \gamma(h-2) + \dots + \phi_p \gamma(h-p) \end{aligned}$$

puisque X_{t-h} ne dépend que des ϵ_{t-h-i} avec $i \geq 0$ alors $E(\epsilon_t X_{t-h}) = 0$. De la même façon :

$$\begin{aligned} \gamma(0) &= E(X_t^2) \\ &= \phi_1 E(X_{t-1} X_t) + \phi_2 E(X_{t-2} X_t) + \dots + \phi_p E(X_{t-p} X_t) + E(\epsilon_t X_t) \\ &= \phi_1 \gamma(-1) + \phi_2 \gamma(-2) + \dots + \phi_p \gamma(-p) + E(\epsilon_t X_t) \\ &= \phi_1 \gamma(1) + \phi_2 \gamma(2) + \dots + \phi_p \gamma(p) + \sigma_\epsilon^2. \end{aligned}$$

■

3.7.2 Fonctions d'autocorrélation et d'autocorrélation partielle

On sait par définition que

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

Proposition 33

1. La fonction d'autocorrélation d'un processus $AR(p)$, notée $\rho(h)$, est donnée par

$$\rho(h) = \begin{cases} 1 & \text{si } h = 0 \\ \phi_1\rho(h-1) + \phi_2\rho(h-2) + \dots + \phi_p\rho(h-p) & \text{si } h \neq 0 \end{cases} \quad (3.4)$$

Ces relations sont connues sous le nom d'équations de Yule-Walker.

2. La fonction d'autocorrélation partielle est nulle pour $|h| > p$ et vaut ϕ_p pour $|h| = p$.

Preuve. 1. Le processus $(X_t, t \in \mathbb{Z})$ s'écrit comme suit

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t.$$

En multipliant par X_t , on obtient

$$\begin{aligned} X_t^2 &= \phi_1 X_{t-1} X_t + \phi_2 X_{t-2} X_t + \dots + \phi_p X_{t-p} X_t + \epsilon_t X_t \\ X_t^2 &= \phi_1 X_{t-1} X_t + \phi_2 X_{t-2} X_t + \dots + \phi_p X_{t-p} X_t \\ &\quad + \epsilon_t (\phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t) \\ &= \phi_1 X_{t-1} X_t + \phi_2 X_{t-2} X_t + \dots + \phi_p X_{t-p} X_t \\ &\quad + \epsilon_t^2 + (\phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p}) \epsilon_t, \end{aligned}$$

en prenant l'espérance

$$\gamma(0) = \phi_1 \gamma(1) + \phi_2 \gamma(2) + \dots + \phi_p \gamma(p) + \sigma_\epsilon^2 + 0$$

le dernier terme est nul car ϵ_t est indépendant de $\{X_{t-1}, X_{t-2}, \dots, X_{t-p}\}$. De plus, en multipliant (3.1) par X_{t-h} , en prenant l'espérance et en divisant par $\gamma(0)$, on obtient

$$\rho(h) = \sum_{i=1}^p \phi_i \rho(h-i)$$

pour tout $h \neq 0$. Cette suite d'équations définit les équations de Yule-Walker.

2. Il suffit d'identifier membres à membres les termes de la fonction ACF et celle du processus $(X_t, t \in \mathbb{Z})$ ce qui peut se récrire sous la forme

$$X_{t+1} - m = \phi_1 (X_t - m) + \phi_2 (X_{t-1} - m) + \dots + \phi_p (X_{t-p+1} - m) + \epsilon_t,$$

le dernier coefficient de la projection linéaire de X_{t+1} sur les p plus récentes valeurs est égal à ϕ_p pour $|h| = p$. Aussi, si $h > p$, le coefficient de X_{t-h} est 0.

■

Proposition 34 Soit $(X_t, t \in \mathbb{Z})$ un processus $AR(p)$ d'autocorrélation $\rho(h)$. Alors

$$\begin{pmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(p-1) \\ \rho(1) & 1 & \rho(1) & \cdots & \rho(p-2) \\ \rho(2) & \rho(1) & 1 & \cdots & \rho(p-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho(p-1) & \rho(p-2) & \rho(p-3) & \cdots & 1 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{p-1} \\ \phi_p \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(p-1) \\ \rho(p) \end{pmatrix}.$$

De plus les $\rho(h)$ décroissent exponentiellement vers 0.

Preuve. En effet, $\forall h > 0$,

$$\rho(h) = \phi_1 \rho(h-1) + \phi_2 \rho(h-2) + \dots + \phi_p \rho(h-p).$$

Le polynôme caractéristique de cette relation de récurrence est

$$\begin{aligned} z^p - \phi_1 z^{p-1} - \phi_2 z^{p-2} - \dots - \phi_{p-1} z - \phi_p &= z^p \left(1 - \frac{\phi_1}{z} - \frac{\phi_2}{z^2} - \dots - \frac{\phi_{p-1}}{z^{p-1}} - \frac{\phi_p}{z^p} \right) \\ &= z^p \Phi \left(\frac{1}{z} \right), \end{aligned}$$

avec $\Phi(B) X_t = \epsilon_t$ où $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$. Les racines du polynôme caractéristique sont les $\lambda_i = \frac{1}{z_i}$ (les z_i étant les racines de Φ) avec $|\lambda_i| > 1$. La forme générale de la solution est, si z_1, \dots, z_n sont des racines distinctes de Φ de multiplicités respectives m_1, \dots, m_n

$$\rho(h) = \sum_{i=1}^n \sum_{k=0}^{m_i-1} \alpha_{ik} \lambda_i^k h^k$$

c'est-à-dire, $\rho(h)$ décroît vers 0 exponentiellement lorsque h tend vers l'infini. Pour plus de détails voir [2]. ■

Fonctions caractéristiques d'un AR(1)

Supposons que $(X_t, t \in \mathbb{Z})$ est un $AR(1)$. On peut donc écrire

$$X_t = \phi X_{t-1} + \epsilon_t$$

D'après les résultats de la section précédente, on peut écrire que si $|\phi| < 1$, on a

$$X_t = \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}$$

L'autocovariance est donnée par

$$\begin{aligned}\gamma(h) &= \text{Cov}(X_t, X_{t+h}) = \lim_{n \rightarrow \infty} E \left[\left(\sum_{i=0}^n \phi^i \epsilon_{t-i} \right) \left(\sum_{j=0}^n \phi^j \epsilon_{t+h-j} \right) \right] \\ &= \sigma_\epsilon^2 \phi^{|h|} \sum_{i=0}^{\infty} \phi^{2i} \text{ (série convergente car } |\phi| < 1) \\ &= \frac{\sigma_\epsilon^2 \phi^{|h|}}{1 - \phi^2}.\end{aligned}$$

On a aussi

$$\text{Var}(X_t) = \gamma(0) = \frac{\sigma_\epsilon^2}{1 - \phi^2} > \sigma_\epsilon^2.$$

Les fonctions d'autocorrélation et d'autocorrélation partielle sont données par

$$\rho(h) = \phi^{|h|} \text{ et } r(h) = \begin{cases} \phi & \text{si } |h| \leq 1 \\ 0 & \text{si } |h| \geq 2 \end{cases}$$

On obtient deux sortes de corrélogrammes suivant que ϕ est positif ou négatif. On voit ainsi que les variables X_t et X_{t-h} sont dépendantes avec une corrélation qui décroît exponentiellement vers 0 lorsque h tend vers l'infini.

Exercice 35 Soit $(X_t, t \in \mathbb{Z})$ un processus AR(2) défini par

$$(1 - 1.1B + 0.18B^2) X_t = \epsilon_t$$

où $\epsilon_t \sim \text{bb}(0, 1)$.

- Ce processus est-il stationnaire ? justifier
- Calculer leurs autocovariances.

Solution 36 Oui ce processus est stationnaire puisque :

$$1 - 1.1z + 0.18z^2 = (1 - 0.9z)(1 - 0.2z)$$

les valeurs propres (0.9 et 0.2) sont à l'intérieures du cercle unité.

Pour calculer les autocovariances, on va utiliser les formules suivantes

$$\begin{aligned}\gamma(j) &= \phi_1 \gamma(j-1) + \phi_2 \gamma(j-2) \\ \rho(j) &= \phi_1 \rho(j-1) + \phi_2 \rho(j-2).\end{aligned}$$

On a

$$\begin{aligned}\rho(1) &= \phi_1 + \phi_2 \rho(1) \Rightarrow \rho(1) = \frac{\phi_1}{1 - \phi_2} \\ \rho(2) &= \phi_1 \rho(1) + \phi_2 \Rightarrow \rho(2) = \frac{\phi_1^2}{1 - \phi_2} + \phi_2.\end{aligned}$$

De plus

$$\begin{aligned}\gamma(0) &= \phi_1\gamma(1) + \phi_2\gamma(2) + \sigma_\epsilon^2 \\ &= \phi_1\rho(1)\gamma(0) + \phi_2\rho(2)\gamma(0) + \sigma_\epsilon^2 \\ &= \left[\frac{\phi_1^2}{1-\phi_2} + \frac{\phi_2\phi_1^2}{1-\phi_2} + \phi_2^2 \right] \gamma(0) + \sigma_\epsilon^2 \\ \gamma(0) &= \frac{(1-\phi_2)\sigma_\epsilon^2}{(1+\phi_2)\left[(1-\phi_2)^2 - \phi_1^2\right]}.\end{aligned}$$

Maintenant, il suffit de remplacer ϕ_1 par 1.1 et ϕ_2 par -0.18 , ce qui nous donne

$$\gamma(0) = 7.89, \gamma(1) = 7.35,$$

et pour $j = 2, 3, \dots$

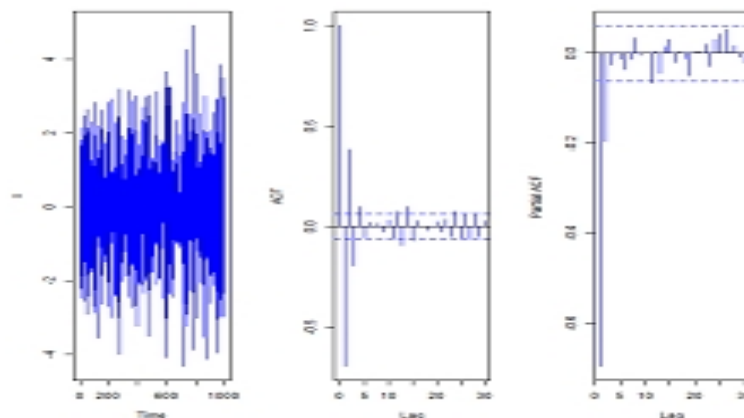
$$\begin{aligned}\gamma(j) &= 1.1\gamma(j-1) - 0.18\gamma(j-2) \\ \gamma(j) &= \gamma(-j).\end{aligned}$$

1. Trajectoire, corrélogramme et corrélogramme partiel du processus $AR(2)$:

$$X_t = -0.8X_{t-1} - 0.15X_{t-2} + \epsilon_t$$

• **Script R**

```
Ar2<-arima.sim(list(order=c(2,0,0),ar=c(-.8,-0.15)),n=1000)
par(mfrow=c(1,3))
plot(Ar2,ylab="x",col="blue",main="")
acf(Ar2,col="blue",main="")
pacf(Ar2,col="blue",main="")
```

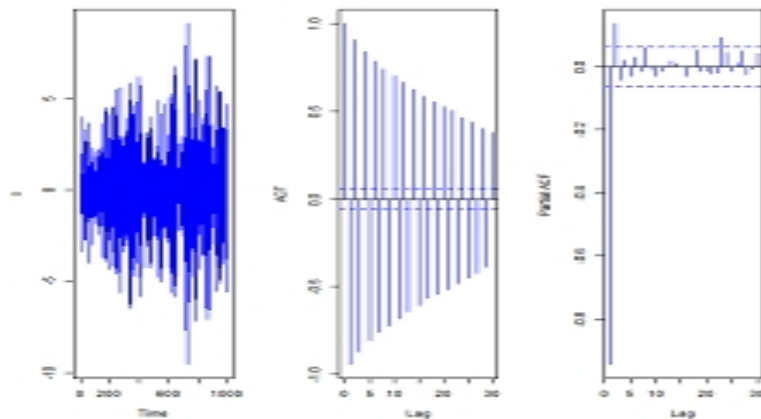


2. Trajectoire, corrélogramme et corrélogramme partiel du processus $AR(2)$:

$$X_t = -0.8X_{t-1} + 0.15X_{t-2} + \epsilon_t$$

- **Script R**

```
Ar2<-arima.sim(list(order=c(2,0,0),ar=c(-.8,0.15)),n=1000)
par(mfrow=c(1,3))
plot(Ar2,ylab="x",col="blue",main="")
acf(Ar2,col="blue",main="")
pacf(Ar2,col="blue",main="")
```

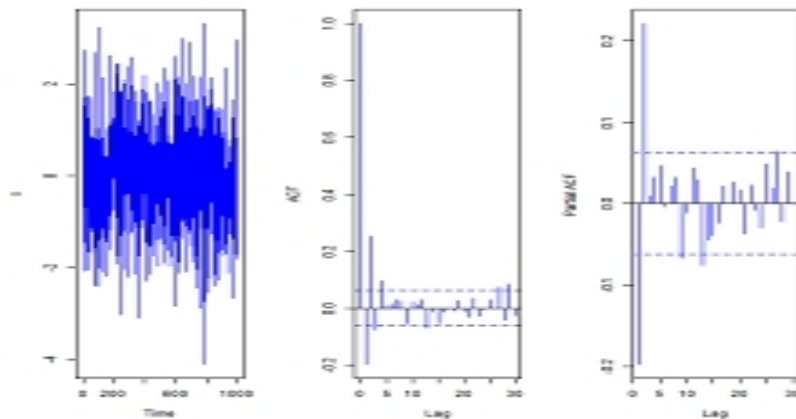


3. Trajectoire, corrélogramme et corrélogramme partiel du processus $AR(2)$:

$$X_t = -0.2X_{t-1} + 0.2X_{t-2} + \epsilon_t$$

- **Script R**

```
Ar2<-arima.sim(list(order=c(2,0,0),ar=c(-.2,0.2)),n=1000)
par(mfrow=c(1,3))
plot(Ar2,ylab="x",col="blue",main="")
acf(Ar2,col="blue",main="")
pacf(Ar2,col="blue",main="")
```

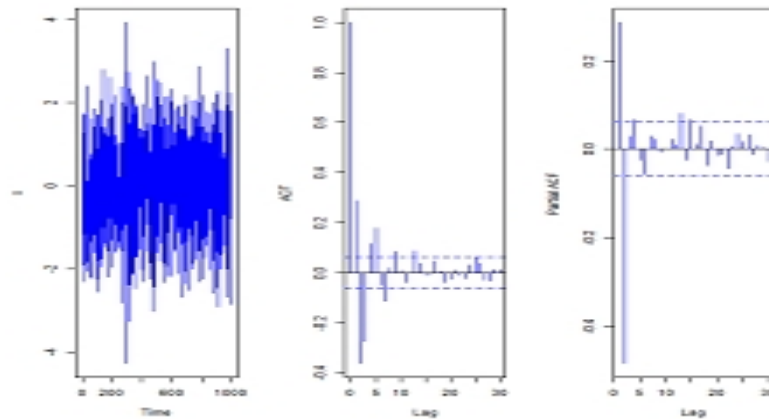


4. Trajectoire, corrélogramme et corrélogramme partiel du processus $AR(2)$:

$$X_t = 0.5X_{t-1} - 0.5X_{t-2} + \epsilon_t$$

• **Script R**

```
Ar2<-arima.sim(list(order=c(2,0,0),ar=c(0.5,-0.5)),n=1000)
par(mfrow=c(1,3))
plot(Ar2,ylab="x",col="blue",main="")
acf(Ar2,col="blue",main="")
pacf(Ar2,col="blue",main="")
```



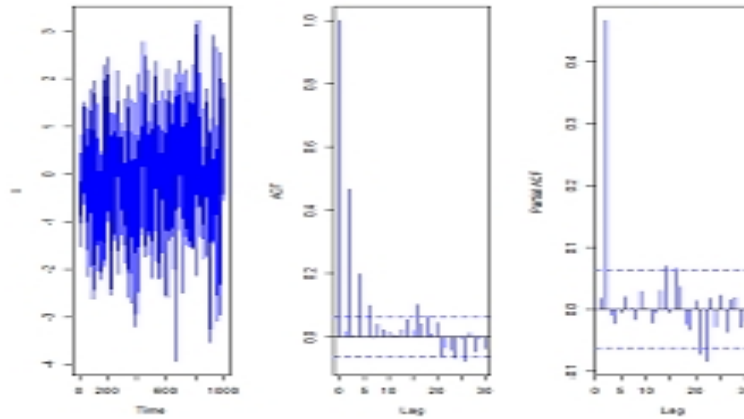
5. Trajectoire, corrélogramme et corrélogramme partiel du processus $AR(2)$:

$$X_t = 0.5X_{t-2} + \epsilon_t$$

• **Script R**

```
Ar2<-arima.sim(list(order=c(2,0,0),ar=c(0,0.5)),n=1000)
par(mfrow=c(1,3))
plot(Ar2,ylab="x",col="blue",main="")
acf(Ar2,col="blue",main="")
```

```
pacf(Ar2,col="blue",main="")
```



6. Trajectoire, corrélogramme et corrélogramme partiel du processus $AR(2)$:

$$X_t = 0.5X_{t-1} + \epsilon_t$$

• **Script R**

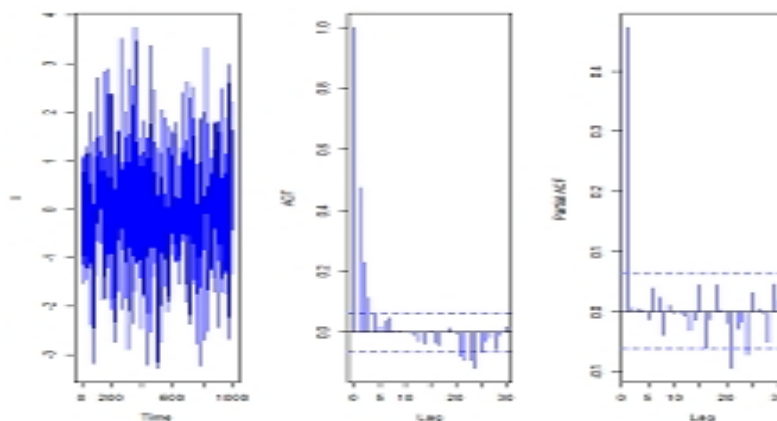
```
Ar2<-arima.sim(list(order=c(2,0,0),ar=c(0.5,0)),n=1000)
```

```
par(mfrow=c(1,3))
```

```
plot(Ar2,ylab="x",col="blue",main="")
```

```
acf(Ar2,col="blue",main="")
```

```
pacf(Ar2,col="blue",main="")
```



7. Trajectoire, corrélogramme et corrélogramme partiel du processus $AR(2)$:

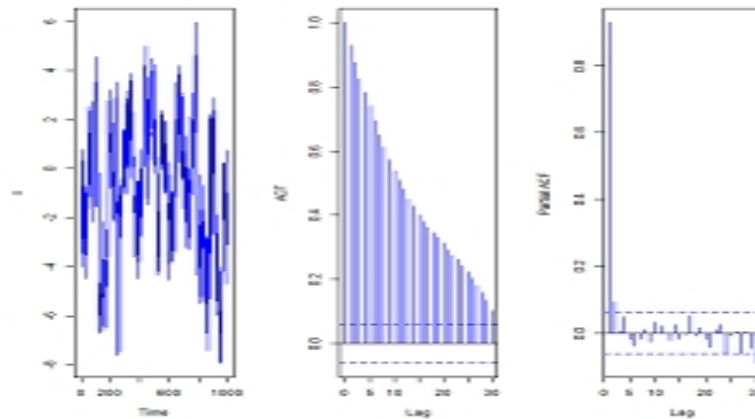
$$X_t = 0.8X_{t-1} + 0.15X_{t-2} + \epsilon_t$$

• **Script R**

```
Ar2<-arima.sim(list(order=c(2,0,0),ar=c(0.8,0.15)),n=1000)
```

```
par(mfrow=c(1,3))
```

```
plot(Ar2,ylab="x",col="blue",main="")
acf(Ar2,col="blue",main="")
pacf(Ar2,col="blue",main="")
```

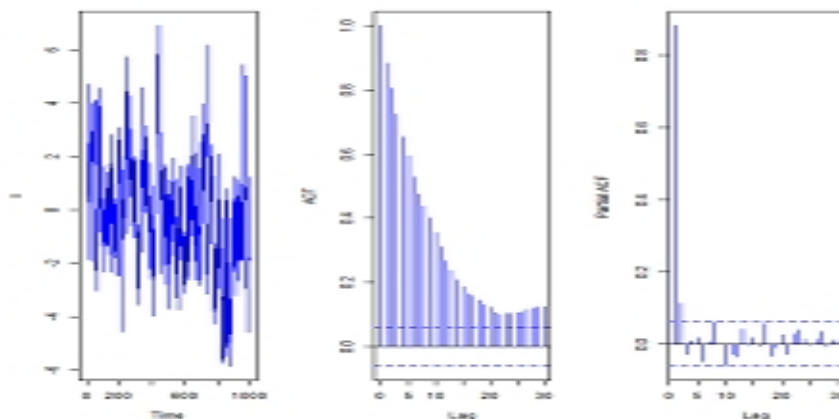


8. Trajectoire, corrélogramme et corrélogramme partiel du processus $AR(2)$:

$$X_t = 0.8X_{t-1} + 0.1X_{t-2} + \epsilon_t$$

• **Script R**

```
Ar2<-arima.sim(list(order=c(2,0,0),ar=c(0.8,0.1)),n=1000)
par(mfrow=c(1,3))
plot(Ar2,ylab="x",col="blue",main="")
acf(Ar2,col="blue",main="")
pacf(Ar2,col="blue",main="")
```



3.8 Stationnarité et causalité

Les processus AR forment une classe flexible de modèles pour de nombreux phénomènes observés. Ils sont définis implicitement par la relation (3.1). Deux notions importantes sont liées à ce type de modèles. La première est la stationnarité ou plus exactement l'existence d'un processus stationnaire $(X_t, t \in \mathbb{Z})$ satisfaisant au modèle. La seconde est la causalité, autrement dit, l'existence d'une solution relie X_t avec les valeurs passées des innovations $(\epsilon_{t-k})_{k>1}$. Cependant, contrairement aux processus moyenne mobile (MA), la définition (3.1) pose quelques problèmes : le processus ainsi défini est-t-il stationnaire ? Quelle est sa forme explicite ?

Pour illustrer ce problème, considérons tout d'abord le cas d'un processus $AR(1)$. On a donc

$$X_t = \phi X_{t-1} + \epsilon_t \quad (3.5)$$

et, en itérant cette équation réursive, on a, pour tout $K \in \mathbb{N}$

$$X_t = \phi^K X_{t-K} + \sum_{k=1}^K \phi^{k-1} \epsilon_{t-k+1}$$

Supposons que $|\phi| < 1$. Dans ce cas, on a

$$\begin{aligned} E \left\{ X_t - \sum_{k=1}^K \phi^{k-1} \epsilon_{t-k+1} \right\}^2 &= \phi^{2K} E \{ X_{t-K} \}^2 \\ &= \phi^{2K} cte, \end{aligned}$$

si X_t est stationnaire (c-à-d, de variance constante). Puisque nous avons supposé que $|\phi| < 1$, alors, on a $\phi^{2K} \rightarrow 0$ lorsque $K \rightarrow \infty$, ce qui nous donne

$$E \left\{ X_t - \sum_{k=1}^K \phi^{k-1} \epsilon_{t-k+1} \right\}^2 \rightarrow 0$$

lorsque $K \rightarrow \infty$. Ceci permet d'écrire

$$X_t = \sum_{k=1}^{\infty} \phi^{k-1} \epsilon_{t-k+1} = \sum_{j=1}^{\infty} \phi^j \epsilon_{t-j}. \quad (3.6)$$

Par le théorème de convergence monotone, on voit que $E \{ X_t \} = 0$. Pour la

fonction d'autocovariance, le même argument permet d'écrire

$$\begin{aligned}
 \text{Cov}(X_t, X_{t+h}) &= \lim_{K \rightarrow \infty} \left(\sum_{k=1}^K \phi^{k-1} \epsilon_{t-k+1}, \sum_{k=1}^K \phi^{k-1} \epsilon_{t+h-k+1} \right) \\
 &= \lim_{K \rightarrow \infty} \left(\sigma_\epsilon^2 \phi^{|h|} \sum_{k=1}^K \phi^{2k} \right) \\
 &= \sigma_\epsilon^2 \phi^{|h|} \sum_{k=1}^{\infty} \phi^{2k} \\
 &= \sigma_\epsilon^2 \frac{\phi^{|h|}}{1 - \phi^2}
 \end{aligned}$$

par la formule de la série d'une suite géométrique, qui converge puisque $|\phi| < 1$.

Maintenant, si $|\phi| > 1$, la série (3.6) ne converge pas. On peut cependant éviter cette difficulté en réécrivant le processus $AR(1)$ comme suit

$$X_t = \frac{1}{\phi} X_{t+1} - \frac{1}{\phi} \epsilon_{t+1}$$

Par les mêmes arguments utilisés pour le cas $|\phi| < 1$, on peut montrer que

$$X_t = - \sum_{j=1}^{\infty} \phi^{-j} \epsilon_{t+j}$$

est une solution de l'équation implicite et que la série converge. Cependant, cette solution relie X_t avec les valeurs futures des innovations $(\epsilon_{t+j})_{j>1}$. Cette solution n'est donc pas naturelle car le processus est corrélé avec les réalisations non encore observées des processus ϵ_t . On dit alors que la solution n'est pas causale.

Définition 37 Un processus $AR(p)$ est dit causal lorsqu'il existe une suite de nombres (α_k) telle que $\sum_{k \in \mathbb{Z}} |\alpha_k| < \infty$ et

$$X_t = \sum_{k=0}^{\infty} \alpha_k \epsilon_{t-k} \quad (3.7)$$

Nous pouvons à présent retenir de notre développement une condition nécessaire et suffisante pour qu'un processus $AR(1)$ soit stationnaire et causal est

$$|\phi| < 1$$

et dans ce cas, les coefficients α_k de sa représentation causale $MA(\infty)$ sont donnés par $\alpha_k = \phi^k$ ce que nous trouvons en comparant (3.7) avec l'équation (3.6).

Observons également que le cas $|\phi| = 1$ ne fournit pas de solution stationnaire ce qui s'explique de soi en vue de la forme de l'autocovariance.

La généralisation de ce résultat au cas où le modèle autorégressif est d'ordre $p > 1$ n'est pas directe. On peut réécrire (3.1)

$$X_t = \sum_{i=1}^p \phi_i (B^i X_t) + \epsilon_t$$

ou encore

$$\left(1 - \sum_{i=1}^p \phi_i B^i\right) X_t = \epsilon_t$$

c-à-d, on obtient

$$\Phi(B) X_t = \epsilon_t \quad (3.8)$$

où $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$.

Remarque 38 En toute généralité, un processus $AR(p)$ vérifie une relation de la forme

$$\Phi(B) X_t = \mu + \epsilon_t$$

où μ est un terme constant. De cette forme générale, il est possible de se ramener à (3.1) par une simple translation : il suffit de considérer non pas X_t mais $Y_t = X_t - m$ où $m = \frac{\mu}{\Phi(1)}$.

En effet,

$$\begin{aligned} \Phi(B)(Y_t + m) &= \mu + \epsilon_t \\ \Phi(B)Y_t + \Phi(1)m &= \mu + \epsilon_t \\ \Phi(B)Y_t &= \epsilon_t. \end{aligned}$$

m correspond ici à l'espérance de $(X_t, t \in \mathbb{Z})$.

Remarque 39 Si le processus $(X_t, t \in \mathbb{Z})$ est stationnaire, alors toutes les racines du polynôme $\Phi(B)$ sont strictement supérieures à 1 en module, ce qui implique en particulier que $\Phi(1) \neq 0$,

$$m = E(X_t) = \frac{\mu}{\Phi(1)} = \frac{\mu}{1 - \phi_1 - \dots - \phi_p} \quad (3.9)$$

Le résultat suivant donne les conditions nécessaires et suffisantes de stationnarité et de causalité des modèles autorégressifs.

