

# Les modèles MA, AR et ARMA multidimensionnels : estimation et causalité

Steven Fortier

RÉSUMÉ Le monde tel qu'on le connaît est rempli de phénomènes dépendant du temps. En statistique, l'une des méthodes fréquemment utilisée est celle des séries chronologiques. À l'aide d'un modèle, il est possible de faire des prédictions sur des événements futurs. Dans cet article, le concept de série chronologique est introduit et l'on s'intéresse à certains modèles importants. Le concept de causalité est abordé ainsi que certains résultats théoriques. On termine le tout avec quelques exemples simples.

## 1 Introduction

Une série chronologique est une suite formée d'observations au cours du temps. L'analyse des séries chronologiques est un outil couramment utilisé de nos jours pour la prédiction de données futures. Ce domaine possède beaucoup d'applications en finance, en médecine, en économétrie et en météorologie et dans bien d'autres domaines. Par exemple, en finance, on s'intéresse à modéliser le taux de change d'une devise. En météorologie, les scientifiques modélisent par exemple la température au Québec dans le dernier mois pour prédire la température qu'il fera demain. L'idée est de prendre un échantillon de données et de construire le meilleur modèle qui ajuste ces données. Ce modèle nous permet de tirer certaines conclusions sur la série.

Par exemple, on peut établir une formule pour la *prédiction de données*, *déceler certains pics* ou *modéliser la tendance (orientation)* de la série. Un autre aspect important de la série est la *composante saisonnière*, c'est-à-dire la présence de cycles. Un autre concept intéressant serait le *phénomène de causalité*, c'est-à-dire l'influence d'une série sur une autre. Cependant, il n'est pas toujours évident de choisir le bon modèle, car, normalement, plusieurs peuvent être de bons candidats. La plupart de ceux-ci auront tendance à minimiser la variance des résidus. Il faut aussi considérer le nombre de paramètres à estimer. Normalement, on utilise le principe de parcimonie qui nous propose de choisir le modèle demandant le moins de paramètres avec une variance faible des résidus. Il y a donc plusieurs critères à considérer dans la création d'un modèle.

---

J'aimerais remercier mon directeur Taoufik Bouezmarni. Je remercie également le professeur Bernard Colin et mes collègues de travail qui m'ont aidé dans la rédaction de cet article.

L'article est structuré de la façon suivante. Dans la deuxième section, on donne les définitions nécessaires à la compréhension du corps de l'article. Ensuite, la troisième section présente trois principaux modèles : AR, MA et ARMA. On définit le concept de causalité entre deux séries et on énonce certains résultats importants. Les sections quatre et cinq se consacrent à l'estimation et à certains exemples.

## 2 Définitions

Dans cette section, nous donnons quelques définitions et notations utiles pour la suite de l'article.

### 2.1 Série chronologique

**Définition 2.1.** Une série chronologique provient de la réalisation d'une famille de variables aléatoires  $\{X_t, t \in I\}$ , où l'ensemble  $I$  est un intervalle de temps qui peut être discret ou continu. Pour cet article, nous utilisons l'ensemble  $I = \{0, 1, \dots, T\}$ , où  $T$  est le nombre total d'observations.

### 2.2 Bruit blanc

**Définition 2.2.** On dit de la suite de variables aléatoires  $\{\varepsilon_t\}$  qu'elle constitue un *bruit blanc faible* si elle possède les propriétés suivantes :

$$\begin{aligned} E[\varepsilon_t] &= 0 && \text{pour tout } t \in Z \\ E[\varepsilon_t^2] &= \sigma_\varepsilon^2 \neq 0 \text{ et constante} \\ \text{Cov}(\varepsilon_s, \varepsilon_t) &= 0 \text{ si } t \neq s. \end{aligned} \tag{2.1}$$

En d'autres termes, les variables aléatoires  $\varepsilon_t$  sont de moyenne nulle, de variance constante et non corrélées. On dit que  $\{\varepsilon_t\}$  est un *bruit blanc fort* s'il est un bruit blanc faible et que les variables aléatoires  $\varepsilon_t$  sont i.i.d. (indépendantes et identiquement distribuées).