Proposition 40 *Le processus (3.8) est causal et stationnaire si et seulement si le polynôme en z : $\Phi(z) = 1 - \sum_{i=1}^p \phi_i z^i$ est tel que*

$$\Phi(z) \neq 0 \text{ pour tout } z \in \mathbb{C} \text{ tel que } |z| \leq 1. \quad (3.10)$$

Les coefficients α_k apparaissant dans la représentation causale $MA(\infty)$ dans (3.7), sont déterminés par

$$1 - \sum_{i=1}^p \alpha_i z^i = \frac{1}{\Phi(z)} \quad (3.11)$$

si $|z| \leq 1$.

Preuve. Voir [2], théorème 3.1.1. ■

Regardons, pour motivation de ces conditions, Dans le cas d'un exemple $AR(1)$, l'étude de racine du polynôme générateur est particulièrement simple $\Phi(z) = 1 - \phi z$ a comme racine $z = \frac{1}{\phi}$. Alors, la condition $|\phi| < 1$ pour la stationnarité et la causalité d'un $AR(1)$ revient à la condition $|z| > 1$ ce qui n'est rien d'autre que notre condition (3.10) : $\Phi(z) = 1 - \phi z \neq 0$ pour tout $|z| \leq 1$.

Pour des processus $AR(p)$ avec $p > 1$, les considérations suivantes permettent d'établir la condition (3.10). On montre le principe pour $p = 2$:

Les racines $z_1, z_2 \neq 0$ du polynôme générateur d'un $AR(2)$

$$\Phi(z) = 1 - \phi_1 z - \phi_2 z^2 = \phi_2 (z - z_1) (z - z_2),$$

sont bien fonction des coefficients ϕ_1, ϕ_2 . Il est plus facile de donner la condition de stationnarité en fonction de ces racines z_1, z_2 , car l'équation d'un $AR(2)$; $\Phi(B) X_t = \epsilon_t$ se transforme comme suit

$$\Phi(B) X_t = z_1 z_2 \left(1 - \frac{B}{z_1}\right) \left(1 - \frac{B}{z_2}\right) X_t = \epsilon_t$$

d'où

$$\left(1 - \frac{B}{z_1}\right) \left(1 - \frac{B}{z_2}\right) X_t = \frac{\epsilon_t}{z_1 z_2} = \eta_t,$$

où $(\eta_t, t \in \mathbb{Z})$ est un bruit blanc de variance σ_η^2 . Donc, le processus défini par $Y_t = \left(1 - \frac{B}{z_2}\right) X_t$ est bien un processus $AR(1)$ avec paramètre $\frac{1}{z_1}$,

$$\left(1 - \frac{B}{z_1}\right) Y_t = \eta_t$$

qui est bien stationnaire et causal si $|z_1^{-1}| < 1 \iff |z_1| > 1$. De plus, le processus X_t défini par

$$\left(1 - \frac{B}{z_2}\right) X_t = Y_t$$

est bien un processus $AR(1)$ avec paramètre $\frac{1}{z_2}$, il existe une solution stationnaire et causale pour X_t si $|z_2^{-1}| < 1 \iff |z_2| > 1$. Alors, un processus $AR(2)$ est stationnaire et causal si les racines z_1, z_2 de son polynôme générateur $\Phi(z)$ vérifient

$$z_1, z_2 \in \{z : |z| > 1\}.$$

Le même raisonnement s'applique au processus $AR(p)$ pour $p > 2$.

Exemple 41 *Le processus $AR(2)$ suivant est stationnaire et causal :*

$$X_t = -\frac{1}{6}X_{t-1} + \frac{1}{6}X_{t-2} + \epsilon_t$$

où $(\epsilon_t, t \in \mathbb{Z})$ est un bruit blanc de variance σ_ϵ^2 . En effet, on peut le réécrire comme

$$\left(-\frac{1}{6}B^2 + \frac{1}{6}B + 1\right) X_t = \epsilon_t$$

nous avons aussi

$$\begin{aligned} \Phi(z) &= -\frac{1}{6}z^2 + \frac{1}{6}z + 1 \\ &= -\frac{1}{6}(z^2 - z - 6) \\ &= -\frac{1}{6}(z+2)(z-3) \end{aligned}$$

Les racines sont à l'extérieur du cercle unité, donc le processus est stationnaire et causal.

3.9 Inversibilité

Par définition, on a directement que tout processus autorégressif est automatiquement inversible. L'équation

$$\epsilon_t = \sum_{i=0}^{\infty} \phi_i X_{t-i} \quad \text{pour tout } t \in \mathbb{Z}$$

est parfois qualifiée de "représentation $AR(\infty)$ " du processus. C'est d'ailleurs une des motivations principales pour l'étudier : Un processus linéaire qui est

inversible permet une telle représentation $AR(\infty)$ avec des coefficients ϕ_i rapidement croissants. Alors, il existe une bonne approximation par un processus $AR(p)$, avec un ordre p assez petit. Celui-ci permettra de traiter le processus linéaire donné dans le cadre simple et efficace de l'estimation et de la prédiction des processus autorégressifs d'ordre p .

Chapitre 4

Processus mixtes : ARMA

Les processus AR et MA ont des caractéristiques qui se révèlent grâce à leurs fonctions d'autocorrélations et leurs fonctions d'autocorrélations partielles. Pour un processus AR , nous avons vu que la fonction d'autocorrélation partielle possède un point de rupture après un certain nombre d'écarts : ce dernier détermine l'ordre du polynôme AR . Pour un processus MA , nous avons vu que c'est la fonction d'autocorrélation qui possède un point de rupture après un certain nombre d'écarts : ce dernier détermine l'ordre du polynôme MA . Cependant pour certains processus, ni la fonction d'autocorrélation, ni la fonction d'autocorrélation partielle ne possèdent de point de rupture. Dans de tels cas, il faut construire un modèle mixte.

Nous définissons dans ce chapitre les séries $ARMA$ qui sont des généralisations directes des deux exemples introductifs, la combinaison des processus autorégressifs et moyennes mobiles. Cette classe de processus $ARMA$ est encore un cas particulier de processus linéaires et jouera un rôle prépondérant dans la modélisation concrète des processus stationnaires. Elle présente l'avantage d'être plus souple à l'utilisation et de fournir généralement de bonnes approximations des séries réelles avec moins de paramètres que les modèles purs.

Pour étudier cette classe importante de processus, nous allons tout d'abord commencer par les fonctions caractéristiques de ce processus en donnant leurs définitions avec une représentation de leurs différents trajectoires.

Définition 42 Un processus $(X_t, t \in \mathbb{Z})$ est un processus $ARMA$ d'ordre p et q noté $ARMA(p, q)$, s'il est défini par :

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \epsilon_t, t \in \mathbb{Z}$$

où les ϕ_i et les θ_j sont des réels et le processus $(\epsilon_t, t \in \mathbb{Z})$ est un bruit blanc de variance σ_ϵ^2 .

Une représentation équivalente est la suivante

$$\Phi(B)X_t = \Theta(B)\epsilon_t,$$

où Φ et Θ sont des polynômes de degrés respectifs p et q tels que

$$\Phi(B) = 1 - \sum_{i=1}^p \phi_i B^i$$

et

$$\Theta(B) = 1 + \sum_{j=1}^q \theta_j B^j$$

et B est l'opérateur de "retard" défini par

$$B^k X_t = X_{t-k}$$

- Les processus *AR* et *MA* sont des processus *ARMA* particuliers.
- Un processus *ARMA* (p, q) est aussi appelé **processus mixte autorégressif moyenne mobile**.
- Il est immédiat qu'un *ARMA* $(p, 0)$ est un *AR* pur et qu'un *ARMA* $(0, q)$ est un *MA* pur. Les seuls processus admettant simultanément une représentation *MA* pure et une représentation *AR* pure correspondent au cas *ARMA* $(0, 0)$, c'est-à-dire, aux bruits blancs.

Remarque 43 Soit $(X_t, t \in \mathbb{Z})$ un processus *ARMA* (p, q)

$$\Phi(B)X_t = \Theta(B)\epsilon_t.$$

La représentation est :

- *minimale* si Φ et Θ n'ont pas de facteurs communs.
- *causale* si Φ a toutes ses racines à l'extérieur du cercle unité.
- *inversible* si Θ a toutes ses racines à l'extérieur du cercle unité.
- *canonique* si la représentation est causale et inversible.

Proposition 44 Le bruit blanc associé à un processus *ARMA* sous sa forme canonique est le processus d'innovation.

Proposition 45 Soit $(X_t, t \in \mathbb{Z})$ un processus ARMA(p, q) canonique : $\Phi(B)X_t = \Theta(B)\epsilon_t$. Si de plus il est minimale alors :

- Il admet une écriture MA (∞)

$$\begin{aligned} X_t &= \Phi^{-1}(B)\Theta(B)\epsilon_t \\ &= \epsilon_t + \sum_{i=1}^{\infty} \psi_i \epsilon_{t-i} \end{aligned}$$

où $(\psi_i)_{i \in \mathbb{N}}$ est une suite réelle.

En posant $\psi_i = 0$ pour $i < 0$, $\theta_0 = 1$ et $\theta_i = 0$ pour $i > q$, on a :

$$\forall i \in \mathbb{N} : \psi_i - \sum_{j=1}^p \phi_j \psi_{i-j} = \theta_i.$$

- Il admet une écriture AR (∞)

$$\epsilon_t = \Theta^{-1}(B)\Phi(B)X_t = X_t + \sum_{i=1}^{\infty} \pi_i X_{t-i} \Leftrightarrow X_t = - \sum_{i=1}^{\infty} \pi_i X_{t-i} + \epsilon_t$$

où $(\pi_i)_{i \in \mathbb{N}}$ est une suite réelle.

En posant $\pi_i = 0$ pour $i < 0$, $\phi_0 = -1$ et $\phi_i = 0$ pour $i > p$, on a :

$$\forall i \in \mathbb{N} : \pi_i - \sum_{j=1}^p \theta_j \pi_{i-j} = -\phi_i.$$

4.1 Fonctions caractéristiques

4.1.1 Fonction d'autocovariance

Proposition 46 La fonction d'autocovariance γ_k d'un processus stationnaire ARMA(p, q) $(X_t, t \in \mathbb{Z})$ satisfait une relation de récurrence de la forme :

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p} \quad \forall k > p.$$

On a donc la même relation de récurrence pour un AR(p) (équations de Yule Walker), mais cette dernière n'est valable que pour des ordres supérieurs à q . Cette relation n'est pas valable pour $k \leq q$ en raison de la corrélation entre X_{t-j} et ϵ_{t-j} .

On peut déterminer les autocorrélations simples à l'aide de deux calculs différents

- On obtient à l'aide de la représentation $MA(\infty)$:

$$\begin{aligned}\gamma_h &= \text{Cov}(X_t, X_{t-h}) \\ &= \text{Cov}\left(\epsilon_t + \sum_{i=1}^{\infty} \psi_i \epsilon_{t-i}, \epsilon_{t-h} + \sum_{i=1}^{\infty} \psi_i \epsilon_{t-h-i}\right) \\ &= \sigma_\epsilon^2 \sum_{i=1}^{\infty} \psi_i \psi_{i+h} \text{ où } \psi_0 = 1.\end{aligned}$$

- On a :

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}.$$

D'où

$$\begin{aligned}\gamma_h - \phi_1 \gamma_{h-1} - \dots - \phi_p \gamma_{h-p} &= \text{Cov}(\epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}, X_{t-h}) \\ &= \text{Cov}\left(\epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}, \sum_{i=0}^{\infty} \psi_i \epsilon_{t-h-i}\right) \\ &= \begin{cases} \sigma_\epsilon^2 \sum_{i=0}^{\infty} \theta_{h+i} \psi_i & \text{si } h \in \{0, \dots, q\} \\ 0 & \text{si } h > q \end{cases}\end{aligned}$$

Remarque 47 Les autocorrélations simples décroissent vers 0.

- Si $p > q$ la décroissance est de type exponentiel ou sinusoidal amorti.
- Si $q \geq p$, les $q - p - 1$ premières valeurs ont un comportement quelconque et les suivantes décroissent.

4.1.2 Densité spectrale d'un processus ARMA

Proposition 48 Soit $(X_t, t \in \mathbb{Z})$ un processus ARMA(p, q) (pas forcément sous sa forme canonique), $\Phi(B)X_t = \Theta(B)\epsilon_t$. Sa densité spectrale vaut :

$$f(w) = \frac{\sigma_\epsilon^2 |\Theta(\exp(-iw))|^2}{2\pi |\Phi(\exp(-iw))|^2}.$$

4.1.3 Fonction d'autocorrélation

Proposition 49 La fonction d'autocorrélation ρ_k d'un processus stationnaire ARMA(p, q) satisfait une relation de récurrence de la forme :

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad \forall k > p.$$

Comme le cas d'un $AR(p)$, on peut obtenir une solution à cette équation. Lorsque les p racines sont distinctes, cette solution est de la forme

$$\rho_k = A_1 \left(\frac{1}{\lambda_1} \right)^k + A_2 \left(\frac{1}{\lambda_2} \right)^k + \dots + A_p \left(\frac{1}{\lambda_p} \right)^k \quad \forall k > p$$

où les paramètres $(A_i)_{i=1}^p$ sont des constantes déterminées par les conditions initiales et les paramètres $(\lambda_i)_{i=1}^p$ désignent les p racines distinctes du polynôme associé à la composante autoregressive du processus : $\Phi(B) = 0$. Mais dans ce cas les valeurs initiales $\{\rho_1 \dots \rho_q\}$ sont différentes de celles obtenues pour un $AR(p)$ et les constantes $(A_i)_{i=1}^p$ sont donc elles mêmes différentes.

Remarque 50 Ainsi, le corrélogramme d'un processus $ARMA(p, q)$ est le même que celui d'un AR à partir de l'ordre $q + 1$.

• Script R

```
par(mfrow=c(1,2))
library(stats)
ts.arma <- arima.sim(list(order=c(1,0,1), ar=0.8, ma=0.6), n=100)
ts.plot(ts.arma, type="l", col="4")
acf(ts.arma, 10, col="4", "correlation")
```

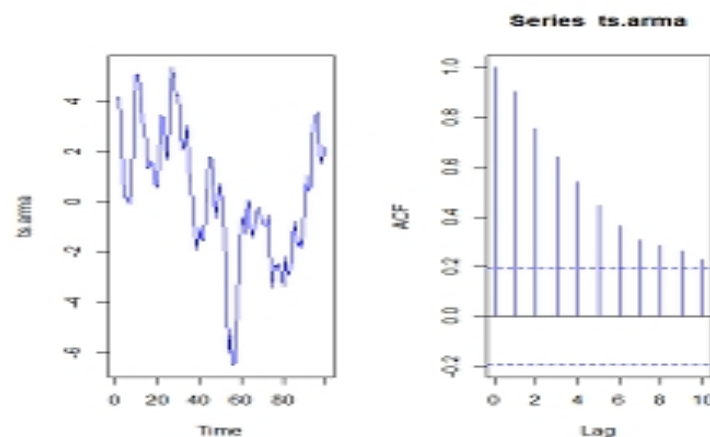


Figure 6. Trajectoire du processus $ARMA(1, 1)$: $X_t = 0.8X_{t-1} + 0.6\epsilon_{t-1} + \epsilon_t$ et sa fonction d'autocorrélation.

4.1.4 Fonction d'autocorrélation partielle

Un processus $ARMA(p, q)$ peut être exprimé tel que : $X_t = \Theta(B) \Phi^{-1}(B) \epsilon_t$ où $\Phi^{-1}(B)$ est un polynôme de B .

Ainsi, la fonction d'autocorrélation partielle d'un processus mixte est infinie en étendue. Elle se comporte finalement comme la fonction d'autocorrélation partielle d'un processus moyenne mobile pure, selon l'ordre de la moyenne mobile et les valeurs des paramètres qu'il contient.

Remarque 51 *Le corrélogramme partiel d'un processus $ARMA(p, q)$ est le même que celui d'un processus MA à partir de l'ordre $p + 1$.*

- **Script R**

```
par(mfrow=c(1,2))
library(stats)
ts.arma<-arima.sim(list(order=c(1,0,1),ar=0.8,ma=0.6),n=100)
ts.plot(ts.arma,type="l",col="4")
pacf(ts.arma,col="4",10)
```

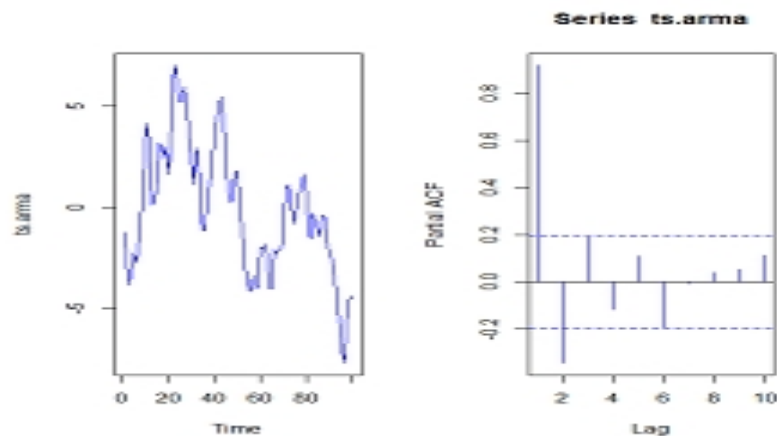


Figure 7. Trajectoire du processus $ARMA(1, 1)$: $X_t = 0.8X_{t-1} + 0.6\epsilon_{t-1} + \epsilon_t$ et sa fonction d'autocorrélation partielle.

4.1.5 Fonctions caractéristiques d'un $ARMA(1,1)$

Considérons un processus $ARMA(1, 1)$ qui a la représentation suivante

$$X_t - \phi X_{t-1} = \epsilon_t - \theta \epsilon_{t-1}.$$

On a donc $|\phi| < 1$ et $|\theta| < 1$. Déterminons maintenant le développement en moyenne mobile infinie. On a

$$\begin{aligned} X_t &= \sum_{j=0}^{+\infty} \phi^j (\epsilon_{t-j} - \theta \epsilon_{t-j-1}) \\ &= \epsilon_t + \sum_{j=1}^{+\infty} (\phi^j - \theta \phi^{j-1}) \epsilon_{t-j} \\ &= \epsilon_t + (\phi - \theta) \sum_{j=1}^{+\infty} \phi^{j-1} \epsilon_{t-j}. \end{aligned}$$

On peut directement calculer la fonction d'autocovariance ou d'autocorrélation à partir de cette expression. La variance $\gamma(0)$ vaut

$$\begin{aligned} \gamma(0) &= \sigma_\epsilon^2 \left(1 + (\phi - \theta)^2 \sum_{j=1}^{+\infty} \phi^{2(j-1)} \right) \\ &= \sigma_\epsilon^2 \left(1 + \frac{(\phi - \theta)^2}{1 - \phi^2} \right). \end{aligned}$$

De plus si $h > 0$, nous avons

$$\text{cov}(X_t, X_{t+h}) = \sigma_\epsilon^2 \left((\phi - \theta)^2 \phi^{h-1} + (\phi - \theta)^2 \phi^h \sum_{j=1}^{+\infty} \phi^{2(j-1)} \right)$$

On voit alors que

$$\gamma(h) = \sigma_\epsilon^2 \left((\phi - \theta) \phi^{h-1} + \frac{(\phi - \theta)^2 \phi^h}{1 - \phi^2} \right).$$

• Script R

```
ARMA<-arima.sim(list(order=c(1,0,1),ma=0.8,ar=0.4),n=1000)
par(mfrow=c(1,3))
plot(ARMA,col="blue",main="")
abline(h=0,col="red")
acf(ARMA,col="blue",main="Autocorrélation")
```

```
pacf(ARMA,col="blue",main="Autocorrélation partiel")
```

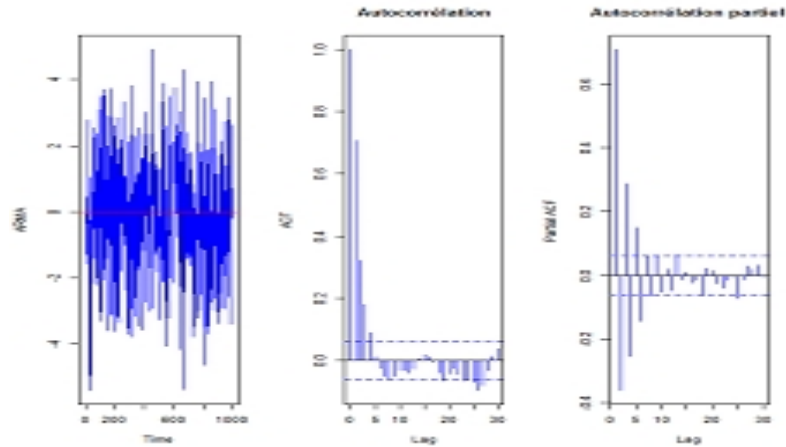


Figure 8. Trajectoire, corrélogramme et corrélogramme partiel du processus $ARMA(1,1)$

- **Script R**

```
ARMA<-arima.sim(list(order=c(3,0,3),ma=c(0.8,-0.4,-0.9),ar=c(-0.9,-0.8,0.4))
,n=1000)
par(mfrow=c(1,3))
plot(ARMA,col="blue",main="")
abline(h=0,col="red")
acf(ARMA,col="blue",main="Autocorrélation")
pacf(ARMA,col="blue",main="Autocorrélation partiel")
```

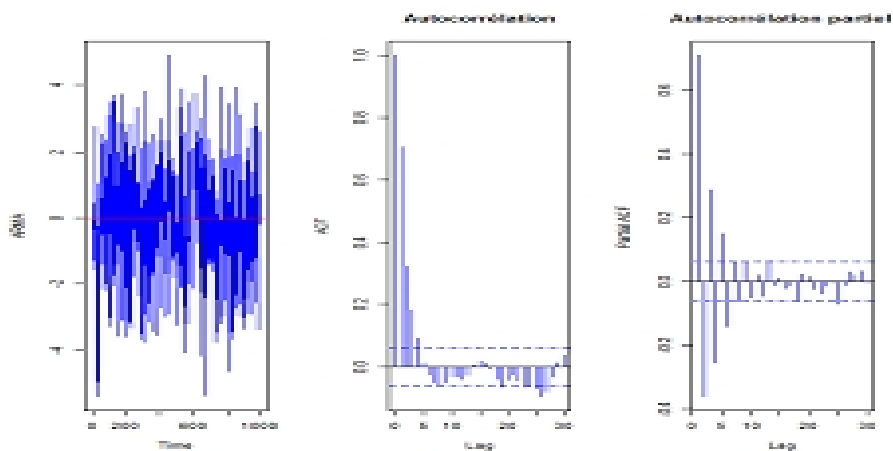


Figure 9. Trajectoire, corrélogramme et corrélogramme partiel du processus $ARMA(3,3)$

$$X_t - 0,9X_{t-1} - 0,8X_{t-2} + 0,4X_{t-3} = \epsilon_t + 0,8\epsilon_{t-1} - 0,4\epsilon_{t-2} - 0,9\epsilon_{t-3}.$$

4.2 Stationnarité et causalité

Soit $(X_t, t \in \mathbb{Z})$ un processus $ARMA(p, q)$. Se demander si ce processus est stationnaire est se questionner si $\{X_t\}$ admet une représentation du genre :

$$X_t = \sum_{j \geq 0} \psi_j \epsilon_{t-j} \text{ où } \forall t \sum_{j \geq 0} |\psi_j| < \infty$$

- On rappelle que : $\Phi(B) X_t = \Theta(B) \epsilon_t, \forall t$.
- On aimerait faire « disparaître » l'opérateur $\Phi(B)$.
- On aimerait multiplier par $\Phi^{-1}(B)$ de chaque côté.

4.2.1 Stationnarité d'un processus ARMA(p,q)

Un processus $ARMA(p, q)$ est stationnaire si le polynôme $\Phi(B)$ admet des racines en dehors du cercle unité en module.

• Si $\Phi^{-1}(B)$ est un opérateur qui existe, on a alors que l'équation $\Phi(B) X_t = \Theta(B) \epsilon_t$ peut être multipliée de chaque côté par l'opérateur $\Phi^{-1}(B)$, ce qui nous donne :

$$\begin{aligned} \Phi^{-1}(B) \Phi(B) X_t &= \Phi^{-1}(B) \Theta(B) \epsilon_t \\ X_t &= \Phi^{-1}(B) \Theta(B) \epsilon_t \\ X_t &= \Psi(B) \epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}. \end{aligned}$$

théorème 52 Si $\Phi(z) \neq 0$ pour tout $z \in \mathbb{C}$ tel que $|z| = 1$, alors l'équation

$$\Phi(B) X_t = \Theta(B) \epsilon_t,$$

d'un modèle ARMA a une unique solution stationnaire définie par :

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}.$$

Preuve. Voir [2]. ■

Remarque 53 Si $(X_t, t \in \mathbb{Z})$ est un processus ARMA pour lequel les polynômes $\Phi(\cdot)$ et $\Theta(\cdot)$ ont des zéros communs, il existe deux possibilités :

• Aucun des zéros communs ne se trouve sur le cercle unité, $(X_t, t \in \mathbb{Z})$ est la unique solution stationnaire d'équation ARMA sans zéro commun, obtenue en annulant les facteurs communs de $\Phi(\cdot)$ et $\Theta(\cdot)$.

• Au moins l'un des zéros communs repose sur le cercle unité, dans ce cas l'équation ARMA peut avoir plus d'une solution stationnaire.

4.2.2 Causalité

Définition 54 Un processus ARMA(p, q) défini par l'équation $\Phi(B)X_t = \Theta(B)\epsilon_t$ est dit causal s'il existe une suite de constantes (ψ_i) telle que

$$\sum_{i=0}^{\infty} |\psi_i| < \infty$$

et X_t peut s'exprimer comme suit :

$$X_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \text{ pour } t = 0, \pm 1, \dots$$

La condition de la causalité est équivalente à :

$$\Phi(x) = 1 - \phi_1 x - \phi_2 x^2 - \dots - \phi_p x^p = 0 \implies |x| > 1.$$

La détermination des coefficients ψ_j se fait par l'identification

$$\Phi(B)X_t = \Theta(B)\epsilon_t \implies X_t = \frac{\Theta(B)}{\Phi(B)}\epsilon_t = \Psi(B)\epsilon_t$$

tel que :

$$\begin{aligned} \Psi(B)\Phi(B) &= \Theta(B) \\ \implies (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) (\psi_0 + \psi_1 B + \psi_2 B^2 + \dots + \psi_p B^p) \\ &= (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_p B^p) \end{aligned}$$

En développant cette relation et par identification, on peut vérifier que les coefficients ψ_j se calculent comme suit :

$$\psi_0 = 1, \psi_1 - \phi_1 \psi_0 = \theta_1, \psi_2 - \phi_1 \psi_1 - \phi_2 \psi_0 = \theta_2$$

et d'une manière générale on a :

$$\psi_j - \sum_{k=1}^p \phi_k \psi_{j-k} = \theta_j \text{ pour } j = 0, 1, 2, \dots$$

où $\theta_0 = 1, \theta_j = 0$ pour $j > q$ et $\psi_j = 0$ pour $j < 0$.

Remarque 55 La définition de la causalité d'un modèle ARMA n'est pas une propriété du processus $(X_t, t \in \mathbb{Z})$ seul, mais elle exprime le rapport entre les deux processus $(X_t, t \in \mathbb{Z})$ et $(\epsilon_t, t \in \mathbb{Z})$ apparaissant dans l'équation du notre modèle. Il a été établi par Yule Walker qu'un processus $(X_t, t \in \mathbb{Z})$ est causal s'il est le résultat à travers un filtre linéaire causal d'un processus bruit blanc $(\epsilon_t, t \in \mathbb{Z})$.

4.2.3 Inversibilité

Définition 56 Un processus ARMA(p, q) défini par l'équation $\Phi(B)X_t = \Theta(B)\epsilon_t$ est dit inversible s'il existe une suite de constantes (π_j) telle que

$$\sum_{j=0}^{\infty} |\pi_j| < \infty$$

et ϵ_t peut s'exprimer tel que :

$$\epsilon_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad t = 0, \pm 1, \dots$$

théorème 57 Un processus ARMA(p, q) est causal (resp. inversible), c'est-à-dire, qu'il peut être représenté sous la forme

$$X_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$$

où

$\sum_{i=0}^{\infty} \psi_i^2 < \infty$ (resp. $\epsilon_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$ où $\sum_{j=0}^{\infty} \pi_j^2 < \infty$), si et seulement si toutes les racines de son polynôme caractéristique $\Phi(B)$ (resp. $\Theta(B)$) sont à l'extérieur du cercle unité. Les coefficients ψ_i (resp. π_j) sont dans ce cas les coefficients de la série de Taylor et nous avons

$$\Psi(B) = \frac{\Theta(B)}{\Phi(B)} \left(\text{resp. } \pi(B) = \frac{\Phi(B)}{\Theta(B)} \right)$$

théorème 58

a) Pour un processus ARMA(p, q) où toutes les racines du polynôme caractéristique $\Phi(B)$ à l'extérieur du cercle unité, les coefficients $\psi_i = \sigma_\epsilon^{-2} E\{X_t \epsilon_{t-i}\}$ de la représentation causale

$$X_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$$

satisfont à l'équation de récurrence suivante :

$$\psi_k = \begin{cases} 1 & \text{si } k = 0 \\ \theta_k + \sum_{i=k}^{\min(k,q)} \phi(i) \psi(k-i) & \text{si } 1 \leq k \leq q \\ \sum_{i=k}^{\min(k,q)} \phi(i) \psi(k-i) & \text{si } k > q \end{cases}$$

b) Pour un processus ARMA(p, q) où toutes les racines du polynôme caractéristique $\Theta(B)$ à l'extérieur du cercle unité, les coefficients $\pi_j = \sigma_\epsilon^{-2} E\{X_t \epsilon_{t-j}\}$ de la représentation inverse

$$\epsilon_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$$

satisfont à l'équation de récurrence suivante :

$$\pi_k = \begin{cases} 1 & \text{si } k = 0 \\ -\phi_k + \sum_{j=k}^{\min(k,q)} \theta(j) \pi(k-j) & \text{si } 1 \leq k \leq p \\ \sum_{j=k}^{\min(k,q)} \theta(j) \pi(k-j) & \text{si } k > p \end{cases}$$

• L'opérateur $\Theta^{-1}(B)$ existe si et seulement si les racines de l'équation $\Theta(x) = 0$ sont supérieures à 1 en module.

• Dans un tel cas :

$$\begin{aligned} \Theta^{-1}(B) \Phi(B) X_t &= \Theta^{-1}(B) \Theta(B) \epsilon_t \\ \Theta^{-1}(B) \Phi(B) X_t &= \epsilon_t \\ \Pi(B) X_t &= \epsilon_t \end{aligned}$$

où

$$\Pi(B) = \Theta^{-1}(B) \Phi(B) = I - \sum_{j \geq 1} \pi_j B^j.$$

De plus, nous avons donc

$$\begin{aligned} \left(I - \sum_{j \geq 1} \pi_j B^j \right) X_t &= \epsilon_t, \\ X_t &= \sum_{j \geq 1} \pi_j B^j X_t + \epsilon_t, \\ X_t &= \sum_{j \geq 1} \pi_j X_{t-j} + \epsilon_t. \end{aligned}$$

Remarque 59 Comme nous allons le constater, la stationnarité et l'inversibilité reposeront sur l'étude des racines du polynôme autorégressif (stationnarité) et du polynôme moyenne mobile (inversibilité).

Chapitre 5

Estimation des paramètres des processus MA , AR et $ARMA$

5.1 Estimation des paramètres d'un processus MA

Etant donné le processus $(X_t, t \in \mathbb{Z})$ admettant une représentation $MA(q)$

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

Le but de ce chapitre, est d'estimer les paramètres d'un modèle moyenne mobile. Premièrement, nous introduisons la méthode du maximum de vraisemblance (Maximum Likelihood) notée MLE , par la suite on introduit la méthode des moindres carrés ordinaires notée OLS .

5.1.1 Estimation par la méthode du maximum de vraisemblance

Dans cette section, en étudiant la méthode du maximum de vraisemblance, qu'elle est couramment utiliser pour estimer les paramètres d'un processus $MA(q)$.

La fonction de vraisemblance associée à un échantillon $X = (x_1, \dots, x_n)$ i.i.d d'une loi dont la densité est $f(X, \theta, \sigma^2)$, avec $\theta = (\theta_1, \dots, \theta_q) \in \mathbb{R}^q$, est définie par :

$$L(X, \theta, \sigma^2) = \prod_{i=1}^n f(X, \theta, \sigma^2).$$

Pour illustrer l'utilisation de la méthode du maximum de vraisemblance dans l'étude des séries chronologiques, il est nécessaire de supposer la connais-

sance de la loi des erreurs : nous supposons les erreurs normalement distribuées. Les erreurs étant normalement distribuées et indépendantes, le processus ε_t est par hypothèse un bruit blanc gaussien, de plus, le processus $MA(q)$ est stationnaire de second ordre ssi : $|\theta_i| < 1$. Donc la fonction de vraisemblance d'un processus $MA(q)$ est donnée par :

$$L(X, \theta, \sigma^2) = (2\pi)^{\frac{-n}{2}} \left| \sigma^2 \Omega \right|^{\frac{-1}{2}} \exp \left(\frac{-1}{2\sigma^2} X' \Omega^{-1} X \right)$$

où $X = (x_1, \dots, x_n)$ est un échantillon de taille n et $\sigma^2 \Omega$ représente la matrice de variance-covariance de n variables consécutives à partir d'un processus $MA(q)$ définie par

$$\Omega = \begin{pmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_q & 0 & \dots & 0 \\ \gamma_1 & \gamma_0 & \gamma_1 & \dots & \gamma_q & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \gamma_q & \dots & \gamma_1 & \gamma_0 \end{pmatrix},$$

le (i, j) ième élément de Ω est donné par $\gamma_{|i-j|}$, où γ_h est l'autocovariance d'ordre h d'un processus $MA(q)$. On introduit le logarithme :

$$\log L(X, \theta, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log(\sigma^2) - \frac{1}{2} \log |\Omega| - \frac{1}{2\sigma^2} X' \Omega^{-1} X$$

L'estimateur du maximum de vraisemblance (*MLE*) est la valeur $\hat{\theta}$ qui maximise la vraisemblance $L(X, \theta, \sigma^2)$. Parfois, il est possible de déduire cet estimateur en dérivant $\log L(X, \theta, \sigma^2)$ par rapport à chacun des paramètres $(\theta_1, \dots, \theta_q, \sigma^2)$ et par la suite en résolvant le système à q équations suivant,

$$\frac{\partial \log L(X, \theta, \sigma^2)}{\partial \theta_j} = 0, \quad j = 1, \dots, q.$$

5.1.2 Comportement asymptotique

L'estimateur du maximum de vraisemblance pour un processus moyenne mobile partagent les propriétés asymptotiques suivantes :

Consistance

Si $\hat{\theta}$ est un estimateur consistant de θ , c'est-à-dire, la loi empirique de $\hat{\theta}$ soit de plus en plus concentrée autour de la vraie valeur du paramètre θ .

Proposition 60 On dit qu'un estimateur de maximum de vraisemblance noté $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_q)$ de $\theta = (\theta_1, \dots, \theta_q)$, obtenu à partir de n observations x_1, \dots, x_n du processus MA(q), est fortement consistant si et seulement si $\hat{\theta}$ converge presque sûrement vers θ , c'est-à-dire,

$$\hat{\theta} \xrightarrow{p.s.} \theta \text{ lorsque } n \rightarrow \infty.$$

Maintenant, nous sommes en position de lancer un résultat sur la normalité asymptotique de notre estimateur.

Normalité asymptotique

Proposition 61 Si le processus MA(q) est stationnaire et causal et si $\hat{\theta}$ est l'estimateur de maximum de vraisemblance de θ , obtenu à partir de n observations x_1, \dots, x_n du processus moyenne mobile, alors

$$\sqrt{n}(\hat{\theta} - \theta) \rightarrow N(0, \sigma^2 \Omega^{-1}) \text{ lorsque } n \rightarrow \infty.$$

Exemple 62 Dans cet exemple, on traite l'estimation d'un processus MA (1) gaussien. Le logarithme de la fonction de vraisemblance du processus MA (1) suivant

$$X_t = \varepsilon_t + \theta \varepsilon_{t-1}$$

où $\varepsilon_t \rightarrow iid N(0, \sigma^2)$ et $|\theta| < 1$, est donnée par :

$$L(\theta, \sigma^2) = -0.5n \ln(2\pi) - 0.5 \ln(\sigma^2) - 0.5 \ln(|\Omega|) - \frac{1}{2\sigma^2} X' \Omega^{-1} X$$

où Ω est la matrice des covariances de x_1, \dots, x_n donnée par :

$$\Omega = \begin{pmatrix} 1 + \theta^2 & \theta & 0 & \dots & 0 \\ \theta & 1 + \theta^2 & \theta & \dots & \dots \\ 0 & \theta & 1 + \theta^2 & \dots & \dots \\ \dots & \dots & \dots & \dots & \theta \\ 0 & 0 & \dots & \theta & 1 + \theta^2 \end{pmatrix}$$

5.1.3 Estimation par la méthode des moindres carrés

Le principe de la méthode des moindres carrés repose sur la recherche de la valeur de θ qui minimise la somme des carrés des erreurs commises lors de l'estimation.

Soit $S(\theta)$ cette somme, définie par

$$S(\theta) = \sum_{t=-\infty}^n E(\varepsilon_t, \theta, X)^2 = \sum_{t=-\infty}^n \widehat{\varepsilon}_t^2.$$

Cette méthode est basée sur une relation de récurrence nécessitant des valeurs initiales de $\widehat{\varepsilon}_t^2$. Le problème tout d'abord est de minimiser :

$$S(\theta) = \sum_{t=1}^n \varepsilon_t^2.$$

On cherche donc à minimiser la quantité suivante :

$$S(\theta_1, \dots, \theta_q) = \sum_{i=1}^n \frac{(x_i - \widehat{x}_i)^2}{r_{i-1}},$$

où $\widehat{x}_i = \sum_{j=1}^i \theta_j (i) (x_{i-j} - \widehat{x}_{i-j})$, $r_i = \frac{\sigma_i}{\sigma^2}$ et $\text{var}(\varepsilon_i) = \text{var}(x_i - \widehat{x}_i) = \sigma_i$. L'estimateur des moindres carrés vérifie :

$$\begin{cases} \widehat{\mu} = (\widehat{\theta}_1, \dots, \widehat{\theta}_q) = \arg \min_{(\theta_1, \dots, \theta_q)} S(\theta_1, \dots, \theta_q) \\ \widehat{\sigma}^2 = \frac{1}{n-q} S(\widehat{\theta}_1, \dots, \widehat{\theta}_q). \end{cases}$$

Les estimateurs obtenus sont efficaces.

5.1.4 Comportement asymptotique

Les estimations exactes des moindres carrés pour un processus moyenne mobile partagent les propriétés asymptotiques suivantes :

La consistance

Si $\widehat{\mu}$ l'estimateur du moindres carrés d'un processus $MA(q)$ est un estimateur consistant, c'est-à-dire, la loi empirique de $\widehat{\theta}$ soit de plus en plus concentrée autour de la vraie valeur du paramètre θ .

Proposition 63 On dit qu'un estimateur des moindres carrés noté $\widehat{\mu} = (\widehat{\theta}_1, \dots, \widehat{\theta}_q)$ de $\theta = (\theta_1, \dots, \theta_q)$, obtenu à partir de n observations x_1, \dots, x_n du processus $MA(q)$, est fortement consistant si et seulement si $\widehat{\mu}$ converge presque sûrement vers θ , c'est-à-dire,

$$\widehat{\mu} \xrightarrow{p.s.} \theta \text{ lorsque } n \rightarrow \infty.$$

Maintenant, on lance un résultat sur la normalité asymptotique.

La normalité asymptotique

Proposition 64 Si le processus $MA(q)$ est stationnaire et causal et si $\hat{\mu}$ est l'estimateur des moindres carrés de θ , obtenu à partir de n observations x_1, \dots, x_n du processus moyenne mobile alors

$$\sqrt{n}(\hat{\mu} - \theta) \xrightarrow{d} N(0, \sigma^2 V(\theta)_q) \text{ lorsque } n \rightarrow \infty$$

où $V(\theta)_q = \sigma^2 [E\mathbf{V}_t \mathbf{V}_t']^{-1}$, $\mathbf{V}_t = (V_t, \dots, V_{t+1-q})'$ et $(V_t, t \in \mathbb{Z})$ est le processus autorégressif,

$$\Theta(B) V_t = \epsilon_t.$$

5.1.5 Normalité asymptotique des différentes méthodes

Soit $(X_t, t \in \mathbb{Z})$ un processus $MA(2)$ défini par :

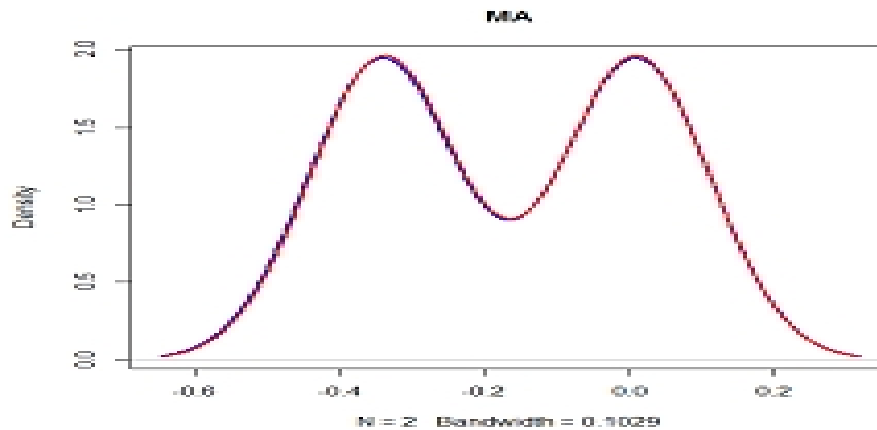
$$X_t = \epsilon_t + 0.7\epsilon_{t-1} - 0.9\epsilon_{t-2}$$

Dans le trajectoire au-dessous, on donne une comparaison entre les deux différentes méthodes d'estimation qu'elles sont représentées dans le mémoire. La distribution de l'estimateur par la méthode des moindres carrés (*OLS*) est en bleu, et la distribution de l'estimateur par la méthode du maximum de vraisemblance (*MLE*) est en rouge.

- **Script R**

```
CFOLS=CFMLE=matrix(NA,5000,2)
for(s in 1 :5000){
X=arima.sim(n=240,list(ma=c(0.7,-0.9)))
CFOLS[s,]=ar(X,order.max=2,method="ols")$ar[1 :2]
CFMLE[s,]=ar(X,order.max=2,method="mle")$ar
}
plot(density(CFOLS[s,]),lwd=2,col="4",main="MA")
```

```
lines(density(CFMLE[s,]),lwd=2,col="2")
```



5.2 Simulation avec le modèle MA(2)

Dans cette section, nous donnons une simulation qui présente les résultats de la section précédente. On exécute cette simulation avec le modèle MA (2)

$$X_t = \begin{pmatrix} \theta_1 & \theta_2 \end{pmatrix} \begin{pmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-2} \end{pmatrix} + \varepsilon_t$$

où ε_t est un bruit blanc, à chaque fois, on change la valeur de $(\theta_i)_{i=1,2}$ et de N avec $N \in (2000, 3000, 4000)$, pour faire cette simulation on utilise le logiciel R avec un script qu'il estime les paramètres par les deux méthodes :

- La méthode du maximum de vraisemblance (*MLE*) .
- La méthode des moindres carrés(*OLS*)

• Script

```
> Ma2=arima.sim(n=2000,list(order=c(0,0,2),ma=c(0.4,-0.4)))
> fit <- arima(order=c(0,0,2),Ma2,method="ML",include.mean=FALSE)
> fit
```

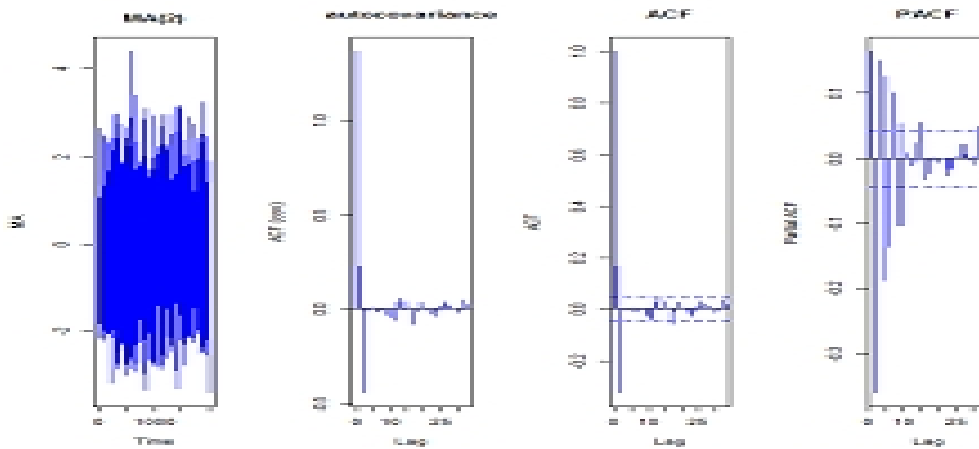
Trajectoire :

• Script R

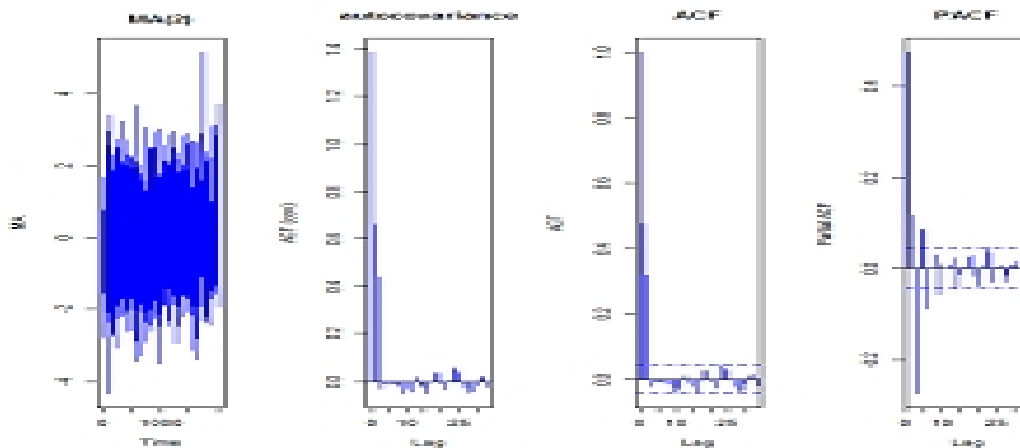
```
MA<-arima.sim(list(order=c(0,0,2),ma=c(0.4,-0.4)),n=2000)
par(mfrow=c(1,4))
plot(MA,col="blue",main="MA(2)")
acf( MA,col="blue",type="covariance",main="autocovariance")
acf( MA,col="blue",main="ACF")
```

pacf(MA,col="blue",main="PACF")

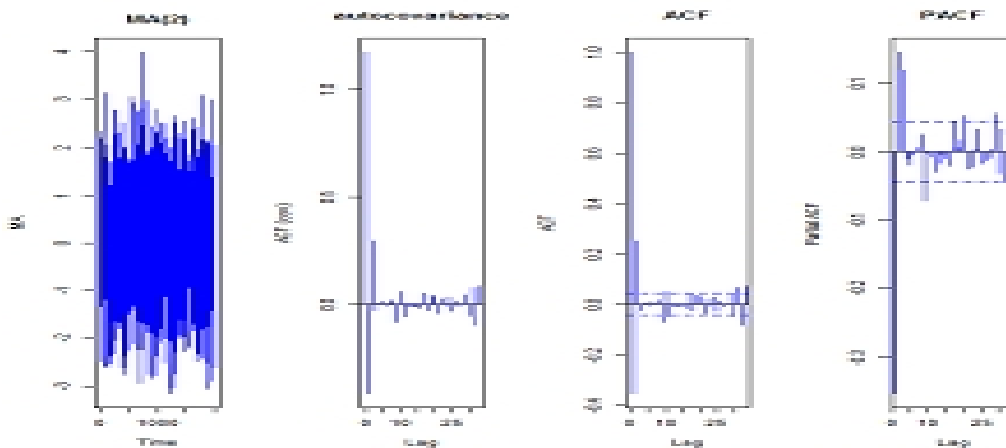
Vraie valeur de θ et σ^2	N	MLE	OLS
$\begin{pmatrix} 0.4 \\ 0 \quad -0.4 \\ \sigma^2 \end{pmatrix}$	2000	$\begin{pmatrix} 0.4139 \\ -0.3990 \\ 1.06 \end{pmatrix}$	$\begin{pmatrix} 0.3974 \\ -0.3848 \\ 1.063 \end{pmatrix}$
	3000	$\begin{pmatrix} 0.4053 \\ -0.4060 \\ 0.9647 \end{pmatrix}$	$\begin{pmatrix} 0.4164 \\ -0.3771 \\ 0.9946 \end{pmatrix}$
	4000	$\begin{pmatrix} 0.3986 \\ -0.3980 \\ 1.007 \end{pmatrix}$	$\begin{pmatrix} 0.4025 \\ -0.3961 \\ 1.02 \end{pmatrix}$



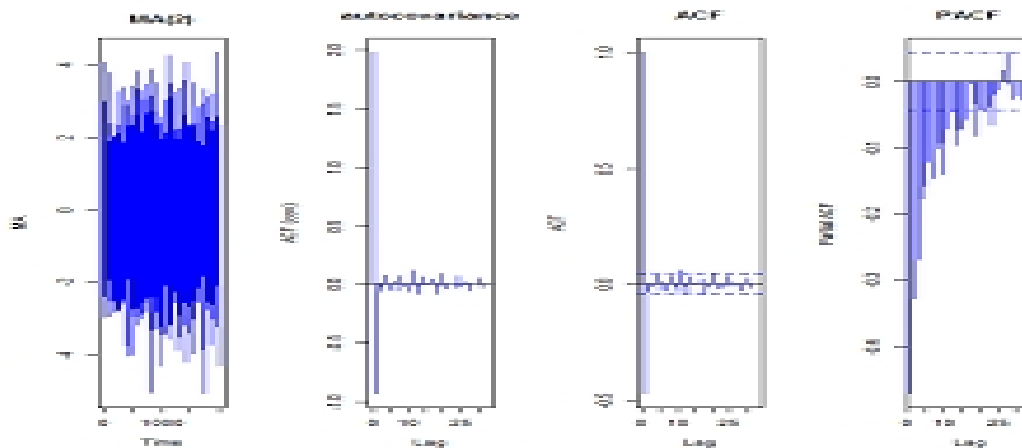
Vraie valeur de θ et σ^2	N	MLE	OLS
$\begin{pmatrix} 0.5 \\ 0.5 \\ \sigma^2 \end{pmatrix}$	2000	$\begin{pmatrix} 0.4904 \\ 0.5252 \\ 0.9921 \end{pmatrix}$	$\begin{pmatrix} 0.4828 \\ 0.5148 \\ 0.9729 \end{pmatrix}$
	3000	$\begin{pmatrix} 0.4971 \\ 0.5042 \\ 1.017 \end{pmatrix}$	$\begin{pmatrix} 0.5008 \\ 0.4887 \\ 1.025 \end{pmatrix}$
	4000	$\begin{pmatrix} 0.4745 \\ 0.4884 \\ 0.9856 \end{pmatrix}$	$\begin{pmatrix} 0.5032 \\ 0.4954 \\ 0.9785 \end{pmatrix}$



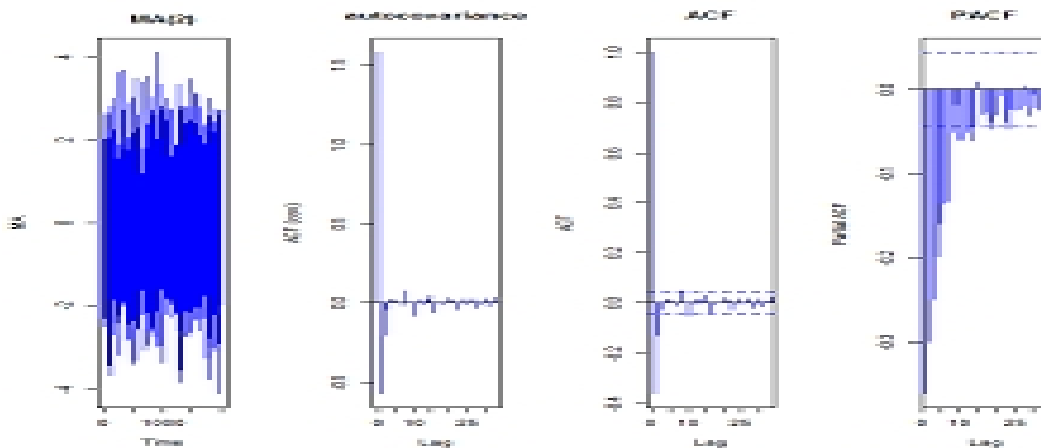
Vraie valeur de θ et σ^2	N	MLE	OLS
$\begin{pmatrix} -0.3 \\ 0.3 \\ \sigma^2 \end{pmatrix}$	2000	$\begin{pmatrix} -0.3030 \\ 0.3161 \\ 1.026 \end{pmatrix}$	$\begin{pmatrix} -0.2791 \\ 0.3089 \\ 1.05 \end{pmatrix}$
	3000	$\begin{pmatrix} -0.3301 \\ 0.2757 \\ 0.9864 \end{pmatrix}$	$\begin{pmatrix} -0.2949 \\ 0.3110 \\ 1.037 \end{pmatrix}$
	4000	$\begin{pmatrix} -0.2852 \\ 0.3008 \\ 0.9839 \end{pmatrix}$	$\begin{pmatrix} -0.2781 \\ 0.2913 \\ 1.015 \end{pmatrix}$



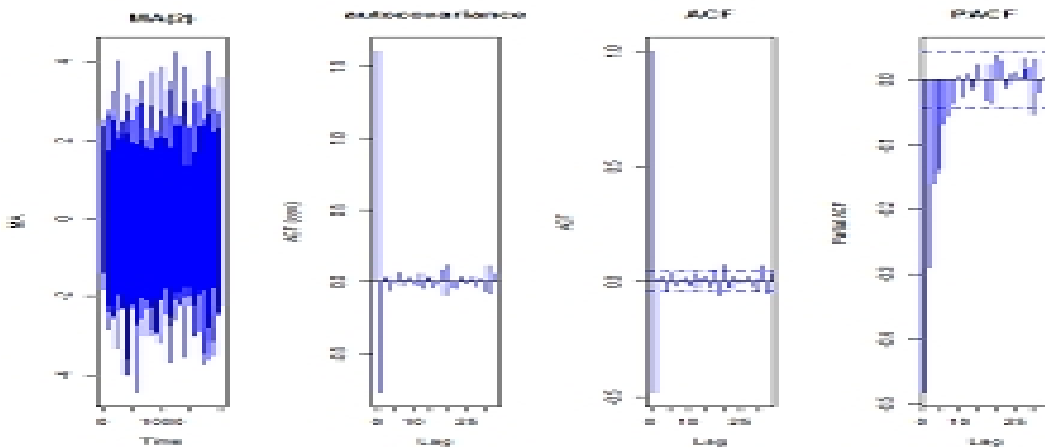
Vraie valeur de θ et σ^2	N	MLE	OLS
$\begin{pmatrix} -0.9 \\ -0.05 \\ \sigma^2 \end{pmatrix}$	2000	$\begin{pmatrix} -0.8675 \\ -0.0742 \\ 1.008 \end{pmatrix}$	$\begin{pmatrix} -0.8972 \\ -0.0453 \\ 1.018 \end{pmatrix}$
	3000	$\begin{pmatrix} -0.9288 \\ -0.0136 \\ 1.034 \end{pmatrix}$	$\begin{pmatrix} -0.9082 \\ -0.0371 \\ 1.016 \end{pmatrix}$
	4000	$\begin{pmatrix} -0.8917 \\ -0.0502 \\ 0.9905 \end{pmatrix}$	$\begin{pmatrix} -0.8784 \\ -0.0794 \\ 1.022 \end{pmatrix}$