*Notation.*

1. Si  $\{\varepsilon_t\}$  bruit blanc faible, alors  $\{\varepsilon_t\} \sim \text{WN}(0, \sigma_\varepsilon^2)$ ;
2. Si  $\{\varepsilon_t\}$  bruit blanc fort, alors  $\{\varepsilon_t\} \sim \text{I. I. D}(0, \sigma_\varepsilon^2)$ .

### 2.3 Stationnarité

Une propriété importante des séries chronologiques est la stationnarité. Cette propriété est nécessaire pour appliquer certains théorèmes sur la causalité. La définition suivante présente le type de stationnarité le plus utilisé.

**Définition 2.3.** Une suite  $\{X_t : t \geq 0\}$  de variables aléatoires est dite *stationnaire du second ordre* si elle vérifie les propriétés suivantes :

$$\begin{aligned} E[X_t] &= \mu < \infty \\ E[X_t^2] &< \infty \\ \text{Cov}(X_s, X_{s+t}) &= \text{Cov}(X_{s-1}, X_{s-1+t}) = \dots = \text{Cov}(X_0, X_t) \quad \text{pour tout } s, t \in \mathbb{N}. \end{aligned}$$

*Remarque 2.4.* On appelle  $\text{Cov}(X_t, X_{t+h})$  l'*autocovariance* (ACV) à l'horizon  $h$  et on la note :

$$r^X(h) = \text{Cov}(X_t, X_{t+h}). \quad (2.2)$$

*Remarque 2.5.* On peut prendre cette fonction et la diviser par la variance de  $\{X_t\}$  pour obtenir une nouvelle fonction que l'on appellera *autocorrélation* (ACF). On la note :

$$\rho^X(h) = \frac{r^X(h)}{r^X(0)} \quad (2.3)$$

avec les propriétés suivantes :

1.  $-1 \leq \rho^X(h) \leq 1$ ;
2.  $\rho^X(h) = 0$  signifie que les observations  $X_t$  et  $X_{t+h}$  sont non corrélées ;
3.  $\rho^X(h) = \pm 1$  signifie que la corrélation est parfaite (négative ou positive).

## 2.4 Opérateurs définis sur une série chronologique

### 2.4.1 Opérateur de retard

**Définition 2.6.** L'*opérateur de retard*  $B$  se définit de la manière suivante :

$$B(X_t) = X_{t-1}.$$

*Remarque 2.7.*  $B^n(X_t) = X_{t-n}$  pour tout  $n \in \mathbb{N}$ .

### 2.4.2 Opérateur de différence d'ordre $d$

**Définition 2.8.** On définit l'opérateur  $\Delta_d$  de différence d'ordre  $d$  comme l'opérateur linéaire tel que

$$\Delta_d X_t = X_t - X_{t-d} = (1 - B^d)X_t.$$

On peut aussi prendre l'opérateur d'ordre un et l'appliquer plusieurs fois :

$$\Delta^2(X_t) = \Delta(\Delta(X_t)) = \Delta(X_t - X_{t-1}) = (1 - B)(X_t - X_{t-1}) = X_t - 2X_{t-1} + X_{t-2}.$$

Ces opérateurs peuvent être utilisés afin de transformer un processus de moyenne non nulle en un processus de moyenne nulle. On peut aussi s'en servir pour enlever la composante saisonnière de la série. Dans la prochaine section, on les utilise pour mieux représenter les modèles.

## 3 Types de modèle

### 3.1 Cas unidimensionnel

Dans cette section, on considère certains modèles fréquemment utilisés pour une série chronologique  $\{X_t\}$ . Nous commençons par deux modèles simples : le modèle autorégressif et le modèle à moyenne mobile. On continue avec le modèle ARMA qui combine ces deux modèles.