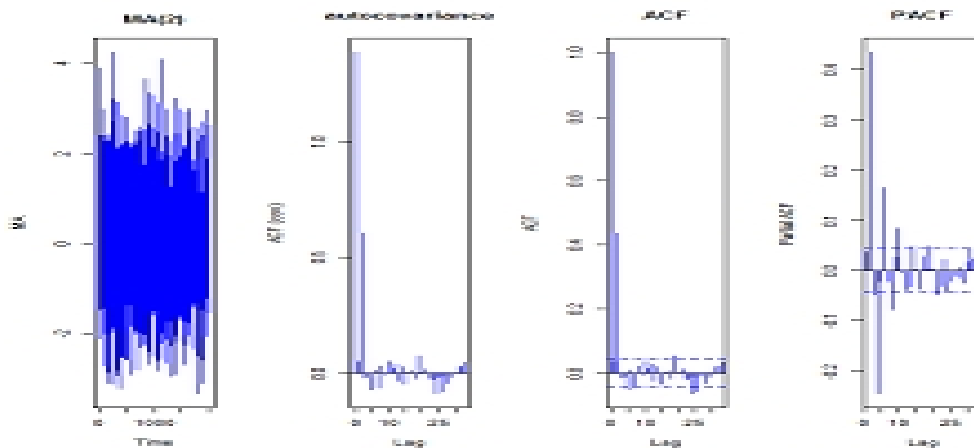
Vraie valeur de θ et σ^2	N	MLE	OLS
$\begin{pmatrix} -0.7 \\ -0.2 \\ \sigma^2 \end{pmatrix}$	2000	$\begin{pmatrix} -0.692 \\ -0.2083 \\ 1.003 \end{pmatrix}$	$\begin{pmatrix} -0.6975 \\ -0.0089 \\ 1.017 \end{pmatrix}$
	3000	$\begin{pmatrix} -0.6821 \\ -0.2134 \\ 1.029 \end{pmatrix}$	$\begin{pmatrix} -0.7117 \\ -0.0084 \\ 1.016 \end{pmatrix}$
	4000	$\begin{pmatrix} -0.6792 \\ -0.2156 \\ 1.01 \end{pmatrix}$	$\begin{pmatrix} -0.7039 \\ -0.0129 \\ 1.019 \end{pmatrix}$



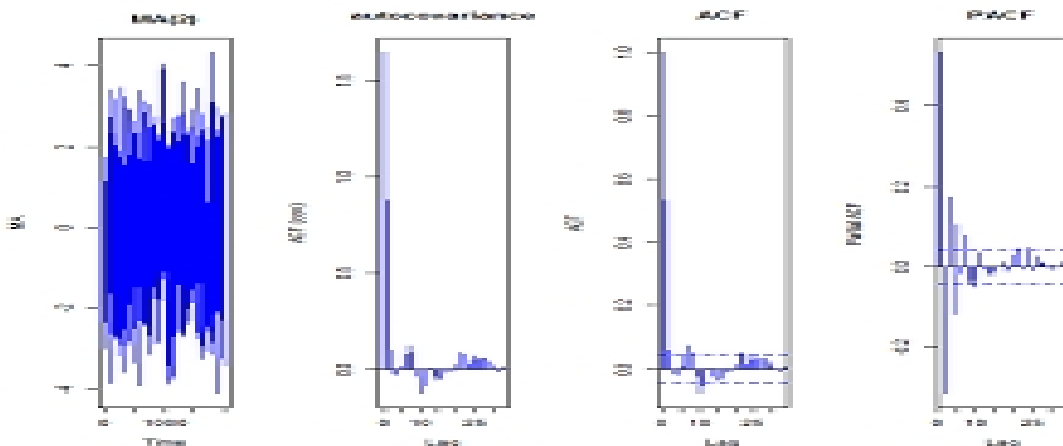
Vraie valeur de θ et σ^2	N	MLE	OLS
$\begin{pmatrix} -0.7 \\ 0 \\ \sigma^2 \end{pmatrix}$	2000	$\begin{pmatrix} -0.6711 \\ -0.0316 \\ 1.063 \end{pmatrix}$	$\begin{pmatrix} 0.6958 \\ -0.0062 \\ 0.9921 \end{pmatrix}$
	3000	$\begin{pmatrix} -0.7005 \\ 0.0217 \\ 1.019 \end{pmatrix}$	$\begin{pmatrix} 0.7027 \\ -0.0108 \\ 0.9911 \end{pmatrix}$
	4000	$\begin{pmatrix} -0.6871 \\ -0.0276 \\ 1.022 \end{pmatrix}$	$\begin{pmatrix} 0.7203 \\ 0.0133 \\ 1.006 \end{pmatrix}$



Vraie valeur de θ et σ^2	N	MLE	OLS
$\begin{pmatrix} 0 \\ 0.6 \\ \sigma^2 \end{pmatrix}$	2000	$\begin{pmatrix} 0.0150 \\ 0.5875 \\ 1.011 \end{pmatrix}$	$\begin{pmatrix} 0.0148 \\ 0.6207 \\ 1.012 \end{pmatrix}$
	3000	$\begin{pmatrix} -0.0121 \\ 0.6029 \\ 1.026 \end{pmatrix}$	$\begin{pmatrix} -0.0004 \\ 0.6038 \\ 0.9955 \end{pmatrix}$
	4000	$\begin{pmatrix} 0.0160 \\ 0.6138 \\ 1.001 \end{pmatrix}$	$\begin{pmatrix} -0.0188 \\ 0.6135 \\ 1.014 \end{pmatrix}$



Vraie valeur de θ et σ^2	N	MLE	OLS
$\begin{pmatrix} 0.8 \\ 0.12 \\ \sigma^2 \end{pmatrix}$	2000	$\begin{pmatrix} 0.8057 \\ 0.1277 \\ 1.012 \end{pmatrix}$	$\begin{pmatrix} 0.8153 \\ 0.1303 \\ 1.042 \end{pmatrix}$
	3000	$\begin{pmatrix} 0.7949 \\ 0.1262 \\ 1.003 \end{pmatrix}$	$\begin{pmatrix} 0.8291 \\ 0.1670 \\ 1.017 \end{pmatrix}$
	4000	$\begin{pmatrix} 0.8239 \\ 0.1275 \\ 1.006 \end{pmatrix}$	$\begin{pmatrix} 0.8150 \\ 0.1443 \\ 0.9927 \end{pmatrix}$



5.2.1 Résultats et commentaires

D'après les résultats empiriques qui sont reportés dans les tables au dessus, on observe que l'augmentation de la taille d'échantillon joue un rôle essentiel pour l'obtention des meilleurs résultats. Cela est prouvé quand l'estimateur $\hat{\theta}$ calculé est plus proche de la vraie valeur, plus la taille d'échantillon est plus élevée.

5.3 Estimation des paramètres d'un processus autorégressifs

Pour prédire les valeurs futures d'un processus AR, il sera nécessaire d'estimer les paramètres. Nous allons à présent traiter les différentes méthodes d'estimation telles que la méthode de Yule-Walker, la méthode de maximum de vraisemblance et la méthode des moindres carrés. Alors, dans cette section, nous supposons que l'ordre du modèle est p . Nous savons donc que les données proviennent, disons, d'un processus $AR(p)$, et en se demande, comment estimer les p paramètre $\phi_1, \phi_2, \dots, \phi_p$ apparaissant dans (3.1) en utilisant l'une de ces méthodes d'estimation.

5.3.1 Méthode de Yule-Walker

Soit $(X_t, t \in \mathbb{Z})$ un processus autorégressif d'ordre p , stationnaire et causal, de moyenne nulle défini par

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t$$

où ε_t est un bruit blanc de variance σ_ε^2 .

L'idée de l'estimateur de Yule-Walker est de trouver des équations reliant les paramètres ϕ_i et la fonction d'autocovariance en utilisant l'hypothèse de processus causal. Commençons d'abord par multiplier l'équation (3.1) par X_{t-h} et prenons l'espérance des deux côtés de l'équation. Nous obtenons

$$\begin{aligned} \gamma(h) &= E(X_t X_{t-h}) = E\left(\sum_{i=1}^p \phi_i X_{t-i} X_{t-h} + \varepsilon_t X_{t-h}\right) \\ &= \sum_{i=1}^p \phi_i E(X_{t-i} X_{t-h}) + E(\varepsilon_t X_{t-h}) \\ &= \sum_{i=1}^p \phi_i \gamma(h-i) + E(\varepsilon_t X_{t-h}) \end{aligned} \quad (5.1)$$

Regardons plus attentivement le terme $E(\varepsilon_t X_{t-h})$ de (5.1). Puisque $(X_t, t \in \mathbb{Z})$ est causal, on peut utiliser la relation suivante

$$X_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

On a donc

$$\begin{aligned}
 E(\varepsilon_t X_{t-h}) &= E\left(\varepsilon_t \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-h-i}\right) \\
 &= E\left(\sum_{i=0}^{\infty} \psi_i \varepsilon_t \varepsilon_{t-h-i}\right) \\
 &= \sum_{i=0}^{\infty} \psi_i E(\varepsilon_t \varepsilon_{t-h-i})
 \end{aligned} \tag{5.2}$$

Pour $h = 0$, on a

$$E(\varepsilon_t \varepsilon_{t-i}) = \begin{cases} \sigma_\varepsilon^2 & \text{si } i = 0 \\ 0 & \text{sinon} \end{cases}$$

Pour $h > 0$, on aura toujours $E(\varepsilon_t \varepsilon_{t-h-i}) = 0$. On obtient donc le système d'équations suivant

$$\gamma(h) = \begin{cases} \sum_{i=0}^p \phi_i \gamma(i-h) & \text{si } 1 \leq h \leq p \\ \sum_{i=0}^p \phi_i \gamma(i-h) + \sigma_\varepsilon^2 & \text{si } h = 0 \end{cases}$$

On a un système d'équations à résoudre avec p équations. Soient les vecteurs $\underline{\gamma}_p = (\gamma(1), \gamma(2), \dots, \gamma(p))'$, $\underline{\phi}_p = (\phi_1, \phi_2, \dots, \phi_p)'$ et la matrice

$$\Omega_p = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma(0) \\ \gamma(p-1) & \cdots & \gamma(1) & \gamma(0) \end{pmatrix}, \tag{5.3}$$

on obtient donc le système

$$\begin{cases} \underline{\gamma}_p = \Omega_p \underline{\phi}_p \\ \gamma(0) = \underline{\phi}_p' \underline{\gamma}_p + \sigma_\varepsilon^2. \end{cases}$$

Ces équations sont les équations de Yule-Walker. Elles créent un lien entre les p paramètres du modèle et la fonction d'autocovariance. Afin d'estimer le vecteur de paramètres $\underline{\phi}_p$, nous devons utiliser la fonction d'autocovariance empirique $\hat{\underline{\gamma}}_p$ comme un estimateur de $\underline{\gamma}_p$. On aura donc un système d'équations qui permettra d'estimer le vecteur $\underline{\phi}_p$. Le système se réécrit de la façon suivante

$$\begin{cases} \hat{\underline{\gamma}}_p = \hat{\Omega}_p \hat{\underline{\phi}}_p \\ \hat{\gamma}(0) = \hat{\underline{\phi}}_p' \hat{\underline{\gamma}}_p + \hat{\sigma}_\varepsilon^2. \end{cases} \tag{5.4}$$

Pour obtenir un estimateur de $\underline{\phi}_p$, il faut donc résoudre (5.4) en inversant (il faut bien entendu s'assurer que $\widehat{\Omega}_p$ soit une matrice inversible) la matrice $\widehat{\Omega}_p$

$$\widehat{\underline{\phi}}_p = \left(\widehat{\Omega}_p\right)^{-1} \widehat{\underline{\gamma}}_p$$

puis en calculant

$$\widehat{\sigma}_\varepsilon^2 = \widehat{\gamma}(0) - \widehat{\underline{\phi}}_p' \widehat{\underline{\gamma}}_p.$$

Remarque 65 Nous signalons déjà ici qu'il existe une méthode pour résoudre ce système d'équations de Yule-Walker sans passer par l'inversion de la matrice $\widehat{\Omega}_p$, c'est l'algorithme itératif de Durbin-Levinson.

Le résultat suivant décrit la normalité asymptotique des estimateurs de Yule-Walker.

Proposition 66 Si le processus $(X_t, t \in \mathbb{Z})$ défini par (3.1) est stationnaire et causal et si $\widehat{\underline{\phi}}_p$ est l'estimateur de Yule-Walker de $\underline{\phi}_p$, obtenu à partir de N observations X_1, \dots, X_N du processus autorégressif, alors

$$\sqrt{N}(\widehat{\underline{\phi}}_p - \underline{\phi}_p) \rightarrow N_{asym}(0, \sigma_\varepsilon^2 \Omega_p^{-1}) \text{ lorsque } N \rightarrow \infty$$

où Ω_p est la matrice de covariance donnée par (5.3) (Cette matrice est inversible si $\gamma(0) > 0$). De plus

$$\widehat{\sigma}_\varepsilon^2 \xrightarrow{p} \sigma_\varepsilon^2 \text{ lorsque } N \rightarrow \infty$$

Preuve. Voir [2], proposition 8.10.1. ■

Remarque 67 Les estimateurs de Yule-Walker sont évidemment consistants dès que les $\widehat{\gamma}(h)$ sont des estimateurs consistants des covariances $\gamma(h)$.

Les estimateurs de Yule-Walker sont donc consistants et, par leur normalité asymptotique, ils sont asymptotiquement efficaces car ils possèdent asymptotiquement les mêmes variances que les estimateurs du maximum de vraisemblance traités ci-dessous. On dit, qu'un estimateur est efficace si sa variance coïncide avec la plus petite variance possible (étant donnée par la borne de Rao-Cramér). En plus, la normalité asymptotique permet de construire des intervalles de confiance pour chaque composante du vecteur $\widehat{\underline{\phi}}_p$.

5.3.2 Méthode du maximum de vraisemblance

Dans cette partie, nous allons appliquer, comme méthode alternative, la méthode générale du maximum de vraisemblance pour estimer les paramètres d'un modèle AR.

Considérons un modèle $AR(p)$ stationnaire et causal défini par (3.1). Notre objectif est d'estimer le vecteur $\underline{\phi}_p$.

On écrit la vraisemblance en fonction des paramètres ϕ_i et σ_ε^2 et on maximise la vraisemblance par rapport à ces paramètres. Pour déterminer la vraisemblance, il est nécessaire de supposer que le processus $(\varepsilon_t, t \in \mathbb{Z})$ est i.i.d (cette condition est vérifiée d'après l'hypothèse du bruit blanc), donc le vecteur $(\varepsilon_1, \dots, \varepsilon_N)'$ est un vecteur gaussien. Les composantes du vecteur $\underline{X}_N = (X_1, \dots, X_N)'$ étant obtenues par combinaisons linéaires des composantes du vecteur $(\varepsilon_1, \dots, \varepsilon_N)'$, d'où, le vecteur \underline{X}_N sera un vecteur gaussien. Cette hypothèse nous permettra d'écrire simplement la vraisemblance du modèle comme suit

$$\mathcal{L}(\underline{X}_N, \underline{\phi}_p, \sigma_\varepsilon^2) = (2\pi)^{-\frac{N}{2}} \left| \det \sigma_\varepsilon^2 \Omega_N \right|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma_\varepsilon^2} \underline{X}_N' \Omega_N^{-1} \underline{X}_N \right\}$$

où $\sigma_\varepsilon^2 \Omega_N$ est la matrice $N \times N$ des covariances du vecteur \underline{X}_N .

Cette méthode peut être utilisée pour n'importe quel modèle tandis que les équations de Yule-Walker sont plus reconnues pour les processus autorégressifs dû à la linéarité des équations. Si la série est stationnaire, causale et inversible, ces estimateurs auront la propriété d'être asymptotiquement normaux. La plupart des logiciels de statistique utilisent des algorithmes itératifs pour calculer ces estimateurs. Alors il est également possible de démontrer un théorème de la normalité asymptotique de ces estimateurs qui sont, comme nous l'avons déjà indiqué, asymptotiquement efficaces.

Proposition 68 Si le processus $(X_t, t \in \mathbb{Z})$ défini par (3.1) est stationnaire et causal et si $\hat{\underline{\phi}}_p$ est l'estimateur du maximum de vraisemblance de $\underline{\phi}_p$, obtenu à partir de N observations X_1, \dots, X_N du processus autorégressif, alors

$$\sqrt{N}(\hat{\underline{\phi}}_p - \underline{\phi}_p) \rightarrow N_{asym}(0, \sigma_\varepsilon^2 \Omega_N^{-1}) \text{ lorsque } N \rightarrow \infty$$

où $\sigma_\varepsilon^2 \Omega_N$ est la matrice $N \times N$ des covariances du vecteur \underline{X}_N .

Preuve. Voir [2]. ■

Remarque 69 La matrice de covariance $\sigma_\varepsilon^2 \Omega_N$ coïncide avec la matrice de covariance asymptotique des estimateurs de Yule-Walker $\sigma_\varepsilon^2 \Omega_p$.

5.3.3 Méthode des moindres carrés

Considérons un modèle $AR(p)$ stationnaire et causal défini par (3.1). Nous allons réécrire le modèle sous la forme suivante

$$X_t = \underline{Z}_t' \underline{\phi}_p + \varepsilon_t \quad (5.5)$$

où $\underline{Z}_t = (X_{t-1}, X_{t-2}, \dots, X_{t-p})'$.

L'estimation des paramètres du modèle (5.5) par la méthode des moindres carrés donne

$$\underline{\phi}_p = \left(\underline{Z}_t \underline{Z}_t' \right)^{-1} \underline{Z}_t X_t$$

et

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N-p} \sum_{t=1}^N \left(X_t - \underline{Z}_t' \hat{\underline{\phi}}_p \right)^2.$$

Il est possible de montrer le résultat suivant concernant la normalité asymptotique de l'estimateur des moindres carrés.

Proposition 70 *Si le processus $(X_t, t \in \mathbb{Z})$ défini par (3.1) est stationnaire et causal et si $\hat{\underline{\phi}}_p$ est l'estimateur des moindres carrés de $\underline{\phi}_p$, obtenu à partir de N observations X_1, \dots, X_N du processus autorégressif, alors*

$$\hat{\underline{\phi}}_p \xrightarrow{p} \underline{\phi}_p \text{ et } \hat{\sigma}_\varepsilon^2 \xrightarrow{p} \sigma_\varepsilon^2$$

et de plus,

$$\sqrt{N}(\hat{\underline{\phi}}_p - \underline{\phi}_p) \rightarrow N_{asym}(0, \sigma_\varepsilon^2 V) \text{ lorsque } N \rightarrow \infty$$

où $V = \lim_{N \rightarrow \infty} \left(\frac{1}{N} \underline{Z}_t \underline{Z}_t' \right)$ en probabilité.

5.3.4 Normalité asymptotique des différentes méthodes

Soit le processus stationnaire $AR(3)$ suivant

$$X_t = 0.8X_{t-1} - 0.5X_{t-3} + \varepsilon_t$$

où $(\varepsilon_t, t \in \mathbb{Z})$ est un bruit blanc de variance 1 et $\underline{\phi} = (0.8, 0, -0.5)'$.

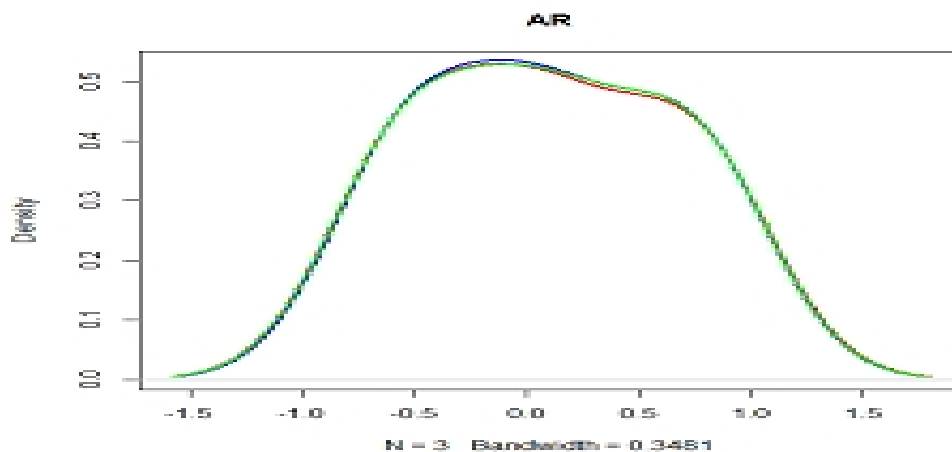
La normalité asymptotique des différentes méthodes est bien illustrée dans la trajectoire au-dessous telle que la distribution de l'estimateur des moindres carrés *OLS* est en rouge, la distribution de l'estimateur de Yule-Walker est en bleue et la distribution de l'estimateur de maximum de vraisemblances est en verte.

Script R

```

YL= OLS= MLE=matrix(NA,5000,3)
for(s in 1 :5000){
X=arima.sim(list(ar=c(.8,0,-.5)),n=240)
YL[s,]=ar(X,order.max=3,method="yw")$ar
OLS[s,]=ar(X,order.max=3,method="ols")$ar[1 :3]
MLE[s,]=ar(X,order.max=3,method="mle")$ar
}
plot(density(OLS[s,]),lwd=2,col="red",main="AR")
lines(density(YL[s,]),lwd=2,col="blue")
lines(density(MLE[s,]),lwd=2,col="green")

```

**5.4 Simulation avec le modèle AR(2)**

Cette section présente des résultats lesquels illustrent les procédures théoriques et les résultats développés de la section précédente, cette simulation a été exécutée avec le modèle $AR(2)$, c'est-à-dire,

$$X_t = \begin{pmatrix} \phi_1 & \phi_2 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ X_{t-2} \end{pmatrix} + \varepsilon_t,$$

pour plusieurs valeurs différentes de $\phi_i, i = 1, 2$ et pour $N \in \{1000, 2000, 3000\}$ où $(\varepsilon_t, t \in \mathbb{Z})$ est un bruit blanc de moyenne nulle et de variance 1. Pour cette simulation, nous avons utilisé le logiciel R et les paramètres ont été estimés par les méthodes :

1. La méthode de Yule-Walker YW.

2. La méthode de maximum de vraisemblance *MLE*.
3. La méthode des moindres carrés *OLS*.

Script R

```
Ar2<-arima.sim(list(order=c(2,0,0),ar=c(+0.5,-0.5)),n=3000)
Ar2.ar <- ar(Ar2, method = "ols", method="mle", method="yw")
Ar2.ar$order
Ar2.ar$ar
rmse<-sqrt(mean(actual-predicted)^2)
```

Vraie valeur $\underline{\phi}$	N	<i>OLS</i>	<i>MLE</i>	<i>YW</i>	<i>RMSE</i>	<i>RMSE</i>	<i>RMSE</i>
		$\hat{\underline{\phi}}^{(ols)}$	$\hat{\underline{\phi}}^{(mle)}$	$\hat{\underline{\phi}}^{(yw)}$	$\hat{\underline{\phi}}^{(ols)}$	$\hat{\underline{\phi}}^{(mle)}$	$\hat{\underline{\phi}}^{(yw)}$
$\begin{pmatrix} -0.8 \\ -0.15 \end{pmatrix}$	1000	$\begin{pmatrix} -0.7752 \\ -0.1177 \end{pmatrix}$	$\begin{pmatrix} -0.8002 \\ -0.1695 \end{pmatrix}$	$\begin{pmatrix} -0.8081 \\ -0.1387 \end{pmatrix}$	$\begin{pmatrix} 0.0248 \\ 0.0323 \end{pmatrix}$	$\begin{pmatrix} 0.0002 \\ 0.0195 \end{pmatrix}$	$\begin{pmatrix} 0.0081 \\ 0.0113 \end{pmatrix}$
	2000	$\begin{pmatrix} -0.7897 \\ -0.1553 \end{pmatrix}$	$\begin{pmatrix} -0.7835 \\ -0.1328 \end{pmatrix}$	$\begin{pmatrix} -0.7486 \\ -0.1281 \end{pmatrix}$	$\begin{pmatrix} 0.0103 \\ 0.0053 \end{pmatrix}$	$\begin{pmatrix} 0.0165 \\ 0.0172 \end{pmatrix}$	$\begin{pmatrix} 0.0514 \\ 0.0219 \end{pmatrix}$
	3000	$\begin{pmatrix} -0.8213 \\ -0.1815 \end{pmatrix}$	$\begin{pmatrix} -0.7895 \\ -0.1587 \end{pmatrix}$	$\begin{pmatrix} -0.8190 \\ -0.1699 \end{pmatrix}$	$\begin{pmatrix} 0.0213 \\ 0.0315 \end{pmatrix}$	$\begin{pmatrix} 0.0105 \\ 0.0087 \end{pmatrix}$	$\begin{pmatrix} 0.019 \\ 0.0199 \end{pmatrix}$

Vraie valeur $\underline{\phi}$	N	<i>OLS</i>	<i>MLE</i>	<i>YW</i>	<i>RMSE</i>	<i>RMSE</i>	<i>RMSE</i>
		$\hat{\underline{\phi}}^{(ols)}$	$\hat{\underline{\phi}}^{(mle)}$	$\hat{\underline{\phi}}^{(yw)}$	$\hat{\underline{\phi}}^{(ols)}$	$\hat{\underline{\phi}}^{(mle)}$	$\hat{\underline{\phi}}^{(yw)}$
$\begin{pmatrix} -0.8 \\ 0.15 \end{pmatrix}$	1000	$\begin{pmatrix} -0.8200 \\ 0.1249 \end{pmatrix}$	$\begin{pmatrix} -0.7927 \\ 0.1949 \end{pmatrix}$	$\begin{pmatrix} -0.8333 \\ 0.1152 \end{pmatrix}$	$\begin{pmatrix} 0.0200 \\ 0.0251 \end{pmatrix}$	$\begin{pmatrix} 0.0073 \\ 0.0449 \end{pmatrix}$	$\begin{pmatrix} 0.0333 \\ 0.0348 \end{pmatrix}$
	2000	$\begin{pmatrix} -0.8169 \\ 0.1366 \end{pmatrix}$	$\begin{pmatrix} -0.8049 \\ 0.1413 \end{pmatrix}$	$\begin{pmatrix} -0.8496 \\ 0.1691 \end{pmatrix}$	$\begin{pmatrix} 0.0170 \\ 0.0134 \end{pmatrix}$	$\begin{pmatrix} 0.0050 \\ 0.0090 \end{pmatrix}$	$\begin{pmatrix} 0.0500 \\ 0.0191 \end{pmatrix}$
	3000	$\begin{pmatrix} -0.8093 \\ 0.1272 \end{pmatrix}$	$\begin{pmatrix} -0.7914 \\ 0.1586 \end{pmatrix}$	$\begin{pmatrix} -0.8030 \\ 0.1479 \end{pmatrix}$	$\begin{pmatrix} 0.0100 \\ 0.0228 \end{pmatrix}$	$\begin{pmatrix} 0.0086 \\ 0.0086 \end{pmatrix}$	$\begin{pmatrix} 0.0300 \\ 0.0021 \end{pmatrix}$

Vraie valeur $\underline{\phi}$	N	<i>OLS</i>	<i>MLE</i>	<i>YW</i>	<i>RMSE</i>	<i>RMSE</i>	<i>RMSE</i>
		$\hat{\underline{\phi}}^{(ols)}$	$\hat{\underline{\phi}}^{(mie)}$	$\hat{\underline{\phi}}^{(yw)}$	$\hat{\underline{\phi}}^{(ols)}$	$\hat{\underline{\phi}}^{(mie)}$	$\hat{\underline{\phi}}^{(yw)}$
$\begin{pmatrix} -0.2 \\ 0.2 \end{pmatrix}$	1000	$\begin{pmatrix} -0.1695 \\ 0.2121 \end{pmatrix}$	$\begin{pmatrix} -0.1681 \\ 0.1872 \end{pmatrix}$	$\begin{pmatrix} -0.1733 \\ 0.2182 \end{pmatrix}$	$\begin{pmatrix} 0.0305 \\ 0.0121 \end{pmatrix}$	$\begin{pmatrix} 0.0319 \\ 0.0128 \end{pmatrix}$	$\begin{pmatrix} 0.0267 \\ 0.0182 \end{pmatrix}$
	2000	$\begin{pmatrix} -0.1991 \\ 0.1824 \end{pmatrix}$	$\begin{pmatrix} -0.1769 \\ 0.2264 \end{pmatrix}$	$\begin{pmatrix} -0.2224 \\ 0.2169 \end{pmatrix}$	$\begin{pmatrix} 0.0009 \\ 0.0176 \end{pmatrix}$	$\begin{pmatrix} 0.0231 \\ 0.0264 \end{pmatrix}$	$\begin{pmatrix} 0.0224 \\ 0.0169 \end{pmatrix}$
	3000	$\begin{pmatrix} -0.1852 \\ 0.2133 \end{pmatrix}$	$\begin{pmatrix} -0.2267 \\ 0.2302 \end{pmatrix}$	$\begin{pmatrix} -0.1750 \\ 0.1968 \end{pmatrix}$	$\begin{pmatrix} 0.0148 \\ 0.0133 \end{pmatrix}$	$\begin{pmatrix} 0.0267 \\ 0.0302 \end{pmatrix}$	$\begin{pmatrix} 0.025 \\ 0.0032 \end{pmatrix}$

Vraie valeur $\underline{\phi}$	N	<i>OLS</i>	<i>MLE</i>	<i>YW</i>	<i>RMSE</i>	<i>RMSE</i>	<i>RMSE</i>
		$\hat{\underline{\phi}}^{(ols)}$	$\hat{\underline{\phi}}^{(mie)}$	$\hat{\underline{\phi}}^{(yw)}$	$\hat{\underline{\phi}}^{(ols)}$	$\hat{\underline{\phi}}^{(mie)}$	$\hat{\underline{\phi}}^{(yw)}$
$\begin{pmatrix} 0.5 \\ -0.5 \end{pmatrix}$	1000	$\begin{pmatrix} 0.4753 \\ -0.4907 \end{pmatrix}$	$\begin{pmatrix} 0.5012 \\ -0.5089 \end{pmatrix}$	$\begin{pmatrix} 0.5286 \\ -0.4722 \end{pmatrix}$	$\begin{pmatrix} 0.0247 \\ 0.0093 \end{pmatrix}$	$\begin{pmatrix} 0.0012 \\ 0.0089 \end{pmatrix}$	$\begin{pmatrix} 0.0286 \\ 0.0278 \end{pmatrix}$
	2000	$\begin{pmatrix} 0.4773 \\ -0.4907 \end{pmatrix}$	$\begin{pmatrix} 0.4814 \\ -0.4828 \end{pmatrix}$	$\begin{pmatrix} 0.5208 \\ -0.4799 \end{pmatrix}$	$\begin{pmatrix} 0.0227 \\ 0.0093 \end{pmatrix}$	$\begin{pmatrix} 0.0186 \\ 0.0172 \end{pmatrix}$	$\begin{pmatrix} 0.0208 \\ 0.0201 \end{pmatrix}$
	3000	$\begin{pmatrix} 0.4992 \\ -0.5149 \end{pmatrix}$	$\begin{pmatrix} 0.4826 \\ -0.5009 \end{pmatrix}$	$\begin{pmatrix} 0.4974 \\ -0.5005 \end{pmatrix}$	$\begin{pmatrix} 0.0000 \\ 0.0149 \end{pmatrix}$	$\begin{pmatrix} 0.0174 \\ 0.0000 \end{pmatrix}$	$\begin{pmatrix} 0.0026 \\ 0.0000 \end{pmatrix}$

Vraie valeur $\underline{\phi}$	N	<i>OLS</i>	<i>MLE</i>	<i>YW</i>	<i>RMSE</i>	<i>RMSE</i>	<i>RMSE</i>
		$\hat{\underline{\phi}}^{(ols)}$	$\hat{\underline{\phi}}^{(mle)}$	$\hat{\underline{\phi}}^{(yw)}$	$\hat{\underline{\phi}}^{(ols)}$	$\hat{\underline{\phi}}^{(mle)}$	$\hat{\underline{\phi}}^{(yw)}$
$\begin{pmatrix} 0 \\ 0.5 \end{pmatrix}$	1000	$\begin{pmatrix} -0.0375 \\ 0.5014 \end{pmatrix}$	$\begin{pmatrix} 0.0196 \\ 0.5317 \end{pmatrix}$	$\begin{pmatrix} -0.0208 \\ 0.5002 \end{pmatrix}$	$\begin{pmatrix} 0.0265 \\ 0.0014 \end{pmatrix}$	$\begin{pmatrix} 0.0275 \\ 0.0317 \end{pmatrix}$	$\begin{pmatrix} 0.0200 \\ 0.0002 \end{pmatrix}$
	2000	$\begin{pmatrix} 0.0007 \\ 0.5074 \end{pmatrix}$	$\begin{pmatrix} 0.0897 \\ 0.4765 \end{pmatrix}$	$\begin{pmatrix} -0.0260 \\ 0.4661 \end{pmatrix}$	$\begin{pmatrix} 0.0195 \\ 0.0074 \end{pmatrix}$	$\begin{pmatrix} 0.0205 \\ 0.0235 \end{pmatrix}$	$\begin{pmatrix} 0.0260 \\ 0.0339 \end{pmatrix}$
	3000	$\begin{pmatrix} 0.0072 \\ 0.4717 \end{pmatrix}$	$\begin{pmatrix} -0.0300 \\ 0.5316 \end{pmatrix}$	$\begin{pmatrix} 0.0250 \\ 0.4727 \end{pmatrix}$	$\begin{pmatrix} 0.0163 \\ 0.0283 \end{pmatrix}$	$\begin{pmatrix} 0.0164 \\ 0.0316 \end{pmatrix}$	$\begin{pmatrix} 0.0200 \\ 0.0273 \end{pmatrix}$

Vraie valeur $\underline{\phi}$	N	<i>OLS</i>	<i>MLE</i>	<i>YW</i>	<i>RMSE</i>	<i>RMSE</i>	<i>RMSE</i>
		$\hat{\underline{\phi}}^{(ols)}$	$\hat{\underline{\phi}}^{(mle)}$	$\hat{\underline{\phi}}^{(yw)}$	$\hat{\underline{\phi}}^{(ols)}$	$\hat{\underline{\phi}}^{(mle)}$	$\hat{\underline{\phi}}^{(yw)}$
$\begin{pmatrix} 0.5 \\ 0 \end{pmatrix}$	1000	$\begin{pmatrix} 0.5134 \\ -0.0878 \end{pmatrix}$	$\begin{pmatrix} 0.5153 \\ -0.0319 \end{pmatrix}$	$\begin{pmatrix} 0.4744 \\ -0.0561 \end{pmatrix}$	$\begin{pmatrix} 0.0134 \\ 0.0290 \end{pmatrix}$	$\begin{pmatrix} 0.0153 \\ 0.0341 \end{pmatrix}$	$\begin{pmatrix} 0.0256 \\ 0.0260 \end{pmatrix}$
	2000	$\begin{pmatrix} 0.5332 \\ -0.0927 \end{pmatrix}$	$\begin{pmatrix} 0.5017 \\ -0.0345 \end{pmatrix}$	$\begin{pmatrix} 0.5228 \\ -0.0447 \end{pmatrix}$	$\begin{pmatrix} 0.0332 \\ 0.0220 \end{pmatrix}$	$\begin{pmatrix} 0.0020 \\ 0.0240 \end{pmatrix}$	$\begin{pmatrix} 0.0228 \\ 0.0340 \end{pmatrix}$
	3000	$\begin{pmatrix} 0.4933 \\ 0.0329 \end{pmatrix}$	$\begin{pmatrix} 0.4713 \\ 0.0417 \end{pmatrix}$	$\begin{pmatrix} 0.4892 \\ 0.0276 \end{pmatrix}$	$\begin{pmatrix} 0.0067 \\ 0.0188 \end{pmatrix}$	$\begin{pmatrix} 0.0287 \\ 0.0193 \end{pmatrix}$	$\begin{pmatrix} 0.0108 \\ 0.0187 \end{pmatrix}$

Vraie valeur $\underline{\phi}$	N	OLS	MLE	YW	$RMSE$	$RMSE$	$RMSE$
		$\hat{\underline{\phi}}^{(ols)}$	$\hat{\underline{\phi}}^{(mie)}$	$\hat{\underline{\phi}}^{(yw)}$	$\hat{\underline{\phi}}^{(ols)}$	$\hat{\underline{\phi}}^{(mie)}$	$\hat{\underline{\phi}}^{(yw)}$
$\begin{pmatrix} 0.8 \\ 0.15 \end{pmatrix}$	1000	$\begin{pmatrix} 0.7703 \\ 0.1819 \end{pmatrix}$	$\begin{pmatrix} 0.8807 \\ 0.1660 \end{pmatrix}$	$\begin{pmatrix} 0.7760 \\ 0.1848 \end{pmatrix}$	$\begin{pmatrix} 0.0297 \\ 0.0319 \end{pmatrix}$	$\begin{pmatrix} 0.0307 \\ 0.020 \end{pmatrix}$	$\begin{pmatrix} 0.0240 \\ 0.0348 \end{pmatrix}$
	2000	$\begin{pmatrix} 0.8013 \\ 0.1372 \end{pmatrix}$	$\begin{pmatrix} 0.8023 \\ 0.1404 \end{pmatrix}$	$\begin{pmatrix} 0.7747 \\ 0.1839 \end{pmatrix}$	$\begin{pmatrix} 0.0130 \\ 0.0128 \end{pmatrix}$	$\begin{pmatrix} 0.0230 \\ 0.0100 \end{pmatrix}$	$\begin{pmatrix} 0.0253 \\ 0.0339 \end{pmatrix}$
	3000	$\begin{pmatrix} 0.7594 \\ 0.1779 \end{pmatrix}$	$\begin{pmatrix} 0.8193 \\ 0.1298 \end{pmatrix}$	$\begin{pmatrix} 0.8053 \\ 0.1399 \end{pmatrix}$	$\begin{pmatrix} 0.0106 \\ 0.0179 \end{pmatrix}$	$\begin{pmatrix} 0.0190 \\ 0.0102 \end{pmatrix}$	$\begin{pmatrix} 0.0153 \\ 0.0101 \end{pmatrix}$

Vraie valeur $\underline{\phi}$	N	OLS	MLE	YW	$RMSE$	$RMSE$	$RMSE$
		$\hat{\underline{\phi}}^{(ols)}$	$\hat{\underline{\phi}}^{(mie)}$	$\hat{\underline{\phi}}^{(yw)}$	$\hat{\underline{\phi}}^{(ols)}$	$\hat{\underline{\phi}}^{(mie)}$	$\hat{\underline{\phi}}^{(yw)}$
$\begin{pmatrix} 0.8 \\ 0.1 \end{pmatrix}$	1000	$\begin{pmatrix} 0.7984 \\ 0.0916 \end{pmatrix}$	$\begin{pmatrix} 0.7632 \\ 0.1234 \end{pmatrix}$	$\begin{pmatrix} 0.7701 \\ 0.1472 \end{pmatrix}$	$\begin{pmatrix} 0.0016 \\ 0.0084 \end{pmatrix}$	$\begin{pmatrix} 0.0368 \\ 0.0234 \end{pmatrix}$	$\begin{pmatrix} 0.0299 \\ 0.0372 \end{pmatrix}$
	2000	$\begin{pmatrix} 0.7743 \\ 0.1181 \end{pmatrix}$	$\begin{pmatrix} 0.7842 \\ 0.0902 \end{pmatrix}$	$\begin{pmatrix} 0.8074 \\ 0.1007 \end{pmatrix}$	$\begin{pmatrix} 0.0257 \\ 0.0181 \end{pmatrix}$	$\begin{pmatrix} 0.0158 \\ 0.0190 \end{pmatrix}$	$\begin{pmatrix} 0.0100 \\ 0.0170 \end{pmatrix}$
	3000	$\begin{pmatrix} 0.7984 \\ 0.1021 \end{pmatrix}$	$\begin{pmatrix} 0.8050 \\ 0.1009 \end{pmatrix}$	$\begin{pmatrix} 0.8057 \\ 0.0836 \end{pmatrix}$	$\begin{pmatrix} 0.0016 \\ 0.0100 \end{pmatrix}$	$\begin{pmatrix} 0.0150 \\ 0.0190 \end{pmatrix}$	$\begin{pmatrix} 0.0157 \\ 0.0164 \end{pmatrix}$

5.4.1 Résultats et commentaires

D'après les résultats empiriques qui sont reportés dans les tables au-dessus, on observe que l'augmentation de la taille d'échantillon joue un rôle essentiel pour l'obtention des meilleurs résultats. Cela est prouvé quand l'estimateur $\hat{\underline{\phi}}$ calculés est plus proche de la vraie valeur, plus la taille d'échantillon est élevée. D'autre part on observe que l'erreur quadratique moyenne $RMSE$ est devenue plus petite lorsque la taille d'échantillon augmente, de plus, on remarque que l'estimateur MLE est proche de l'estimateur YW .

5.5 Estimation des paramètres d'un ARMA

L'estimation des paramètres d'un modèle $ARMA(p, q)$, lorsque les ordres p et q sont supposés connus, peut se réaliser par différentes méthodes dans le domaine temporel.

Dans cette section on va présenter deux méthodes d'estimation : la méthode de maximum de vraisemblance et la méthode des moindres carrés ordinaires. Le but est d'estimer les coefficients des polynômes Φ et Θ , ainsi que σ^2 . Une première approche consiste à déterminer ces valeurs à partir des auto-corrélations. En général, ces estimateurs ne sont pas efficaces. C'est pourquoi, on utilise des estimations préliminaires comme première étape dans des méthodes itératives, du type maximum de vraisemblance ou moindres carrés. Si on a une partie moyenne mobile, on utilise l'algorithme des innovations.

On considère ici le cas d'un processus $ARMA(p, q)$ canonique minimal. Le cas d'un processus $MA(q)$ n'est qu'un cas particulier. Considérons l'écriture $MA(\infty)$ du processus :

$$X_t = \epsilon_t + \sum_{i=1}^{\infty} \psi_i \epsilon_{t-i}.$$

On peut estimer les coefficients $(\psi_i)_{i \in \{1, \dots, N\}}$ par $(\hat{\psi}_1(N), \dots, \hat{\psi}_{p+q}(N))$ à l'aide de l'algorithme des innovations. On se fixe N a priori ; on le prend suffisamment grand pour avoir assez d'équations nécessaires à la résolution du problème, sachant que $\psi_i \xrightarrow{i \rightarrow \infty} 0$.

En posant $\psi_i = 0$ pour $i < 0$, $\theta_0 = 1$ et $\theta_i = 0$ pour $i > q$, on a :

$$\forall i \in \mathbb{N} : \psi_i - \sum_{j=1}^p \phi_j \psi_{i-j} = \theta_i.$$

On utilise ces relations pour obtenir une première estimation de (ϕ_1, \dots, ϕ_p) et $(\theta_1, \dots, \theta_q)$ à partir de $(\hat{\psi}_1(N), \dots, \hat{\psi}_{p+q}(N))$.

5.5.1 Estimation par la méthode du maximum de vraisemblance

En général, on ajoute une hypothèse sur la loi des résidus. On suppose ici que les résidus sont distribués selon une loi normale $N(0, \sigma^2)$, hypothèse qu'il faudrait vérifier à l'aide d'un test (Kolmogorov par exemple).

Si l'hypothèse de normalité n'est pas vérifiée, on peut tout de même considérer que la vraisemblance normale est un critère d'ajustement qui peut convenir.

Si on dispose d'un échantillon $\{X_1, \dots, X_N\}$, la vraisemblance $L(X_1, \dots, X_N, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma^2)$ s'écrit :

$$L(X_1, \dots, X_N, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma^2) = \frac{1}{\sqrt{(2\pi)^N \det(\Gamma_N)}} \exp\left\{-\frac{1}{2} X^t \Gamma_N^{-1} X\right\},$$

où Γ_N est la matrice de variance covariance de $\{X_1, \dots, X_N\}$. Remarquons que (X_1, \dots, X_N) est un vecteur gaussien, en tant que transformation linéaire du vecteur gaussien $(\epsilon_1, \dots, \epsilon_N)$.

La maximisation de cette vraisemblance n'est pas simple et elle nécessite l'utilisation d'algorithmes d'optimisation.

Notons pour tout $i \in \{2, \dots, N\}$, $\hat{X}_i = E\{L(X_i / X_{i-1}, \dots, X_1)\}$.

On peut utiliser l'algorithme des innovations pour calculer les erreurs de prévision au pas 1, $\epsilon_i = X_i - \hat{X}_i$, ainsi que leur variance σ_i . On évite ainsi le calcul direct de Γ_N^{-1} et de $\det(\Gamma_N)$. On a

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{pmatrix} = C_N \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{pmatrix}$$

avec la matrice C_N (estimée par l'algorithme des innovations) définie par

$$C_N = \begin{pmatrix} 1 & 0 & \dots & \dots & 0 \\ \theta_1(1) & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \theta_{N-1}(N-1) & \theta_{N-2}(N-1) & \dots & \theta_1(N-1) & 1 \end{pmatrix}.$$

Les $(\epsilon_i)_{i \in \{1, \dots, N\}} = (X_i - \hat{X}_i)_{i \in \{1, \dots, N\}}$ ne sont pas corrélés, la matrice de covariance de $(\epsilon_i)_{i \in \{1, \dots, N\}}$ vaut

$$V_N = \begin{pmatrix} \sigma_0 & 0 & \dots & 0 \\ 0 & \sigma_1 & \dots & \dots \\ \dots & \dots & \dots & 0 \\ 0 & \dots & 0 & \sigma_{N-1} \end{pmatrix}.$$

On a $\Gamma_N = C_N V_N C_N^t$ et $\det(\Gamma_N) = (\det C_N)^2 \det V_N = \sigma_0 \sigma_1 \dots \sigma_{N-1}$, et donc

$$X^t \Gamma_N^{-1} X = \sum_{i=1}^N \frac{(X_i - \hat{X}_i)^2}{\sigma_{i-1}},$$

où les prévisions \widehat{X}_i sont données par l'algorithme des innovations.

La vraisemblance s'écrit donc :

$$L(X_1, \dots, X_N, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma^2) = \frac{1}{\sqrt{(2\pi)^N \sigma_0 \sigma_1 \dots \sigma_{N-1}}} \exp \left(-\frac{1}{2} \sum_{i=1}^N \frac{(X_i - \widehat{X}_i)^2}{\sigma_{i-1}} \right).$$

L'algorithme des innovations nous indique que :

$$\widehat{X}_{i+1} = \begin{cases} \sum_{j=1}^i \theta_j(i) (X_{i+1-j} - \widehat{X}_{i+1-j}) & \text{si } 1 \leq i \leq \max(p, q) \\ \phi_i X_i + \dots + \phi_p X_{i+1-p} + \sum_{j=1}^q \theta_j(i) (X_{i+1-j} - \widehat{X}_{i+1-j}) & \text{si } i > \max(p, q). \end{cases}$$

On définit $r_i = \frac{\sigma_i}{\sigma^2}$. On peut réécrire la vraisemblance comme suit :

$$\begin{aligned} & L(X_1, \dots, X_N, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma^2) \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \frac{1}{\sqrt{r_0 r_1 \dots r_{N-1}}} \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^N \frac{(X_i - \widehat{X}_i)^2}{r_{i-1}} \right). \end{aligned}$$

On utilise la log-vraisemblance. L'estimateur du maximum de vraisemblance vérifie :

$$\begin{cases} \widehat{\beta} = (\widehat{\phi}_1, \dots, \widehat{\phi}_p, \widehat{\theta}_1, \dots, \widehat{\theta}_q) \\ = \arg \min_{(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)} \left\{ \log \left[\frac{1}{N} S(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q) \right] + \frac{1}{N} \sum_{i=1}^N \log r_{i-1} \right\} \\ \widehat{\sigma}^2 = \frac{1}{N} S(\widehat{\phi}_1, \dots, \widehat{\phi}_p, \widehat{\theta}_1, \dots, \widehat{\theta}_q) \text{ où } S(\widehat{\phi}_1, \dots, \widehat{\phi}_p, \widehat{\theta}_1, \dots, \widehat{\theta}_q) = \sum_{i=1}^N \frac{(X_i - \widehat{X}_i)^2}{r_{i-1}}. \end{cases}$$

Remarque 71 La maximisation s'effectue de manière itérative, et l'estimation préliminaire fournit des valeurs initiales. Les estimateurs obtenus sont efficaces.

5.5.2 Comportement asymptotique

Consistance

Maintenant nous sommes en position de lancer un résultat très important sur la consistance.

Proposition 72 On dit qu'un estimateur de maximum de vraisemblance noté $\hat{\beta} = (\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q)$ de $\beta = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$, obtenu à partir de N observations X_1, \dots, X_N du processus ARMA(p, q), est fortement consistant si et seulement si la suite $\hat{\beta}$ converge presque sûrement vers β , c'est-à-dire,

$$\hat{\beta} \xrightarrow{p.s.} \beta \text{ lorsque } N \rightarrow \infty.$$

Remarque 73 Ce résultat signifie que $\hat{\beta}$ est assurément proche de β quand N est assez grand.

Remarque 74 La convergence presque sûre implique la convergence en probabilité, un estimateur fortement consistant est consistant.

Normalité asymptotique

Le MLE a la propriété d'être normalement réparti de façon asymptotique. C'est énoncé dans les propositions suivantes :

Proposition 75 L'estimateur du maximum de vraisemblance d'un processus ARMA(p, q) causal et inversible, obtenu à partir de N observations X_1, \dots, X_N du processus ARMA(p, q), est asymptotiquement normal, c'est-à-dire,

$$\sqrt{N}(\hat{\beta} - \beta) \rightarrow N(0, V(\beta)_{p+q})$$

où la matrice de variance covariance asymptotique $V(\beta)_{p+q}$ dépend de paramètres du modèle. De plus, elle peut être calculée explicitement. Plus précisément pour $p \geq 1$ et $q \geq 1$, on a

$$V(\beta)_{p+q} = \sigma^2 \begin{bmatrix} E\mathbf{U}_t\mathbf{U}_t' & E\mathbf{U}_t\mathbf{V}_t' \\ E\mathbf{V}_t\mathbf{U}_t' & E\mathbf{V}_t\mathbf{V}_t' \end{bmatrix}^{-1} \quad (5.6)$$

où $\mathbf{U}_t = (U_t, \dots, U_{t+1-p})'$, $\mathbf{V}_t = (V_t, \dots, V_{t+1-q})'$ et $(U_t, t \in \mathbb{Z})$, $(V_t, t \in \mathbb{Z})$ sont les processus autorégressifs,

$$\Phi(B)U_t = \epsilon_t \quad \text{et} \quad \Theta(B)V_t = \epsilon_t.$$

Remarque 76 Pour $p = 0$, $V(\beta)_q = \sigma^2 [E\mathbf{V}_t\mathbf{V}_t']^{-1}$, et pour $q = 0$, $V(\beta)_p = \sigma^2 [E\mathbf{U}_t\mathbf{U}_t']^{-1}$.

Exemple 77 Considérons le cas d'un modèle ARMA (1, 1). Nous avons dans ce cas

$$V(\phi, \theta) = \sigma^2 \begin{bmatrix} EU_t^2 & EU_t V_t \\ EU_t V_t & EV_t^2 \end{bmatrix}^{-1}$$

où

$$U_t - \phi U_{t-1} = \epsilon_t \text{ et } V_t + \theta V_{t-1} = \epsilon_t.$$

Un calcul simple nous donne,

$$V(\phi, \theta) = \begin{bmatrix} (1 - \phi^2)^{-1} & (1 + \phi\theta)^{-1} \\ (1 + \phi\theta)^{-1} & (1 - \theta^2)^{-1} \end{bmatrix}^{-1}$$

D'où l'estimateur $\begin{bmatrix} \hat{\phi} \\ \hat{\theta} \end{bmatrix}$ est asymptotiquement normal

$$\left(\begin{bmatrix} \phi \\ \theta \end{bmatrix}, n^{-1} \frac{1 + \phi\theta}{(\phi + \theta)^2} \begin{bmatrix} (1 - \phi^2)(1 + \phi\theta) & -(1 - \theta^2)(1 - \phi^2) \\ -(1 - \theta^2)(1 - \phi^2) & (1 - \theta^2)(1 + \phi\theta) \end{bmatrix} \right).$$

Remarque 78 Ces distributions asymptotiques fournissent nous, avec une technique générale, à calculer des intervalles de confiance asymptotique pour des estimateurs des moindres carrés ou des estimateurs du maximum de vraisemblance.

Partant de la matrice de variance covariance des estimateurs du maximum de vraisemblance, il est possible de reconstruire les statistiques de Student associées aux différents paramètres du modèle ARMA. Sur le plan appliqué, lorsque le processus $(X_t, t \in \mathbb{Z})$ est stationnaire, on peut montrer que la distribution asymptotique de ces statistiques de Student est asymptotiquement distribuée selon une loi normale. Donc, on peut appliquer à ces estimateurs les méthodes d'inférence traditionnelles.

5.5.3 Estimation par la méthode des moindres carrés ordinaires

On cherche cette fois à minimiser la quantité suivante :

$$S(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q) = \sum_{i=1}^N \frac{(X_i - \hat{X}_i)^2}{r_{i-1}}.$$

L'estimateur des moindres carrés vérifie :

$$\begin{cases} \hat{\mu} = (\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q) = \arg \min_{(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)} S(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q) \\ \hat{\sigma}^2 = \frac{1}{N-p-q} S(\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q). \end{cases}$$

Les estimateurs obtenus sont efficaces.

5.5.4 Comportement asymptotique

Consistance

Maintenant nous sommes en position de donner un résultat très important sur la consistance de l'estimateur obtenu.

Proposition 79 *On dit qu'un estimateur des moindres carrés noté $\hat{\mu} = (\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q)$ de $\mu = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$, obtenu à partir de N observations X_1, \dots, X_N du processus ARMA(p, q), est fortement consistant si et seulement si la suite $\hat{\mu}$ converge presque sûrement vers μ , c'est-à-dire,*

$$\hat{\mu} \xrightarrow{p.s.} \mu \text{ lorsque } N \rightarrow \infty.$$

Remarque 80 *Ce résultat signifie que $\hat{\mu}$ est assurément proche de μ quand N est assez grand.*

Normalité asymptotique

La propriété de la normalité asymptotique de l'estimateur des moindres carrés est énoncée dans la proposition suivante :

Proposition 81 *L'estimateur des moindres carrés d'un processus ARMA(p, q) causal et inversible, obtenu à partir de N observations X_1, \dots, X_N du processus ARMA(p, q), est asymptotiquement normal, c'est-à-dire,*

$$\sqrt{N}(\hat{\mu} - \mu) \rightarrow N(0, V(\mu)_{p+q})$$

où la matrice de variance covariance asymptotique $V(\mu)_{p+q}$ est donnée par (5.6).

Remarque 82 *Il existe d'autres méthodes d'estimation qu'on ne présente pas ici, notamment la méthode du maximum de vraisemblance conditionnel et la méthode des moindres carrés conditionnels.*

5.5.5 Comparaison des différentes méthodes d'inférence

Soit le processus stationnaire ARMA(3, 2)

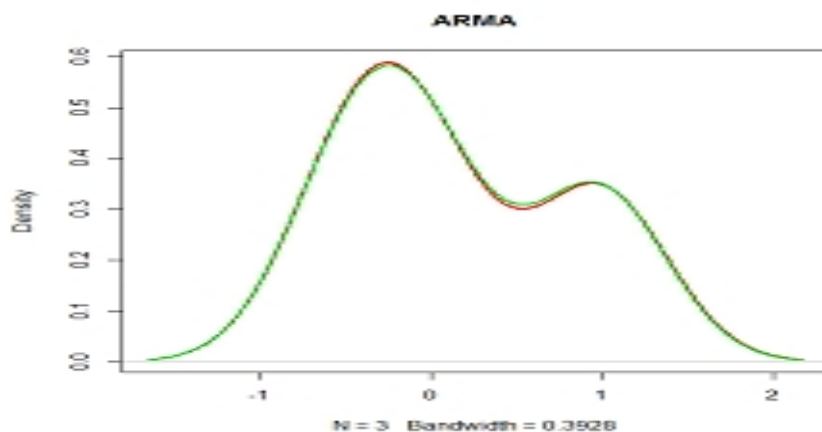
$$X_t = 0.8X_{t-1} - 0.5X_{t-3} + \epsilon_t + 0.4\epsilon_{t-1} + 0.6\epsilon_{t-2}$$

La normalité asymptotique des différentes méthodes est bien illustrée dans la trajectoire au-dessous telle que la distribution de l'estimateur des moindres

carrés *OLS* est en rouge, la distribution de l'estimateur de maximum de vraisemblance *MLE* est en verte.

script R

```
CFOLS= CFMLE=matrix(NA,5000,3)
for(s in 1 :5000){
X=arima.sim(list(ar=c(.8,0,-.5),ma = c(.4,.6)),n=240)
CFOLS[s,]=ar(X,order.max=3,method="ols")$ar[1 :3]
CFMLE[s,]=ar(X,order.max=3,method="mle")$ar
}
plot(density(CFOLS[s,]),lwd=2,col="2",main="ARMA")
lines(density(CFMLE[s,]),lwd=2,col="3")
```



5.6 Simulation avec le modèle $ARMA(1, 1)$

Dans cette section on va présenter les résultats développés dans la section précédente, cette simulation a été exécutée avec le modèle $ARMA(1, 1)$, c'est-à-dire

$$X_t = \phi_1 X_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t$$

pour plusieurs valeurs différentes de ϕ_i et θ_j tq $i = 1$ et $j = 1$ et pour $N \in \{1000, 2000, 3000\}$ ou $(\epsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc de moyenne nulle et de variance σ^2 .