### 3.1.1 Processus autorégressif AR( $p$ )

Les premiers processus autorégressifs ont été introduits par George Udny Yule dans [Yul27]. Dans cet article, Yule utilise le premier modèle autorégressif pour modéliser la série chronologique du nombre de taches solaires plutôt que la méthode du périodogramme de Schuster [Sch06]. Un processus autorégressif est un processus où l'on écrit une observation au temps  $t$  comme une combinaison linéaire des observations passées plus un certain bruit blanc.

**Définition 3.1.** La suite  $\{X_t : t \geq 0\}$  est un processus autorégressif d'ordre  $p$  ( $p > 0$ ) s'il peut s'écrire sous la forme suivante :

$$X_t = \sum_{k=1}^p \phi_k X_{t-k} + \varepsilon_t, \quad \text{où } \{\varepsilon_t\} \sim \text{WN}(0, \sigma_\varepsilon^2). \quad (3.1)$$

Les  $\phi_k$  ( $k = 1, 2, \dots, p$ ) constituent les paramètres du modèle,

Dans ce cas, on note  $\{X_t\} \sim \text{AR}(p)$ . De la même façon, on peut réécrire un processus AR( $p$ ) avec un polynôme  $\phi(B)$  qui multipliera  $X_t$  cette fois-ci :

$$\phi(B)X_t = \varepsilon_t \text{ avec } \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p. \quad (3.2)$$

**Exemple 3.2.** Un processus AR(1) prend la forme suivante :

$$X_t = \phi X_{t-1} + \varepsilon_t, \quad \text{où } \{\varepsilon_t\} \sim \text{WN}(0, \sigma_\varepsilon^2).$$

### 3.1.2 Processus à moyenne mobile MA( $q$ )

C'est Eugen Slutsky qui, en 1927, dans son article [Slu27], a introduit pour la première fois les processus à moyenne mobile. La définition suivante présente ce processus.

**Définition 3.3.** On dit que la suite  $\{X_t : t \geq 0\}$  est un processus à moyenne mobile d'ordre  $q$  ( $q > 0$ ) si celui-ci peut s'écrire sous la forme suivante :

$$X_t = \sum_{k=1}^q \theta_k \varepsilon_{t-k} + \varepsilon_t, \quad \text{où } \{\varepsilon_t\} \sim \text{WN}(0, \sigma_\varepsilon^2), \quad (3.3)$$

où les  $\theta_k$  ( $k = 1, 2, \dots, q$ ) sont les paramètres du modèle.

Dans ce cas, on note  $\{X_t\} \sim \text{MA}(q)$ .

**Exemple 3.4.** Un processus MA(1) prend la forme suivante :

$$X_t = \theta \varepsilon_{t-1} + \varepsilon_t, \quad \text{où } \{\varepsilon_t\} \sim \text{WN}(0, \sigma_\varepsilon^2). \quad (3.4)$$

On peut utiliser l'opérateur de retard  $B$  pour écrire ce processus sous une autre forme. On aura donc un polynôme en  $B$  qui multipliera  $\varepsilon_t$  :

$$X_t = \theta(B)\varepsilon_t \text{ avec } \theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q. \quad (3.5)$$

*Remarque 3.5.* Si  $\{X_t\} \sim \text{MA}(q)$ , alors :

1.  $X_t$  est stationnaire ;
2.  $\rho^X(h) = 0$  pour tout  $h > q$ .

**Définition 3.6.** Un processus est dit *causal* s'il existe une suite  $\{a_k\}$  réelle telle que  $\sum_{k=0}^{\infty} |a_k| < \infty$  et que

$$X_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}. \quad (3.6)$$

Parfois, lorsque l'on parle d'un processus causal, on dit que celui-ci a une représentation  $\text{MA}(\infty)$ .

*Remarque 3.7.* Tout processus  $\text{MA}(q)$  est causal.

**Définition 3.8.** Un processus est dit *inversible* s'il existe une suite  $\{b_k\}$  réelle telle que  $\sum_{k=0}^{\infty} |b_k| < \infty$  et

$$\varepsilon_t = \sum_{k=0}^{\infty} b_k X_{t-k}. \quad (3.7)$$

Une autre façon de dire qu'un processus est inversible est de dire qu'il possède une représentation  $\text{AR}(\infty)$ .

*Remarque 3.9.* Avec cette définition, tout processus  $\text{AR}(p)$  est inversible.

**Théorème 3.10.** *Un processus autoregressif  $\text{AR}(p)$  est causal et stationnaire si et seulement si son polynôme  $\phi(z)$  est tel que*

$$\phi(z) \neq 0 \text{ avec } z \in \mathbb{C} \text{ tel que } |z| \leq 1. \quad (3.8)$$

*En d'autres mots, toutes les racines de  $\phi(z)$  sont de norme plus grande que 1.*

Ce théorème se retrouve dans le livre de Brockwell et Davis [BD09] (théorème 3.1.1). La démonstration y est aussi présente.

**Exemple 3.11.** Le processus  $\text{AR}(2)$  suivant est stationnaire et causal :

$$X_t = -\frac{1}{6}X_{t-1} + \frac{1}{6}X_{t-2} + \varepsilon_t, \quad \text{où } \{\varepsilon_t\} \sim \text{WN}(0, \sigma_\varepsilon^2). \quad (3.9)$$

En effet, on peut le réécrire comme :

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

Ainsi, on a que

$$\phi(z) = -\frac{1}{6}z^2 + \frac{1}{6}z + 1 = \frac{1}{6}(z^2 - z - 6) = \frac{1}{6}(z+2)(z-3).$$

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

**Théorème 3.12.** *Un processus à moyenne mobile MA( $q$ ) est inversible si et seulement si son polynôme  $\theta(z)$  est tel que*

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

*On note la ressemblance de cet énoncé avec le théorème de stationnarité et de causalité pour les processus autorégressifs.*

La démonstration se trouve dans le livre de Brockwell et Davis [BD09] (Théorème 3.1.2).

### 3.1.3 Modèle ARMA( $p, q$ )

Les modèles AR et MA peuvent être parfaits dans certain cas, mais il peut arriver que l'on ait besoin d'estimer un grand nombre de paramètres pour ajuster le modèle. Si l'on dispose de peu d'observations, ces estimations auront tendance à être peu précises. De plus, si un modèle contenant  $p$  paramètres convient pour la situation, il n'est pas bon de tenter d'ajuster un modèle qui en contiendra plus que  $p$ . Les modèles ARMA consistent à avoir une partie autorégressive et une partie à moyenne mobile. Herman Wold [Wol38] a montré que les processus ARMA pouvaient être utilisés pour modéliser n'importe quelle série stationnaire pour autant que les ordres  $p$  et  $q$  soient bien choisis. Box et Jenkins [BJ70] ont travaillé pour développer une méthodologie pour l'estimation du modèle d'une série chronologique. Il peut être facile de vérifier la causalité et l'inversibilité du modèle ARMA à l'aide des théorèmes 3.10 et 3.12.

**Définition 3.13.** Un processus est dit ARMA( $p, q$ ) s'il existe des suites réelles  $\{\phi_k\}$  et  $\{\theta_k\}$  telles que

$$X_t - \sum_{k=1}^p \phi_k X_{t-k} = \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}, \quad \text{avec } \{\varepsilon_t\} \sim \text{WN}(0, \sigma_\varepsilon^2).$$

On peut aussi utiliser les polynômes  $\phi(B)$  et  $\theta(B)$  pour réécrire ce modèle sous la forme :

$$\phi(B)X_t = \theta(B)\varepsilon_t$$

avec

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \text{ et } \theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q.$$

On note  $\{X_t\} \sim \text{ARMA}(p, q)$ .