Pour cette simulation, nous avons utilisé le logiciel R et les paramètres ont été estimés par les méthodes suivantes :

- 1-la méthode du maximum de vraisemblance *MLE*
- 2- la méthode des moindres carrés *OLS*.

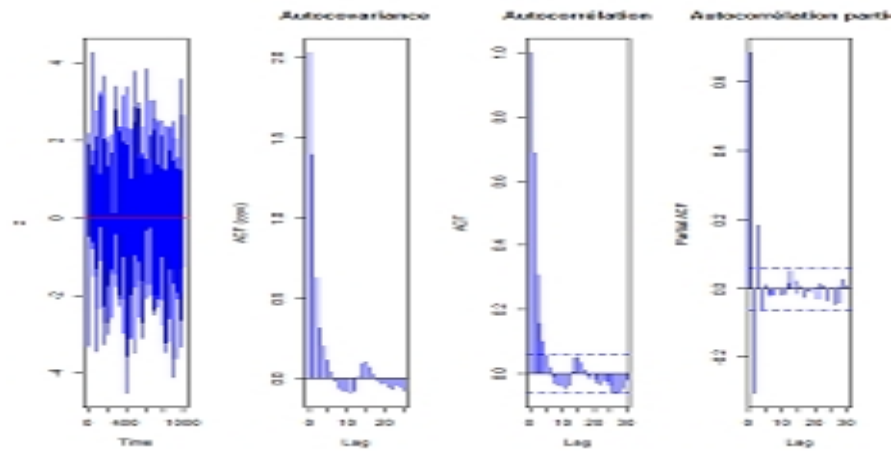
• **Script R**

```

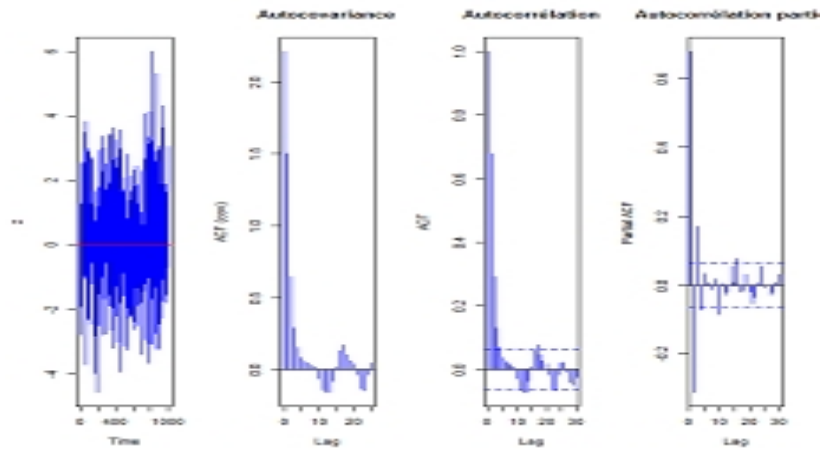
z<-arima.sim(list(order=c(1,0,1),ar=0.5,ma=0.5),n=1000)
fit<-arima(order=c(1,0,1),z,method="ML",include.mean=FALSE)
fit
z<-arima.sim(list(order=c(1,0,1),ar=0.5,ma=0.5),n=1000)
fit<-arima(order=c(1,0,1),z,method="CSS",include.mean=FALSE)
fit
par(mfrow=c(1,4))
plot(z,col="blue")
abline(h=0,col="red")
acf(z,lag=25,type="covariance",main="Autocovariance")
acf(z,col="blue",lag=25,type="covariance",main="Autocovariance")
pacf(z,col="blue",main="Autocorrélation partiel")

```

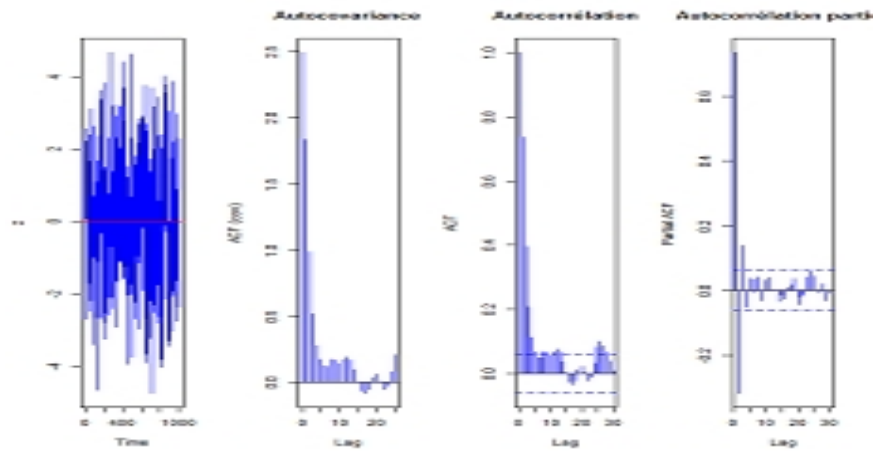
Vraie valeur de ϕ, θ et σ^2	N	MLE	OLS
$\begin{pmatrix} 0.5 \\ 0.5 \\ \sigma^2 \end{pmatrix}$	1000	$\begin{pmatrix} 0.5265 \\ 0.4623 \\ 0.9647 \end{pmatrix}$	$\begin{pmatrix} 0.4761 \\ 0.5581 \\ 1.033 \end{pmatrix}$
	2000	$\begin{pmatrix} 0.5353 \\ 0.4974 \\ 1.032 \end{pmatrix}$	$\begin{pmatrix} 0.5061 \\ 0.5049 \\ 1.036 \end{pmatrix}$
	3000	$\begin{pmatrix} 0.5338 \\ 0.4900 \\ 0.994 \end{pmatrix}$	$\begin{pmatrix} 0.4840 \\ 0.5247 \\ 0.9759 \end{pmatrix}$



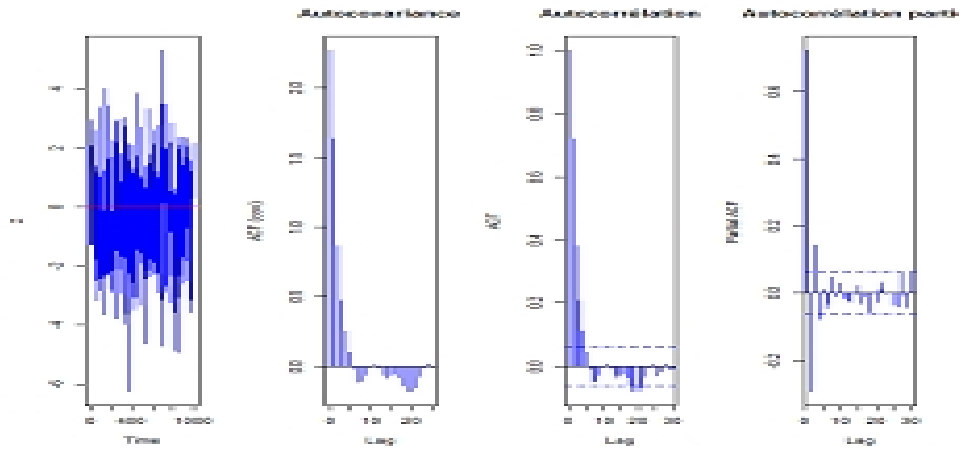
Vraie valeur de ϕ, θ et σ^2	N	MLE	OLS
$\begin{pmatrix} -0.4 \\ -0.8 \\ \sigma^2 \end{pmatrix}$	1000	$\begin{pmatrix} -0.4072 \\ -0.7967 \\ 1.065 \end{pmatrix}$	$\begin{pmatrix} -0.3618 \\ -0.7964 \\ 1.011 \end{pmatrix}$
	2000	$\begin{pmatrix} -0.4290 \\ -0.7862 \\ 0.9531 \end{pmatrix}$	$\begin{pmatrix} -0.3839 \\ -0.7998 \\ 1.03 \end{pmatrix}$
	3000	$\begin{pmatrix} -0.3865 \\ -0.8202 \\ 1.022 \end{pmatrix}$	$\begin{pmatrix} -0.4237 \\ -0.7956 \\ 0.9477 \end{pmatrix}$



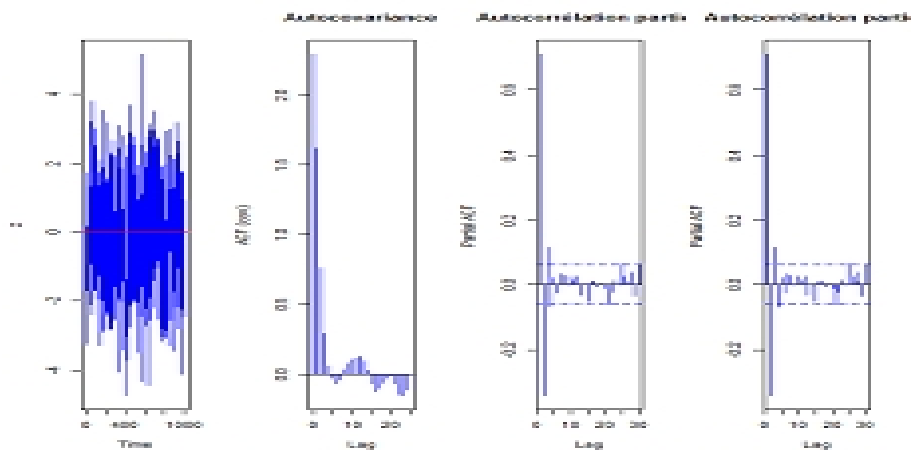
Vraie valeur de ϕ, θ et σ^2	N	MLE	OLS
$\begin{pmatrix} -0.3 \\ 0.7 \\ \sigma^2 \end{pmatrix}$	1000	$\begin{pmatrix} -0.2792 \\ 0.7098 \\ 0.966 \end{pmatrix}$	$\begin{pmatrix} -0.3013 \\ 0.7108 \\ 0.9807 \end{pmatrix}$
	2000	$\begin{pmatrix} -0.2801 \\ 0.6885 \\ 1.013 \end{pmatrix}$	$\begin{pmatrix} -0.3157 \\ 0.6991 \\ 0.9844 \end{pmatrix}$
	3000	$\begin{pmatrix} -0.3009 \\ 0.6865 \\ 0.981 \end{pmatrix}$	$\begin{pmatrix} -0.2874 \\ -0.7087 \\ 1.038 \end{pmatrix}$



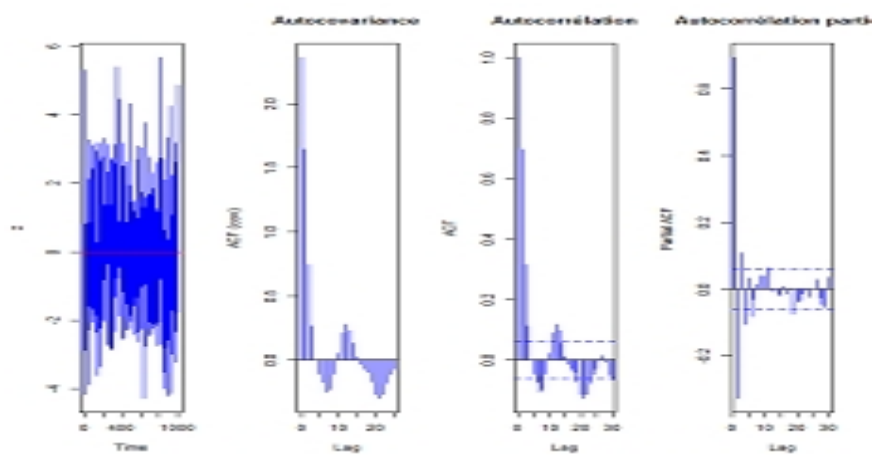
Vraie valeur de ϕ, θ et σ^2	N	MLE	OLS
$\begin{pmatrix} 0.2 \\ -0.5 \\ \sigma^2 \end{pmatrix}$	1000	$\begin{pmatrix} 0.2553 \\ -0.5559 \\ 1.025 \end{pmatrix}$	$\begin{pmatrix} 0.2699 \\ -0.5823 \\ 1.006 \end{pmatrix}$
	2000	$\begin{pmatrix} 0.1001 \\ -0.4095 \\ 0.989 \end{pmatrix}$	$\begin{pmatrix} 0.1631 \\ -0.4950 \\ 1.002 \end{pmatrix}$
	3000	$\begin{pmatrix} 0.2136 \\ -0.5369 \\ 1.003 \end{pmatrix}$	$\begin{pmatrix} 0.2265 \\ -0.5278 \\ 1.011 \end{pmatrix}$



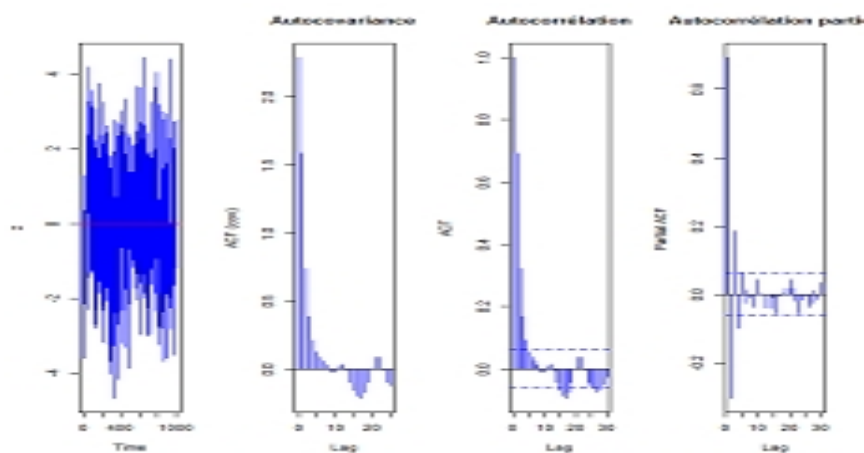
Vraie valeur de ϕ, θ et σ^2	N	MLE	OLS
$\begin{pmatrix} 0.12 \\ 0 \\ \sigma^2 \end{pmatrix}$	1000	$\begin{pmatrix} 0.0896 \\ 0 \\ 0.9614 \end{pmatrix}$	$\begin{pmatrix} 0.1698 \\ 0 \\ 0.9958 \end{pmatrix}$
	2000	$\begin{pmatrix} 0.1060 \\ 0 \\ 0.965 \end{pmatrix}$	$\begin{pmatrix} 0.1194 \\ 0 \\ 1.04 \end{pmatrix}$
	3000	$\begin{pmatrix} 0.1272 \\ 0 \\ 0.9964 \end{pmatrix}$	$\begin{pmatrix} 0.1418 \\ 0 \\ 1.034 \end{pmatrix}$



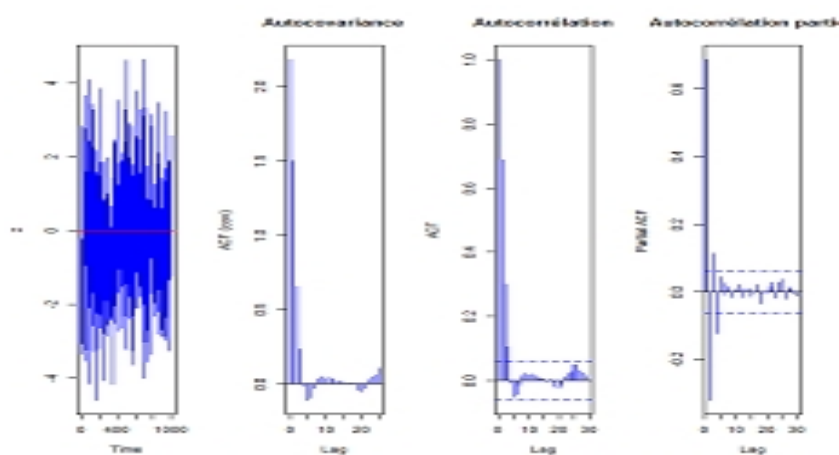
Vraie valeur de ϕ, θ et σ^2	N	MLE	OLS
$\begin{pmatrix} -0.12 \\ 0 \\ \sigma^2 \end{pmatrix}$	1000	$\begin{pmatrix} -0.1515 \\ 0 \\ 1.006 \end{pmatrix}$	$\begin{pmatrix} -0.0928 \\ 0 \\ 1.002 \end{pmatrix}$
	2000	$\begin{pmatrix} -0.1029 \\ 0 \\ 1.026 \end{pmatrix}$	$\begin{pmatrix} -0.1443 \\ 0 \\ 0.9737 \end{pmatrix}$
	3000	$\begin{pmatrix} -0.1122 \\ 0 \\ 0.9949 \end{pmatrix}$	$\begin{pmatrix} -0.1338 \\ 0 \\ 1.047 \end{pmatrix}$



Vraie valeur de ϕ, θ et σ^2	N	MLE	OLS
$\begin{pmatrix} -0.25 \\ -0.25 \\ \sigma^2 \end{pmatrix}$	1000	$\begin{pmatrix} -0.2681 \\ -0.2261 \\ 1.014 \end{pmatrix}$	$\begin{pmatrix} -0.2785 \\ -0.2014 \\ 0.9603 \end{pmatrix}$
	2000	$\begin{pmatrix} -0.2211 \\ -0.2672 \\ 1.026 \end{pmatrix}$	$\begin{pmatrix} -0.1818 \\ -0.3498 \\ 0.9896 \end{pmatrix}$
	3000	$\begin{pmatrix} -0.2198 \\ -0.2798 \\ 1.059 \end{pmatrix}$	$\begin{pmatrix} -0.2738 \\ -0.2365 \\ 1.013 \end{pmatrix}$



Vraie valeur de ϕ, θ et σ^2	N	MLE	OLS
$\begin{pmatrix} 0.15 \\ 0.9 \\ \sigma^2 \end{pmatrix}$	1000	$\begin{pmatrix} 0.1400 \\ 0.9153 \\ 0.9858 \end{pmatrix}$	$\begin{pmatrix} 0.1279 \\ 0.9032 \\ 0.996 \end{pmatrix}$
	2000	$\begin{pmatrix} 0.1361 \\ 0.9224 \\ 0.9939 \end{pmatrix}$	$\begin{pmatrix} 0.1037 \\ 0.91 \\ 0.9715 \end{pmatrix}$
	3000	$\begin{pmatrix} 0.1355 \\ 0.9060 \\ 0.9834 \end{pmatrix}$	$\begin{pmatrix} 0.1458 \\ 0.9171 \\ 0.9779 \end{pmatrix}$



5.6.1 Résultats et commentaires

D'après les résultats empiriques qui sont reportés dans les tables au-dessus, on observe que l'augmentation de la taille d'échantillon joue un rôle essentiel pour l'obtention des meilleurs résultats. Cela est prouvé quand les estimateurs $\hat{\phi}$ et $\hat{\theta}$ calculés est plus proche de la vraie valeur et $\hat{\sigma}^2$ plus proche de 1, plus la taille d'échantillon est élevée.

Chapitre 6

Modèles (G)ARCH

L'observation de certaines séries chronologiques provenant du monde économique et financier (par exemple, des taux d'échanges, des taux d'actions, des indices, ...) montre des caractéristiques spécifiques qui ne sont pas théoriquement prises en compte dans la modélisation de Box-Jenkins. Plus particulièrement, nous observons un comportement de cette variabilité en cluster, c'est-à-dire, une tendance de la série à grouper les périodes de grande variabilité, tandis que de petites variations sont plutôt suivies par de petits changements. Un tel comportement (qualifié d'hétéroscédastique) ne peut certainement pas être expliqué avec les modèles linéaires à variance constante (ou homoscédastique) ou ceux du type Box-Jenkins. L'aggrégation de ces périodes de variabilité élevée fait partie des caractéristiques des séries financières appelées "stylized facts".

Un autre phénomène de ce type concerne la distribution probabiliste de ces données. L'apparition de chocs (c'est-à-dire d'observations isolées d'extrêmement élevées valeur absolue) n'est pas compatible avec une loi normale, mais nécessite plutôt une modélisation avec une distribution plus étalée (heavy tails). Nous verrons selon l'étude des moments d'ordre élevé des observations que l'on doit constater une non compatibilité avec les moments impliqués par une loi normale.

Dans la littérature des quinze dernières années, la modélisation du type (G)ARCH s'est montrée utile dans beaucoup d'applications du monde économique et financier. Nous en citerons seulement quelques unes à titre d'exemple.

Les séries typiques du monde financier possédant ces stylized facts sont : le pricing des assets, des tests de volatilité d'efficacité du marché, l'estimation du risque systématique des modèles du marché et de la structure des taux

d'intérêts, le développement des stratégies dynamiques de hedging, l'option pricing, la modélisation des primes de risque,... Dans la macro-économie, on peut citer la construction des portefeuilles, des mesures d'incertitude d'inflation, l'étude des effets des décisions des banques centrales, la caractérisation de la relation entre macro-économie et le marché des actions.

Nous allons voir dans ce chapitre comment il est possible de prendre en compte d'autres caractéristiques importantes d'une série chronologique observée. Dans un premier temps nous décrirons empiriquement quelles sont ces nouvelles caractéristiques, avant d'introduire une famille de processus qui permettent de tenir compte de ces nouvelles observations : la famille des modèles *ARCH*.

Pour bien décrire ces modèles dans les observations, nous nous servirons de la notion de moment conditionnel que nous expliquerons en détails. Elle permettra de modéliser par exemple la variabilité instantanée des observations. Ensuite, nous discuterons des propriétés principales de la classe des processus non linéaires *ARCH* et de sa généralisation naturelle : la classe des processus *GARCH*. Alors dans ce chapitre nous avons présenté d'abord la définition et la représentation des modèles *GARCH*. En suite, nous établissons la condition de stationnarité strict et au second-ordre, à partir du modèle *GARCH* de première ordre, car les preuves sont plus facile et les résultats sont plus explicites. Ensuite, nous étudierons le problème de l'estimation et de l'ajustement d'un modèle de type (G)*ARCH*, où l'estimation par maximum de vraisemblance jouera un rôle prépondérant.

6.1 Généralité sur les modèles GARCH(p,q)

Pour limiter le problème de la représentation *ARMA*(p, q), Engle [3] propose une nouvelle classe de modèles autorégressifs conditionnellement hétéroscédastique (*ARCH*) et leur extension *GARCH* est due à Bollerslev [1]. Dans ces modèles, celle-ci s'écrit comme une fonction affine des valeurs passées du carré de la série.

Nous donnons une première définition d'un processus *GARCH* fondée sur les deux premiers moments conditionnels.

Définition 83 On dit que $(\epsilon_t, t \in \mathbb{Z})$ est un processus *GARCH*(p, q) semi-fort si

$$i) E(\epsilon_t / F_{t-1}) = 0, \forall t \in \mathbb{Z}.$$

ii) il exist des constantes $w, \alpha_i, i = 1, \dots, q$ et $\beta_j, j = 1, \dots, p$ telles que

$$\sigma_t^2 = \text{Var}(\epsilon_t / F_{t-1}) = w + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \quad t \in \mathbb{Z} \quad (6.1)$$

où F_{t-1} est l'information disponible jusqu'à l'instant $(t-1)$ donc F_{t-1} est la σ -algèbre engendrée par $(\epsilon_{t-1}, \epsilon_{t-2}, \dots)$.

La définition plus restrictive suivante permettra d'obtenir explicitement des processus solutions.

Définition 84 Soit (η_t) une suite de variables i.i.d tel que, $E(\eta_t) = 0, E(\eta_t^2) = 1$. On dit que (ϵ_t) est un processus GARCH(p, q) au sens fort s'il vérifie :

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = w + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \end{cases} \quad t \in \mathbb{Z} \quad (6.2)$$

où les α_i et β_j sont positives et w est une constante strictement positive.

Il est clair qu'un processus GARCH fort tel que σ^2 est mesurable par rapport à la tribu F_{t-1} est un processus GARCH semi-forts mais la réciproque n'est cependant pas vraie.

Remarque 85

1) Si $p = 0$, on a

$$\sigma_t^2 = w + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2$$

dans ce cas le processus est appelé ARCH(q).

2) Dans le cas GARCH(p, q) fort en remplaçant ϵ_{t-i} par $\sigma_{t-i} \eta_{t-i}$ dans l'équation suivante

$$\sigma_t^2 = w + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2$$

on obtient

$$\begin{aligned} \sigma_t^2 &= w + \sum_{i=1}^q \alpha_i \sigma_{t-1}^2 \eta_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \\ \implies \sigma_t^2 &= w + \sum_{i=1}^r a_i (\eta_{t-i}^2) \sigma_{t-i}^2 \end{aligned} \quad (6.3)$$

où $a_i(z) = \alpha_i z + \beta_i, i = 1, \dots, r$ et $r = \max(p, q)$. Cette représentation montre que dans le cas d'un GARCH fort, le processus de volatilité vérifie une équation autorégressive, mais avec coefficients aléatoires.

L'innovation du processus (ϵ_t^2) est par définition la variable $v_t = \epsilon_t^2 - \sigma_t^2$. En remplaçant, dans la représentation $GARCH(p, q)$ suivante

$$\sigma_t^2 = w + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, t \in \mathbb{Z},$$

les variables σ_{t-j}^2 par $\epsilon_{t-j}^2 - v_{t-j}$, on obtient la représentation suivante

$$\epsilon_t^2 = w + \sum_{i=1}^r (\alpha_i + \beta_i) \epsilon_{t-i}^2 + v_t - \sum_{j=1}^p \beta_j v_{t-j}, t \in \mathbb{Z}$$

où $r = \max(p, q)$ avec la convention $\alpha_i = 0$ (resp $\beta_j = 0$) si $i > q$ (resp $j > p$). On retrouve ainsi dans cette équation la structure linéaire des modèles $ARMA$, permettant par exemple un calcul très simple des prévisions linéaires. Sous des hypothèses supplémentaires (impliquant la stationnarité de ϵ_t^2), on peut affirmer que si (ϵ_t) est un $GARCH(p, q)$, (ϵ_t^2) est un processus $ARMA(r, p)$. En particulier, le carré d'un processus $ARCH(q)$ admet, s'il est stationnaire, une représentation $AR(q)$. Ces représentations $ARMA$ seront utiles pour l'estimation et l'identification des processus $GARCH$.

6.2 Etude de Stationnarité

La stationnarité joue un rôle majeur en séries temporelles car elle remplace de manière naturelle l'hypothèse d'observations *i.i.d* en statistique standard. Nous allons chercher sous quelles conditions il existe des processus stationnaires (au sens strict et au second ordre) vérifiant l'une des définitions précédentes. On s'intéresse plus particulièrement aux solutions non anticipatives du modèle (6.2), c'est-à-dire, aux processus (ϵ_t) tel que ϵ_t soit une fonction mesurable des variables $\eta_{t-s}, s \geq 0$.

Nous examinons d'abord le cas du modèle $GARCH(1, 1)$ qui peut se traiter plus explicitement. On notera, pour $x > 0, \log^+ x = \max(\log x, 0)$.

6.2.1 Cas du modèle $GARCH(1,1)$

Dans le cas $p = q = 1$, le modèle $GARCH(1, 1)$ s'écrit

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = w + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \end{cases} \quad (6.4)$$

tels que $w > 0, \alpha \geq 0, \beta \geq 0$.

6.2.2 Stationnarité stricte

La stationnarité stricte du modèle $GARCH(1,1)$ a été étudiée par Nelson [4].

théorème 86 *On dit que le modèle $GARCH(1,1)$ strictement stationnaire non anticipative ssi :*

$$-\infty \leq \gamma = E \log\{\alpha\eta_t^2 + \beta\} < 0.$$

Alors le processus ϵ_t défini par $\sqrt{h_t}\eta_t$ est l'unique solution strictement stationnaire non anticipative, où

$$h_t = \lim_{n \rightarrow +\infty} [1 + \sum_{n=1}^N a(\eta_{t-1}) \dots a(\eta_{t-n})] w \quad (6.5)$$

Si $\gamma \geq 0$ il n'existe pas de processus $GARCH(1,1)$ fort strictement stationnaire et non anticipative.

Preuve. Supposons que $w > 0$, $\alpha \geq 0$, $\beta \geq 0$, en oubliant l'indice 1 pour α et β

On a

$$\sigma_t^2 = w + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \iff \sigma_t^2 = w + a(\eta_{t-1})\sigma_{t-1}^2$$

tel que

$$\begin{aligned} a(\eta_t) &= \alpha\eta_{t-1}^2 + \beta = w + a(\eta_{t-1})(w + a(\eta_{t-2})\sigma_{t-2}^2) \\ &= w[1 + \sum_{n=1}^N a(\eta_{t-1}) \dots a(\eta_{t-n})] + a(\eta_{t-1}) \dots a(\eta_{t-N-1})\sigma_{t-N-1}^2 \\ &= h_t(N) + a(\eta_{t-1}) \dots a(\eta_{t-N-1})\sigma_{t-N-1}^2. \end{aligned} \quad (6.6)$$

Posons

$$\begin{aligned} h_t &= \lim_{N \rightarrow +\infty} h_t(N) \\ h_t &= \lim_{N \rightarrow +\infty} [1 + \sum_{n=1}^N a(\eta_{t-1}) \dots a(\eta_{t-n})] w. \end{aligned} \quad (6.7)$$

La limite existant toujours dans $\overline{\mathbb{R}^+} = [0, +\infty]$ puisque les termes de la sommes sont positifs. De plus, en faisant tendre N vers l'infini dans la relation suivante

$$h_t(N) = w + a(\eta_{t-1}) h_{t-1}(N-1)$$

on obtient

$$h_t = w + a(\eta_{t-1}) h_{t-1}.$$

Nous allons montrer que h_t est presque sûrement finie si et seulement si $\gamma < 0$.

Supposons $\gamma < 0$ et $\alpha > 0$ ou $\beta > 0$, car si ces deux coefficients sont nuls, le processus est un bruit blanc fort. On a donc $a(\eta_t) > 0$ p.s, posons

$$\gamma = E \log\{a(\eta_t)\}$$

$\gamma \in [-\infty, +\infty[$ car

$$E \log^+\{a(\eta_t)\} \leq E a(\eta_t) = \alpha + \beta$$

On utilise la règle de Cauchy pour les séries à termes positifs. On a

$$\begin{aligned} [a(\eta_{t-1}) \dots a(\eta_{t-n})]^{1/n} &= \exp\left(\log [a(\eta_{t-1}) \dots a(\eta_{t-n})]^{1/n}\right) \\ &= \exp\left[\frac{1}{n} \log \left\{ \prod_{i=1}^n a(\eta_{t-i}) \right\}\right] \\ &= \exp\left[\frac{1}{n} \sum_{i=1}^n \log \{a(\eta_{t-i})\}\right] \rightarrow e^\gamma \text{ p.s.} \end{aligned}$$

quand $n \rightarrow \infty$, par application de la loi forte des grands nombres à la suite *i.i.d* ($\log\{a(\eta_t)\}$).

D'après la règle de Cauchy pour (h_t) presque sûrement fini, il suffit que $\gamma < 0$ dans ce cas la série converge, de plus, h_t est à valeurs réelles positives vérifie :

$$h_t = w + a(\eta_{t-1}) h_{t-1}$$

Par la suite le processus (ϵ_t) défini par :

$$\epsilon_t = \sqrt{h_t} \eta_t = \left\{ w + \sum_{n=1}^{\infty} a(\eta_{t-1}) \dots a(\eta_{t-i}) w \right\}^{\frac{1}{2}} \eta_t$$

est strictement stationnaire.

Nous allons montrer l'unicité en raisonnant par l'absurde. Soit $\tilde{\epsilon}_t = \sigma_t \eta_t$ une autre solution strictement stationnaire non anticipative alors :

$$\tilde{\epsilon}_t - \epsilon_t = \eta_t (\sigma_t - \sqrt{h_t})$$

et on a

$$\sigma_t^2 = h_t(N) + a(\eta_{t-1}) \dots a(\eta_{t-N-1}) \sigma_{t-N-1}^2$$

alors

$$\sigma_t^2 - h_t = (h_t(N) - h_t) + a(\eta_{t-1}) \dots a(\eta_{t-N-1}) \sigma_{t-N-1}^2$$

Les terme entre parenthèses à droit tends vers 0 p.s quand $N \rightarrow \infty$ puisque la série h_t converge p.s. Nous avons $a(\eta_{t-1}) \dots a(\eta_{t-n}) \rightarrow 0$ avec une probabilité 1, comme $N \rightarrow \infty$ de plus, la distribution de σ_{t-N-1}^2 est indépendant de N par la stationnarité donc

$$a(\eta_{t-1}) \dots a(\eta_{t-N-1}) \sigma_{t-N-1}^2 \rightarrow 0$$

lorsque $N \rightarrow \infty$. Nous avons prouvé que

$$\sigma_t^2 - h_t \rightarrow 0$$

en probabilité lorsque $N \rightarrow \infty$, ce terme est indépendant de N donc nous avons nécessairement $\sigma_t^2 = h_t$ quel que soit t .

Si $\gamma > 0$ d'après la règle de Cauchy la série

$$\sum_{n=1}^N a(\eta_{t-1}) \dots a(\eta_{t-N}) \rightarrow +\infty$$

p.s. lorsque $N \rightarrow +\infty$. Donc si $w > 0$, $h_t = +\infty$ p.s. D'après (6.6), il est clair que $\sigma_t^2 = +\infty$ p.s. Par suite, il n'existe pas de solution finie presque sûrement de (6.4). Si $w = 0$, on a la marche aléatoire

$$\log \sigma_t^2 = \gamma + \log \sigma_{t-1}^2 + u_t,$$

où (u_t) est un bruit blanc fort. En se limitant aux solution non anticipatives on voit que u_t est indépendant de $\log \sigma_{t-1}^2$ et l'on conclut à nouveau que $\sigma_t^2 = +\infty$ p.s. d'où, il n'existe pas de solution finie p.s. de (6.4).

Pour $\gamma = 0$, nous procéderons par l'absurde. Supposons qu'il existe une solution strictement stationnaire (ϵ_t, σ_t^2) de (6.4). Nous avons pour $n > 0$,

$$\sigma_0^2 \geq w \left\{ 1 + \sum_{i=1}^n a(\eta_{-1}) \dots a(\eta_{-i}) \right\}$$

d'où, on déduit que $a(\eta_{-1}) \dots a(\eta_{-i}) w$ converge vers zéro p.s, quand $n \rightarrow +\infty$ et de manière équivalente

$$\sum_{n=1}^n \log a(\eta_i) + \log w \rightarrow -\infty \text{ p.s} \quad (6.8)$$

lorsque $n \rightarrow +\infty$. Selon le théorème de Chung Fuchs¹, nous avons

$$\limsup \sum_{i=1}^n \log a(\eta_i) = +\infty$$

¹Si X_1, \dots, X_n est une suite iid telle que $EX_1 = 0$ et $E|X_1| > 0$ alors $\lim_{n \rightarrow \infty} \sup \sum_{i=1}^n X_i = +\infty$ et $\lim_{n \rightarrow \infty} \inf \sum_{i=1}^n X_i = -\infty$.

avec probabilité 1, qui contredit (6.8). ■

Remarque 87

1) Dans le cas où $w = 0$ et $\gamma < 0$, il est clair que la seule solution strictement stationnaire du modèle est $\epsilon_t = 0$.

2) On voit que la condition $\gamma < 0$ dépend de la loi du processus (η_t) et qu'elle n'est pas symétrique en α et β .

3) Dans le cas ARCH(1) ($\beta = 0$), la contrainte de la stationnarité stricte s'écrit

$$0 \leq \alpha < \exp\{-E(\log \eta_t^2)\}$$

par exemple dans le cas où $\eta_t \rightsquigarrow N(0, 1)$ la condition est $\alpha < 3.56$.

6.2.3 Stationnarité au second ordre

théorème 88 Supposons $w > 0$ on dit que le modèle GARCH(1, 1) fort stationnaire au second ordre et non anticipatif

(i) si

$$\alpha + \beta < 1$$

tels que $\alpha \geq 0$, $\beta \geq 0$. Alors le processus (ϵ_t) défini par $\epsilon_t = \sqrt{h_t}\eta_t$ où h_t est défini par (6.7) est l'unique solution stationnaire au second ordre.

(ii) si

$$\alpha + \beta \geq 1$$

il n'existe pas de solution GARCH(1, 1) fort stationnaire au second et non anticipatif.

Preuve. Si (ϵ_t) est un processus GARCH(1, 1) au sens de la définition (6.4), stationnaire au second ordre et non anticipatif, on a

$$\begin{aligned} E(\epsilon_t^2) &= E(E(\epsilon_t^2/\epsilon_{t-1}) = E(\sigma_t^2) \\ &= E(w + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2) \\ &= w + (\alpha + \beta)E(\epsilon_{t-1}^2) \end{aligned}$$

soit

$$(1 - \alpha - \beta)E(\epsilon_t^2) = w$$

est par l'hypothèse $w > 0$ donc il faut que $\alpha + \beta < 1$, on obtient de plus $E(\epsilon_t^2) > 0$.

Inversement : supposons que

$$w > 0, \quad \alpha \geq 0, \quad \beta \geq 0, \quad \alpha + \beta < 1.$$

On recherche une solution stationnaire de la forme (6.4), on remarque d'abord que la condition de stationnarité strict est également vérifiée car d'après l'inégalité de Jensen

$$E \log \{a(\eta_t)\} \leq \log E \{a(\eta_t)\} = \log(\alpha + \beta) < 0.$$

Soit $\epsilon_t = \sqrt{h_t} \eta_t$, où h_t défini en (6.7), la solution strictement stationnaire est montrons qu'elle admet une variance finie, la variable étant une limite croissante de variable aléatoire positives, on peut intervertir espérance et somme infinie et écrire

$$\begin{aligned} E(\epsilon_t^2) &= E(h_t) \\ &= \left[1 + \sum_{n=1}^{+\infty} E \{a(\eta_{t-1}) \dots a(\eta_{t-n})\} \right] w \\ &= \left[1 + \sum_{n=1}^{+\infty} E \{a(\eta_{t-1})\}^n \right] w \\ &= \left[1 + \sum_{n=1}^{+\infty} (\alpha + \beta)^n \right] w = \frac{w}{1 - \alpha - \beta} \end{aligned}$$

Cela suffit à prouver la stationnarité au second ordre de la solution. De plus cette solution est un bruit blanc car $E(\epsilon_t) = E(E(\epsilon_t/\epsilon_{t-1})) = 0$ et pour tout $h > 0$,

$$\text{Cov}(\epsilon_t, \epsilon_{t-h}) = E(\epsilon_{t-h} E(\epsilon_t/\epsilon_{t-1})) = 0.$$

L'unicité de la solution stationnaire au second ordre non anticipative, supposons que $\epsilon'_t = \sqrt{h'_t} \eta_t$ une autre solution stationnaire au second ordre non anticipative. On a

$$|h_t - h'_t| = a(\eta_{t-1}) \dots a(\eta_{t-n}) |h_{t-n-1} - h'_{t-n-1}|.$$

En prenant l'espérance, il vient

$$\begin{aligned} E|h_t - h'_t| &= E \{a(\eta_{t-1}) \dots a(\eta_{t-n})\} E|h_{t-n-1} - h'_{t-n-1}| \\ &= (\alpha + \beta)^n E|h_{t-n-1} - h'_{t-n-1}|. \end{aligned}$$

Notez que la seconde égalité utilise le fait que les solutions sont non anticipative, hypothèse qui n'était pas nécessaire pour établir le caractère unique

de la solution strictement stationnaire. L'espérance de $|h_{t-n-1} - h'_{t-n-1}|$ étant bornée par $E|h_{t-n-1}| + E|h'_{t-n-1}|$, quantité finie et indépendante de n par stationnarité, et puisque $(\alpha + \beta)^n$ tend vers 0 lorsque $n \rightarrow \infty$, on obtient $E|h_t - h'_t| = 0$ et donc $h_t = h'_t$ pour tous t , p.s. ■

6.3 Stationnarité du modèle GARCH(p,q)

Dans le cas général d'un GARCH(p,q) fort, l'écriture vectorielle suivante sera très utile. On a

$$\underline{z}_t = \underline{b}_t + A_t \underline{z}_{t-1} \quad (6.9)$$

où $\underline{b}_t = \underline{b}(\eta_t) = (w\eta_t^2, 0, \dots, w, 0, \dots, 0)' \in \mathbb{R}^{p+q}$, $\underline{z}_t = (\epsilon_t^2, \dots, \epsilon_{t-q+1}^2, \sigma_t^2, \dots, \sigma_{t-p+1}^2)'$ $\in \mathbb{R}^{p+q}$ et

$$A_t = \begin{pmatrix} \alpha_1 \eta_t^2 & \dots & \alpha_q \eta_t^2 & \beta_1 \eta_t^2 & \dots & \beta_p \eta_t^2 \\ & I_{q-1} & & & O_{p-1} & \\ \alpha_1 & \dots & \alpha_q & \beta_1 & \dots & \beta_p \\ & O_{q-1} & & & I_{p-1} & \end{pmatrix} \in \mathbb{R}^{p+q} \quad (6.10)$$

est une matrice de dimension $(p+q) \times (p+q)$. Dans le cas ARCH(q), \underline{z}_t ne contient que ϵ_t^2 et ses $q-1$ premières valeurs passées, et A_t se limite au bloc supérieur gauche de la matrice ci-dessous. L'équation (6.9) constitue un modèle vectoriel autorégressif d'ordre un, avec coefficients positifs et *i.i.d.*

En itérant (6.9), on obtient

$$\begin{aligned} \underline{z}_t &= \underline{b}_t + A_t \underline{z}_{t-1} \\ \underline{z}_t &= \underline{b}_t + A_t (\underline{b}_{t-1} + A_{t-1} \underline{z}_{t-2}) \\ \underline{z}_t &= \underline{b}_t + \sum_{k=1}^{+\infty} A_t A_{t-1} \dots A_{t-k+1} \underline{b}_{t-k} \end{aligned} \quad (6.11)$$

6.3.1 Stationnarité stricte

L'outil principal pour l'étude de la stationnarité stricte est le concept d'exposant de Lyapounov. Soit A une matrice de dimension $(p+q) \times (p+q)$. Son rayon spectral, noté $\rho(A)$, est le plus grand module de ses valeurs propres. Soit $\|\cdot\|$ une norme quelconque sur l'espace des matrices de dimension $(p+q) \times (p+q)$. On a le résultat d'algèbre suivant :

$$\lim_{t \rightarrow +\infty} \frac{1}{t} \log \|A^t\| = \log \rho(A). \quad (6.12)$$

Cette propriété s'étend aux matrices aléatoires.

théorème 89 Soit $(A_t, t \in \mathbb{Z})$ une suite de matrices aléatoires, strictement stationnaire, telle que $E \{\log^+ \|A_t\|\}$ est finie, on a

$$\begin{aligned}\gamma &= \lim_{t \rightarrow +\infty} \frac{1}{t} E (\log \|A_t A_{t-1} \dots A_1\|) \\ &= \inf_{t \in \mathbb{N}^*} \frac{1}{t} E (\log \|A_t A_{t-1} \dots A_1\|) \\ &= \lim_{t \rightarrow +\infty} \frac{1}{t} \log \|A_t A_{t-1} \dots A_1\| \text{ p.s.}\end{aligned}$$

et γ s'appelle l'exposant de Lyapounov associé à la suite de matrices aléatoires $(A_t, t \in \mathbb{Z})$.

Remarque 90

- 1) On a toujours $\gamma \leq E (\log \|A_1\|)$, avec égalité en dimension 1.
- 2) Dans le cas $A = A_t$ pour tout $t \in \mathbb{Z}$, on a

$$\begin{aligned}\gamma &= \lim_{t \rightarrow +\infty} \frac{1}{t} \log \|A_t A_{t-1} \dots A_1\| \\ &= \lim_{t \rightarrow +\infty} \frac{1}{t} \log \|A^t\| \\ &\stackrel{\text{d'après (6.12)}}{\Rightarrow} = \log \rho(A)\end{aligned}$$

3) Toutes les normes sont équivalentes sur un espace de dimension fini, il est facile de voir que γ est indépendant du choix de la norme.

Lemme 91 Soit $(A_t, t \in \mathbb{Z})$ une suite de matrices aléatoires i.i.d telle que $E \{\log^+ \|A_t\|\}$ est finie. Alors

$$\lim_{t \rightarrow \infty} \text{p.s.} \|A_0 \dots A_{-t}\| = 0 \Rightarrow \gamma < 0.$$

théorème 92 On dit que le modèle GARCH(p, q) est strictement stationnaire non anticipative, solution du modèle (6.2) ssi

$$\gamma < 0$$

où γ est l'exposant de Lyapounov de la suite $(A_t, t \in \mathbb{Z})$, cette solution strictement stationnaire est unique.

Preuve. Nous utiliserons la norme définie par $\|A\| = \sum_{ij} |a_{ij}|$. Par commodité la norme sera notée de manière identique quelle que soit la dimension de A . Avec cette convention, la norme est clairement multiplicative : $\|AB\|$

$\leq \|A\| \|B\|$ pour toutes matrices A et B telles que AB existe. Puisque les variables η_t étant de variance finie, toutes les composantes de la matrice A_t sont intégrables. On a donc

$$E \log^+ \|A_t\| \leq E \|A_t\| < \infty.$$

Supposons $\gamma < 0$ et la série

$$\tilde{z}_t = \underline{b}_t + \sum_{n=0}^{\infty} A_t A_{t-1} \dots A_{t-n} \underline{b}_{t-n-1}$$

converge presque sûrement pour t , en effet, en utilisant la multiplicativité de la norme,

$$\|\tilde{z}_t\| \leq \|\underline{b}_t\| + \sum_{n=0}^{+\infty} \|A_t A_{t-1} \dots A_{t-n}\| \|\underline{b}_{t-n-1}\|$$

et

$$\begin{aligned} & \|A_t \dots A_{t-n}\|^{\frac{1}{n}} \|\underline{b}_{t-n-1}\|^{\frac{1}{n}} \\ &= \exp \left\{ \frac{1}{n} \log \|A_t \dots A_{t-n}\| + \frac{1}{n} \log \|\underline{b}_{t-n-1}\| \right\} \xrightarrow{p.s.} e^\gamma < 1. \end{aligned}$$

car, on a

$$n^{-1} \log \|\underline{b}_{t-n-1}\| \rightarrow 0 \text{ p.s.}$$

D'où, et selon la règle de Cauchy, \tilde{z}_t est bien défini dans $(\mathbb{R}_+^*)^{p+q}$. Soit $\tilde{z}_{q+1,t}$ la $q+1$ -ème composante de \tilde{z}_t . En posant $\epsilon_t = \sqrt{\tilde{z}_{q+1,t} \eta_t}$ on définit une solution strictement stationnaire du modèle (6.2) d'après (6.11), ϵ_t s'exprime comme fonction mesurable de $\eta_t, \eta_{t-1}, \dots$. Supposons maintenant qu'il existe une autre solution strictement stationnaire du modèle (6.2), et de manière équivalente une autre solution strictement stationnaire positive \underline{z}'_t de (6.8). Alors,

$$\|\underline{z}_0 - \underline{z}'_0\| = \|A_0 (\underline{z}_{-1} - \underline{z}'_{-1})\| = \dots \leq \|A_0 \dots A_{1-t}\| \|\underline{z}_{-t} - \underline{z}'_{-t}\|$$

On sait que $\|A_0 \dots A_{1-t}\| \rightarrow 0$ presque sûrement quand $t \rightarrow \infty$ et que la loi de $\underline{z}_t - \underline{z}'_{-t}$ est indépendante de t d'où converge vers zéro en probabilité et par conséquent $\underline{z}_0 = \underline{z}'_0$. ■

Le résultat suivant fournit une condition nécessaire simple de stationnarité stricte, sous trois formes différentes.

Corollaire 93 (Conséquence de la stationnarité strict) Soit γ l'exposant de Lyapounov de la suite $(A_t, t \in \mathbb{Z})$ définie par (6.10). Si $\gamma < 0$, les propriétés suivantes sont équivalentes :

- 1) $\sum_{j=1}^p \beta_j < 1$.
- 2) $1 - \beta_1 z_1 - \dots - \beta_p z^p = 0 \Rightarrow |z| > 1$.
- 3) $\rho(B) < 1$. où B est la sous matrice de A_t définie par

$$\begin{pmatrix} \beta_1 & \beta_2 & \dots & \beta_p \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 1 & 0 \end{pmatrix}.$$

6.3.2 Stationnarité au second ordre

Le théorème suivant donne des conditions nécessaires et suffisantes de la stationnarité au second ordre.

théorème 94 *S'il existe un processus GARCH(p, q), au sens fort vérifie la définition (6.2), stationnaire au second ordre, et si $w > 0$, alors*

$$\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1 \quad (6.13)$$

Preuve. Montrons d'abord que la condition (6.13) est nécessaire. Soit (ϵ_t) un processus GARCH(p, q), stationnaire au second ordre. Alors la quantité

$$E(\epsilon_t^2) = E(E(\epsilon_t^2 / \epsilon_{t-1})) = E(\sigma_t^2)$$

est un réel positif, indépendant de t . En prenant l'espérance des deux membres de l'égalité (6.1), on tire donc

$$E(\epsilon_t^2) = w + \sum_{i=1}^q \alpha_i E(\epsilon_t^2) + \sum_{j=1}^p \beta_j E(\epsilon_t^2),$$

soit

$$\left(1 - \sum_{i=1}^q \alpha_i - \sum_{j=1}^p \beta_j\right) E(\epsilon_t^2) = w$$

Puisque w est strictement positif, on doit avoir (6.13).

Supposons maintenant que (6.13) soit vraie et cherchons une solution GARCH au sens fort (6.2) qui soit stationnaire. Pour $t \in \mathbb{Z}$. Définissons les vecteurs à valeurs dans \mathbb{R}^d suivants :

$$Z_k(t) = \begin{cases} 0 & \text{si } k < 0 \\ \underline{b}_t + A_t Z_{k-1}(t-1) & \text{si } k \geq 0 \end{cases}$$

On a

$$Z_k(t) - Z_{k-1}(t) = \begin{cases} 0 & \text{si } k < 0 \\ \underline{b}_t & \text{si } k = 0 \\ A_t Z_{k-1}(t-1) - Z_{k-2}(t-1) & \text{si } k > 0 \end{cases}$$

En itérant ces relations on obtient, pour $k > 0$

$$Z_k(t) - Z_{k-1}(t) = A_t A_{t-1} \dots A_{t-k+1} \underline{b}_{t-k}.$$

Par ailleurs, pour la norme $\|A\| = \sum_{i,j} |a_{ij}|$, on a pour toute matrice aléatoire A à coefficients positifs,

$$E \|A\| = E \sum_{i,j} |a_{ij}| = E \sum_{i,j} a_{ij} = \|E(A)\|$$

Donc pour $k > 0$

$$\begin{aligned} E \|Z_k(t) - Z_{k-1}(t)\| &= E \|A_t A_{t-1} \dots A_{t-k+1} \underline{b}_{t-k}\| \\ &= \|E(A_t A_{t-1} \dots A_{t-k+1} \underline{b}_{t-k})\| \end{aligned}$$

car la matrice $A_t A_{t-1} \dots A_{t-k+1} \underline{b}_{t-k}$ est positive (donc sa norme est la somme de ses éléments). Remarquons que dans le produit $A_t A_{t-1} \dots A_{t-k+1} \underline{b}_{t-k}$ tous les termes sont indépendants (car le processus (η_t) est *i.i.d* et chaque terme du produit est fonction d'une variable η_{t-i} , les dates $t-i$ étant distinctes). Par ailleurs $A = E(A_t)$ et $\underline{b} = E(\underline{b}_t)$ ne dépend évidemment pas de t finalement, pour $k > 0$

$$E \|Z_k(t) - Z_{k-1}(t)\| = \|A^k \underline{b}\| = (1, \dots, 1) A^k \underline{b}$$

puisque tous les termes du vecteur $A^k \underline{b}$ sont positifs. La condition (6.13) implique que les valeurs propres de A sont de module strictement inférieur à 1. En effet, on vérifie que

$$\det(\lambda \mathbb{I}_{p+q} - A) = \lambda^{p+q} \left(1 - \sum_{i=1}^q \alpha_i \lambda^{-i} - \sum_{j=1}^p \beta_j \lambda^{-j} \right)$$

par exemple en retranchant la $q-1$ -ème ligne de A à la première, puis en développant le déterminant par rapport à la première ligne. Donc, si $|\lambda| \geq 1$, en utilisant l'inégalité $|a-b| \geq |a| - |b|$, on obtient

$$|\det(\lambda \mathbb{I}_{p+q} - A)| \geq \left| 1 - \sum_{i=1}^q \alpha_i \lambda^{-i} - \sum_{j=1}^p \beta_j \lambda^{-j} \right| \geq 1 - \sum_{i=1}^q \alpha_i - \sum_{j=1}^p \beta_j > 0$$

d'où $\rho(A) < 1$. Par la suite, en utilisant la décomposition de Jordon et (6.12), il est facile de voir que $A^k \rightarrow 0$ quand $k \rightarrow \infty$. Donc pour t fixé, $Z_k(t)$ converge à la fois aux sens L^1 , en utilisant le critère de Cauchy, et presque-sûr quand $k \rightarrow \infty$. Soit z_t la limite de $(Z_k(t))_k$. A k fixé, le processus $(Z_k(t))_{k \in \mathbb{Z}}$ est strictement stationnaire. Le processus limite z_t l'est donc également. Par ailleurs il est clair que z_t vérifie l'équation (6.9). ■

6.3.3 Région de stationnarité

Pour un modèle $ARCH(2)$ la matrice A_t prend la forme

$$A_t = \begin{pmatrix} \alpha_1 \eta_t^2 & \alpha_2 \eta_t^2 \\ 1 & 0 \end{pmatrix}$$

et la région de stationnarité peut être estimée par simulation. Le tableau suivant donne, pour quelques valeurs des coefficients α_1 et α_2 , la valeur moyenne et l'écart-type (entre parenthèses) obtenus pour 1000 simulation de taille 1000, c'est-à-dire 1000 simulation de

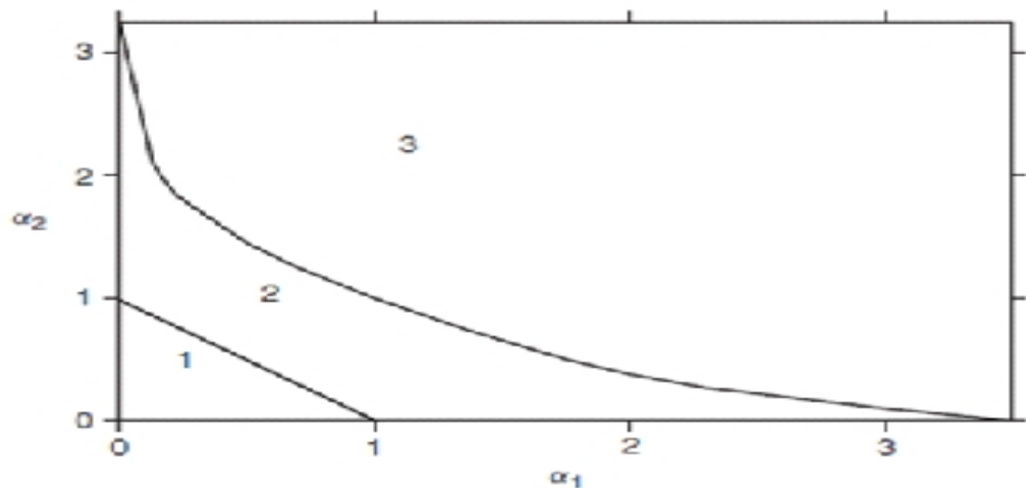
$$\hat{\gamma} = \frac{1}{1000} \log ||A_{1000} A_{999} \dots A_1||$$

les η_t étant tirés selon une loi $N(0, 1)$. Sauf ce cas, il n'existe pas de condition explicite portant sur les coefficients α_1 et α_2 .

	α_1								
α_2	0.25	0.3	1	1.2	1.7	1.8	3.4	3.5	3.6
0	—	—	—	—	—	—	-0.49	-0.18	-0.10
							(0.071)	(0.71)	(0.071)
0.5	—	—	—	-0.175	-0.021	0.006	—	—	—
				(0.040)	(0.042)	(0.044)			
1	—	—	-0.011	0.046	—	—	—	—	—
			(0.038)	(0.038)					
1.75	-0.015	—	—	—	—	—	—	—	—
	(0.035)	(0.032)							

TAB : Estimations de γ à partir de 1000 simulations de taille $t = 1000$

La figure suivante, construite à partir de ces simulation, donne une idée plus précise de la région de stationnarité stricte et second-order du processus $ARCH(2)$.



Régions de stationnarité du modèle ARCH(2)

1 et 2 :stationnarité au second-ordre, 2 :stationnarité stricte, 3 :non stationnarité.

6.4 Estimation des paramètres du modèle GARCH

Dans cette section, nous allons traiter l'estimation des paramètres d'un modèle (G)ARCH. Les modèles introduits reposent sur des formulations des moyennes et variances conditionnelles.

En pratique celle-ci souvent paramétrées de façon que la moyenne conditionnelle $m_t(\underline{\theta})$ et la variance conditionnelle $h_t^2(\underline{\theta})$ apparaissent comme des fonctions de paramètres inconnus et de valeurs passées du processus. La connaissance de ces moments ne suffit cependant pas sans hypothèse supplémentaire à caractériser la loi conditionnelle du processus, la méthode d'estimation envisagée est :

–La méthode de maximum de vraisemblance (MLE), Dans la suite, nous indexons par E_0 les espérances et les variances calculées par rapport à la vraie loi du processus. Soit un modèle ϵ_t tel que :

$$\begin{aligned} E_0(\epsilon_t / F_{t-1}, X_T) &= E_0(\epsilon_t / F_{t-1}) = m_t(\underline{\theta}_0) \\ V_0(\epsilon_t / F_{t-1}, X_T) &= V_0(\epsilon_t / F_{t-1}) = h_t^2(\underline{\theta}_0) \end{aligned}$$

où $\underline{\theta}_0$ est la vraie valeur inconnue du paramètre $\underline{\theta}$ où $\underline{\theta} \in \Theta$.

6.4.1 Méthode de maximum de vraisemblance

Pour comprendre cette approche, nous allons tout d'abord, considérer le cas le plus simple d'un processus ARCH pour ϵ_t . Nous étudierons ensuite le cas des processus GARCH.

L'estimateur des paramètres de modèle ARCH se base très souvent sur la maximisation de la fonction de vraisemblance. Nous supposons que le processus ϵ_t est conditionnellement gaussien. La vraisemblance associée à ϵ_t conditionnellement au passé F_{t-1} est donc

$$L(\epsilon_t / F_{t-1}, \underline{\theta}) = \frac{1}{h_t \sqrt{2\pi}} \exp\left(-\frac{(\epsilon_t - m_t(\underline{\theta}))^2}{2h_t^2(\underline{\theta})}\right)$$

et dépend du vecteur $\underline{\theta} = (\alpha_0, \dots, \alpha_p)' \in \mathbb{R}^p$. La fonction de vraisemblance de $(\epsilon_1, \epsilon_2, \dots, \epsilon_t)$ conditionnelle est par conséquent

$$L(\epsilon_1, \epsilon_2, \dots, \epsilon_t, \underline{\theta}) = \prod_{t=1}^T L(\epsilon_t / F_{t-1}, \underline{\theta})$$

L'estimateur est alors défini comme le vecteur $\hat{\underline{\theta}}_T = (\hat{\alpha}_0, \dots, \hat{\alpha}_p)'$ qui maximise le logarithme de cette fonction de vraisemblance :

$$\hat{\underline{\theta}}_T = \arg \max \log L(\epsilon_1, \epsilon_2, \dots, \epsilon_T, \underline{\theta})$$

Maintenant pour un modèle GARCH, la vraisemblance associée à ϵ_t conditionnellement au passé F_{t-1} s'écrit

$$L(\epsilon_t / F_{t-1}, \underline{\theta}) = \frac{1}{h_t \sqrt{2\pi}} \exp\left(-\frac{(\epsilon_t - m_t(\underline{\theta}))^2}{2h_t^2(\underline{\theta})}\right)$$

mais cette fois, la variance h_t^2 suit un processus ARMA et dépend donc des valeurs passées de la variance conditionnelle h_1^2, \dots, h_t^2 . Ces valeurs n'étant pas observées en pratique, la maximisation direct de la vraisemblance est rendue impossible. En pratique, on estime successivement les valeurs de h_1^2, \dots, h_t^2 avant de calculer la vraisemblance. Ainsi, pour un vecteur $\underline{\theta}_0 = (\alpha_0^0, \dots, \alpha_p^0, \beta_1^0, \dots, \beta_q^0)'$ fixées de paramètres, on calcul récursivement

$$\hat{h}_s^2 = \alpha_0^0 + \sum_{i=1}^p \alpha_i^0 \epsilon_{s-i}^2 + \sum_{j=1}^q \beta_j^0 \hat{h}_{s-j}^2$$

avec la convention $\epsilon_i = 0$ et $h_i^2 = 0$ si $i = 0$. On remplace donc la fonction de vraisemblance par

$$L(\epsilon_t / F_{t-1}, \underline{\theta}) = \frac{1}{\hat{h}_t \sqrt{2\pi}} \exp\left(-\frac{(\epsilon_t - m_t(\underline{\theta}))^2}{2\hat{h}_t^2}\right)$$

et la fonction de vraisemblance totale est donnée par

$$L(\epsilon_1, \epsilon_2, \dots, \epsilon_t, \underline{\theta}_0) = \prod_{t=1}^T L(\epsilon_t / F_{t-1}, \underline{\theta}_0)$$

Cette fonction de vraisemblance peut être calculée pour différentes valeurs du vecteur $\underline{\theta}_0$ et sa maximisation donne l'estimateur de maximum de vraisemblance.

théorème 95 *Sous certaines conditions de régularité, l'estimateur du MLE est asymptotiquement convergent et normal*

$$\sqrt{T}(\hat{\underline{\theta}}_T - \underline{\theta}) \rightarrow N(0, J^{-1}IJ^{-1}) \text{ lorsque } T \rightarrow \infty$$

où la matrice de variance covariance asymptotique de l'estimateur MLE est calculée à partir de

$$\begin{cases} J = E_0 \left[-\frac{\partial^2 \log L(\underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}'} \right] \\ I = E_0 \left[\frac{\partial \log L(\underline{\theta})}{\partial \underline{\theta}} \frac{\partial \log L(\underline{\theta})}{\partial \underline{\theta}'} \right] \end{cases}$$

où E_0 désigne l'espérance prise par rapport à la vraie loi.

Naturellement dans la pratique les matrices I et J sont directement estimées en remplaçant l'espérance E_0 par la moyenne empirique et le paramètre inconnu $\underline{\theta}$ par son estimateur convergent $\hat{\underline{\theta}}_T$. Ainsi, on utilise :

$$\begin{cases} \hat{I} = \left(\frac{1}{T} \sum_{i=0}^T \frac{\partial \log L(\underline{\theta})}{\partial \underline{\theta}} \right) \Big|_{\underline{\theta}=\hat{\underline{\theta}}_T} \left(\frac{\partial \log L(\underline{\theta})}{\partial \underline{\theta}'} \right) \Big|_{\underline{\theta}=\hat{\underline{\theta}}_T} \\ \hat{J} = \left(-\frac{1}{T} \sum_{i=1}^T \frac{\partial^2 \log L(\underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}'} \right) \Big|_{\underline{\theta}=\hat{\underline{\theta}}_T} \end{cases}$$

et la variance estimée de $\hat{\underline{\theta}}_T$ vérifie alors

$$\text{Var} \left[\sqrt{T}(\hat{\underline{\theta}}_T - \underline{\theta}) \right] = \hat{J}^{-1} \hat{I} \hat{J}^{-1}$$

6.5 Simulation

On utilise dans cette section le logiciel RStudio 1.1.463 et les packages "caschrono", "fBasics", "xtable", "fGarch" et "FinTS".

Le logiciel R prend comme notation des processus *GARCH*

- ε_t est un processus *GARCH*(1,1)

$$\varepsilon_t = \sigma_t \eta_t$$

où $\eta_t \rightsquigarrow iid$ de loi $N(0,1)$ indépendant de $F_{t-1}, \forall t$ et

$$\sigma^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta^2 \eta_{t-1}^2$$

avec $\omega > 0, \alpha > 0$ et $\beta > 0$.

6.5.1 Simulation d'un GARCH(1,1)

L'estimation d'un *GARCH* se fait par la méthode du maximum de vraisemblance. Cette méthode nécessite de préciser la loi du bruit blanc $\forall t$. Nous avons supposé $\eta_t \rightsquigarrow N(0,1)$ bruit blanc gaussien. Nous illustrons la méthode en estimant le modèle d'une série simulée.

D'abord, nous simulons 1000, 2000, 3000 observations suivant le modèle

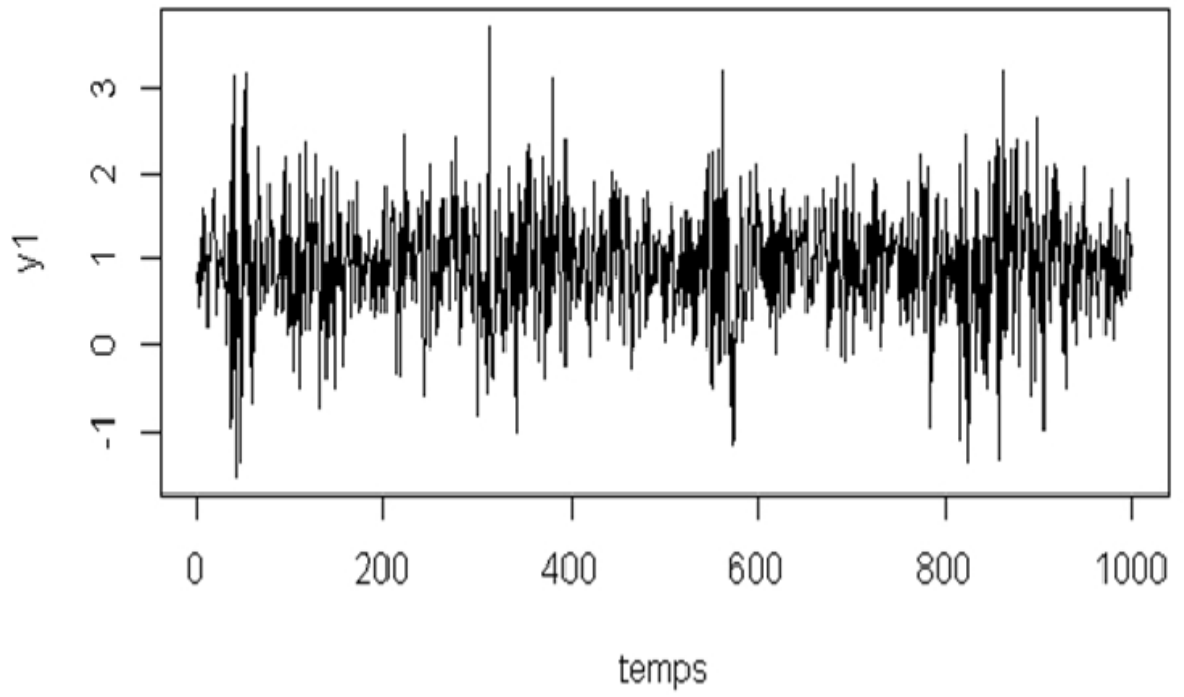
$$\begin{aligned} y_t &= 1 + \eta_t h_t \\ h_t &= \sqrt{0.02 + 0.1 \varepsilon_{t-1}^2 + 0.85 h_{t-1}^2} \end{aligned}$$

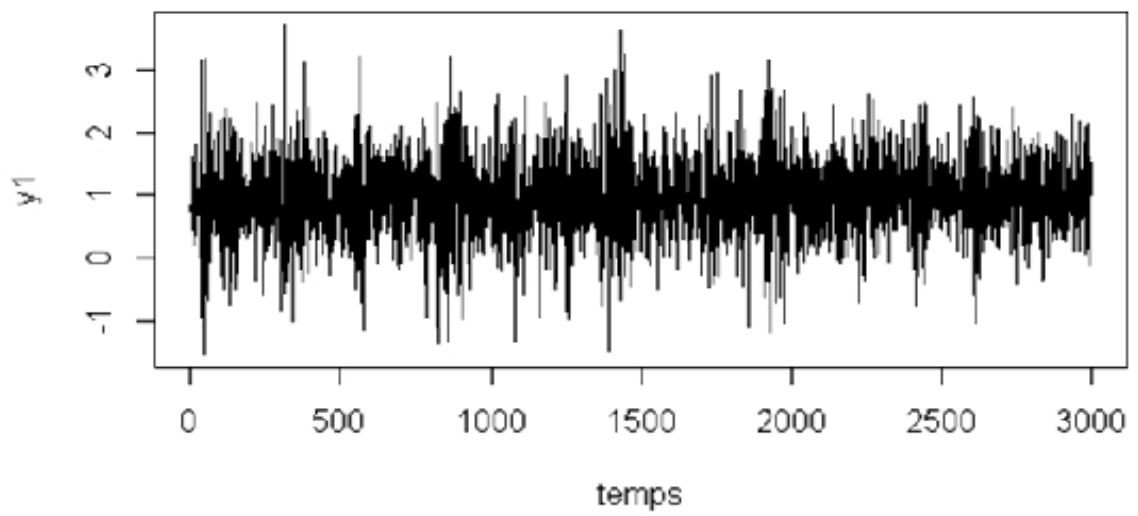
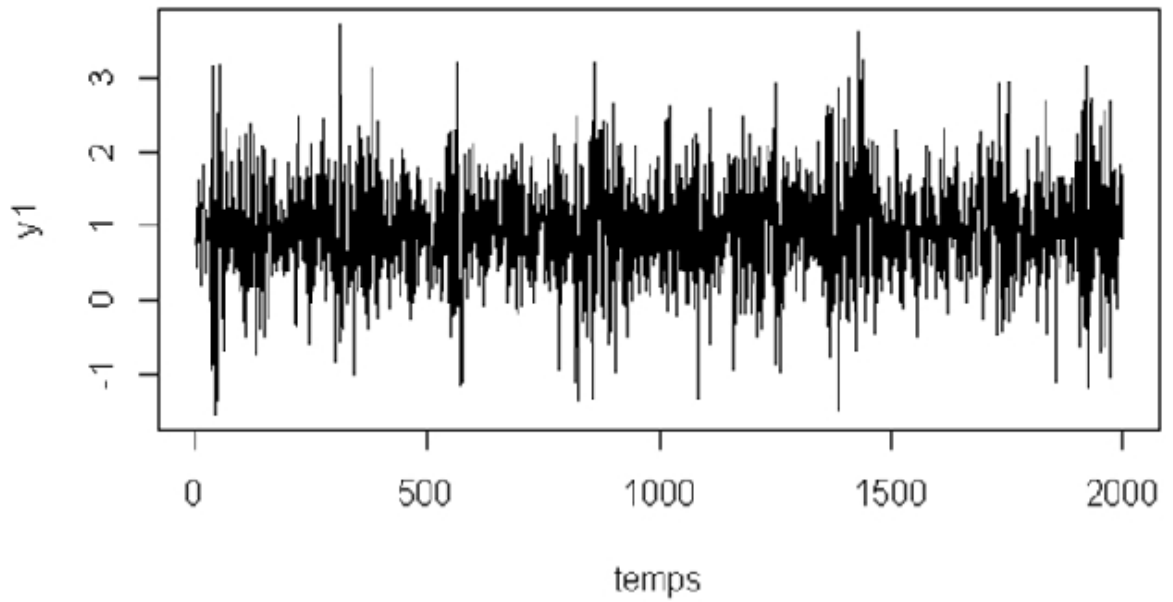
à l'aide de `garchSim()`.

6.5.2 Script R

```
spec=garchSpec(model=list(mu=1,omega=0.02,alph=0.1,beta=0.85),rseed=9647)
var.margi=0.02/(1-0.1-0.85)
y=garchSim(spec,n=1000,extended=TRUE)
y1=y[1:1000,1]
plot.ts(y1,xlab='temps')
```

A chaque fois en remplace 1000 par 2000 et 3000. La série à étudier est formée des 1000, 2000, 3000 observations contenues dans la première colonne de y . Avant d'ajuster un modèle à cette série, examinons son chronogramme (fig : 1.1 pour 1000, fig : 1.2 pour 2000, fig : 1.3 pour 3000).





Il y a plusieurs valeurs extrêmes. Ne pas en tenir compte risquerait de fausser considérablement l'estimation, aussi décidons-nous de les remplacer par des valeurs plus raisonnables. Il faut commencer par repérer ces valeurs extrêmes et les situer par rapport à des quantiles extrêmes de la distribution des écarts à la moyenne, en valeur absolue. Nous retenons 5 ordres quantiles élevés et calculons les quantiles correspondants, q_5 .

6.5.3 Script R

```
> m1=mean(y1)
> (q5=quantile(abs(y1-m1),probs=c(.975,.98,.985,.99,.995))).
> extrem985=which(abs(y1-m1)>q5[3])
> y1b=y1
> y1b[extrem985]=q5[3]*sign(y1[extrem985]-m1)
> plot.ts(y1b,xlab='temps')
```

1. $N = 1000$

97.5%	98%	98.5%	99%	99.5%
1.542952	1.709958	1.912828	2.093440	2.280911

2. $N = 2000$

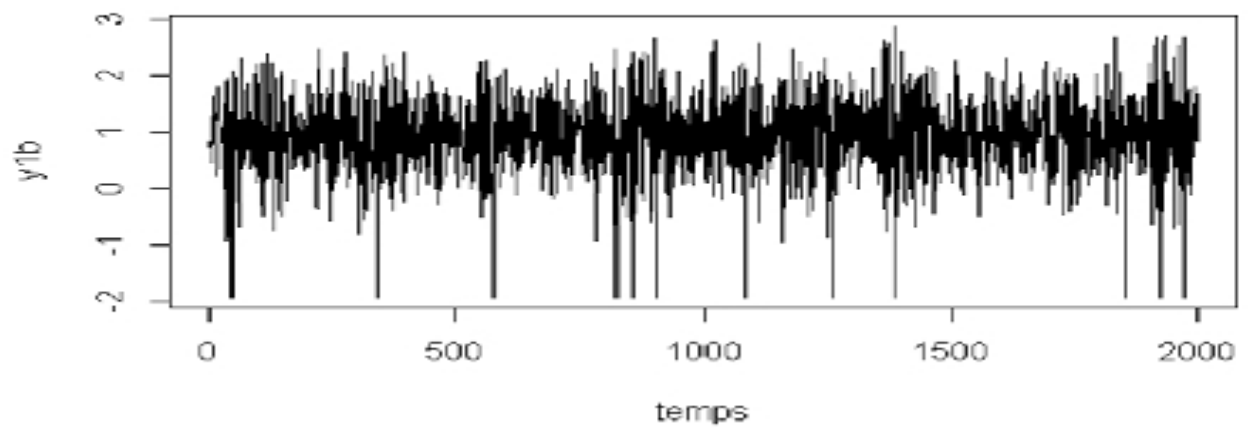
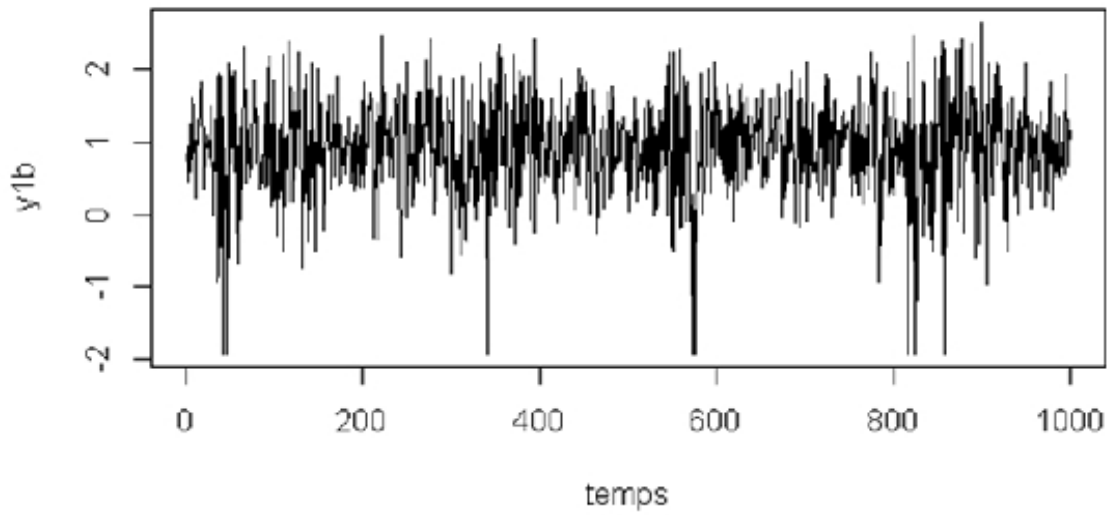
97.5%	98%	98.5%	99%	99.5%
1.632451	1.718765	1.917385	2.043769	2.264712

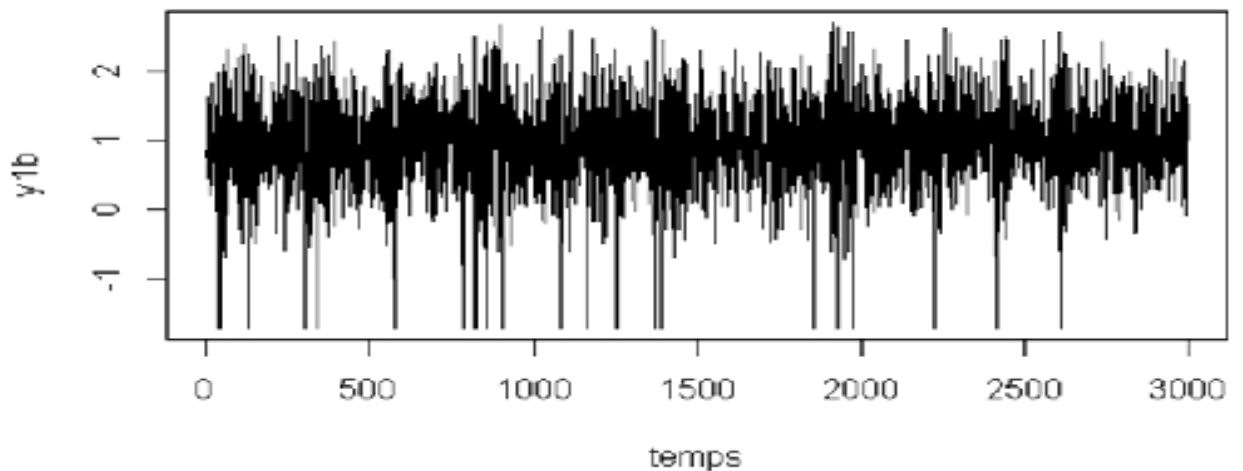
3. $N = 3000$

97.5%	98%	98.5%	99%	99.5%
1.496125	1.598511	1.691911	1.942191	2.135320

D'après l'examen de ces quantiles, nous décidons de ramener les points de valeur absolue supérieure au quantile d'ordre 98.5% à ce percentile, en conservant le signe de la valeur modifiée. Pour ce faire, nous repérons à l'aide de `which()` les indices des points dépassant ce seuil en valeur absolue, puis nous affectons les nouvelles valeurs à ces points. L'effet de cette modification de valeur s'observe en comparant les graphes (fig :2.1 pour 1000, fig :2.2 pour

2000 et fig :2.3 pour 3000) avec les graphes précédents.





Nous pouvons maintenant estimer le modèle à l'aide de `garchFit()` et de `fGarch`

```
> mod1=garchFit(~garch(1,1),data=y1b,trace=FALSE,include.mean=TRUE)
> summary(mod1)
```

1. $N = 1000$

Coefficient(s) :

<i>mu</i>	<i>omega</i>	<i>alpha1</i>	<i>beta1</i>
0.944458	0.021559	0.144686	0.815522

Std. Errors :

based on Hessian

Error Analysis :

	<i>Estimate</i>	<i>Std.Error</i>	<i>tvalue</i>	<i>Pr(> t)</i>
<i>mu</i>	0.944458	0.018565	50.873	$< 2e - 16$ ***
<i>omega</i>	0.021559	0.007649	2.819	0.00482 **
<i>alpha1</i>	0.144686	0.029668	4.877	$1.08e - 06$ ***
<i>beta1</i>	0.815522	0.034335	23.752	$< 2e - 16$ ***

2. $N = 2000$

Coefficient(s) :

<i>mu</i>	<i>omega</i>	<i>alpha1</i>	<i>beta1</i>
0.947481	0.020116	0.105606	0.854914

Std. Errors :

based on Hessian

Error Analysis :

	<i>Estimate</i>	<i>Std.Error</i>	<i>tvalue</i>	<i>Pr(> t)</i>
<i>mu</i>	0.94748	0.01362	69.541	$< 2e - 16$ ***
<i>omega</i>	0.02012	0.00540	3.725	0.000195 ***
<i>alpha1</i>	0.10561	0.01686	6.264	$3.76e - 10$ ***
<i>beta1</i>	0.85491	0.02184	39.144	$< 2e - 16$ ***

3. $N = 3000$

Coefficient(s) :

<i>mu</i>	<i>omega</i>	<i>alpha1</i>	<i>beta1</i>
0.979235	0.021009	0.100700	0.851364

Std. Errors :

based on Hessian

Error Analysis :

	<i>Estimate</i>	<i>Std.Error</i>	<i>tvalue</i>	<i>Pr(> t)</i>
<i>mu</i>	0.979235	0.010694	91.567	$< 2e - 16$ ***
<i>omega</i>	0.021009	0.004463	4.707	$2.51e - 06$ ***
<i>alpha1</i>	0.100700	0.013788	7.303	$2.81e - 13$ ***
<i>beta1</i>	0.851364	0.019588	43.463	$< 2e - 16$ ***

6.5.4 Conclusion

Le but de ce chapitre a été mis en évidence l'utilité des modèles non linéaire et l'hétéroscédasticité conditionnelle qui possède, maintenant, des outils puissants d'analyse et de modélisation fondés sur des bases théoriques solides pour modéliser des séries chronologiques stationnaire présentant une dynamique non linéaire. Le concept de variance conditionnelle a commencé à

jouer un grand rôle au début des années quatre vingt avec l'article fondateur de Engle [3]. Il caractérise les modèles venus élargir la classe des modèles classiques fondés essentiellement sur une structure de dépendance linéaire entre une variable à l'instant t et ses valeurs passées et celles d'un bruit blanc et de ses valeurs passées.

Les modèles *ARCH* permettent de prendre en compte des faits stylisés inhérents à la volatilité comme, par exemple, un excès d'aplatissement, de faibles autocorrélations des rentabilités journalières des actifs considérés, et des autocorrélations positives et significatives pour le carré de ces rentabilités, c'est-à-dire, non stationnarité des variations de volatilité avec l'existence de la volatilité, plusieurs chercheurs essayent d'étudier la raison de la volatilité du marché. La raison de volatilité n'est pas seulement les informations existantes sur le marché mais aussi le comportement des investisseurs, les bulles spéculatives et plusieurs autres facteurs. Avec ses influences, la volatilité fait stabiliser les marchés financiers et fait stabiliser aussi l'économie mondiale.

Les modèles *ARCH* et *GARCH* ont l'avantage de permettre de modéliser avec assez peu de paramètres des séries temporelles complexes et sont pour cette raison très utilisés pour certaines séries financières, en particulier pour prédire la volatilité. Il est cependant de plus en plus admis que la prise en compte de non stationnarités dans les séries financières est inévitable pour mettre en pratique ce type de modèles sur des données réelles de longue durée.

Les modèles *ARCH* posent problème lorsque le nombre de données historiques devient grand auquel cas les variances conditionnelles ont tendance à devenir négatives. En effet, le problème des modèles *ARCH* vient du fait que la volatilité est prédite par les carrés des innovations.

Dans ce chapitre, nous avons écrit en détail la méthode du maximum de vraisemblance. Dans un premier temps, nous avons présenté une procédure itérative de calcul de la log vraisemblance gaussienne, conditionnellement à des valeurs initiales fixes ou aléatoires. Cette vraisemblance est écrite comme si la loi des variables η_t était normale centrée réduite, mais cette hypothèse n'est pas nécessaire pour la convergence forte de l'estimateur. Elle a évidemment un effet sur la variance de la loi normale asymptotique de l'estimateur.

Bibliographie

- [1] Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307-327. 1986.
- [2] Brockwell. P J. and . Davis. R A. *Time Series : Theory and Methods*, (Second ed.). Springer, 1991.
- [3] Engle, R.F. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation. *Econometrica*, 50, 987-1008. 1982.
- [4] Nelson. D.B. Conditional Heteroskedasticity in Asset Returns : a New Approach. *Econometrica*, 59, 347-370. 1991