*Remarque 3.14.* On note certaines propriétés pour les modèles ARMA( $p, q$ ) :

1. Si  $p = q = 0$ , on a  $X_t = \varepsilon_t \sim \text{WN}(0, \sigma_\varepsilon^2)$  ;
2. Si  $p = 0$  et  $q \neq 0$ , on a  $\{X_t\} \sim \text{MA}(q)$  ;
3. Si  $q = 0$  et  $p \neq 0$ , on a  $\{X_t\} \sim \text{AR}(p)$ .

### 3.2 Cas bidimensionnel et critère de causalité entre deux séries

Dans cette section, on considère deux séries  $\{X_{1t}\}$  et  $\{X_{2t}\}$  et on reprend les notions de processus autorégressif et à moyenne mobile de la deuxième section. Le but est donc d'avoir un modèle pour représenter la structure du vecteur  $U_t = (X_{1t}, X_{2t})'$ . Par la suite, on donnera des conditions nécessaires et suffisantes pour tester la causalité entre  $\{X_{1t}\}$  et  $\{X_{2t}\}$ . Tout d'abord, il faut bien comprendre ce que l'on veut dire lorsqu'on parle de causalité entre deux séries.

#### 3.2.1 Définitions de causalité au sens de Granger

La définition de causalité au sens de Granger a été introduite pour la première fois dans un article de Clive Granger [Gra69].

**Définition 3.15.** Soient :

1.  $X_1 = \{X_{1t}\}$  et  $X_2 = \{X_{2t}\}$  deux séries stationnaires et inversibles ;
2.  $\bar{X}_{1t} = \{X_{1s} \mid s < t\}$  et  $\bar{X}_{2t} = \{X_{2s} \mid s < t\}$  (Historique de  $X_{it}$  pour  $i = 1, 2$ ) ;
3.  $\{\varepsilon_t\}$  un bruit blanc faible à deux dimensions avec matrice de covariance  $\Sigma$  ;
4.  $A_t =$  Ensemble dénombrable de variables aléatoires de variances finies et comprenant  $X_{1t}$  et  $X_{2t}$  (Ensemble d'information) ;
5.  $\bar{A}_t = \bigcup_{s < t} A_s$   
 $=$  Ensemble de variables aléatoires de variances finies comprenant  $\bar{X}_{1t}$  et  $\bar{X}_{2t}$  ;
6.  $\bar{A}_t \setminus \bar{X}_{1t}$  constitue l'information disponible au temps  $t - 1$  sans l'historique de  $X_{1t}$  ;
7.  $A = \bigcup_{t \in \mathbb{Z}} A_t$  ;
8.  $B$  est un sous-ensemble de  $A$  ;
9.  $\hat{X}_{it} = E(X_{it}|B)$  est le meilleur prédicteur linéaire sans biais de  $X_{it}$  sachant l'information contenu dans  $B$  ;
10.  $\varepsilon(X_{it}|B) = X_{it} - \hat{X}_{it}$  est l'écart entre la véritable valeur et la prédiction ;
11.  $\sigma_\varepsilon^2(X_{it}|B) = E[(\varepsilon(X_{it}|B))^2]$ .

**Définition 3.16.** On dit que  $X_1$  cause  $X_2$  si et seulement si

$$\sigma_\varepsilon^2(X_{2t}|\bar{A}_t) < \sigma_\varepsilon^2(X_{2t}|\bar{A}_t \setminus \bar{X}_{1t}) \text{ pour au moins une valeur de } t.$$

On note alors  $X_1 \rightarrow X_2$ .

On dit qu'il y a *rétroaction*<sup>1</sup> entre  $X_1$  et  $X_2$  si et seulement si  $X_1$  cause  $X_2$  et  $X_2$  cause  $X_1$ .

**Définition 3.17.** On dit que  $X_1$  ne cause pas  $X_2$  si et seulement si

$$\sigma_\varepsilon^2(X_{2t}|\bar{A}_t) = \sigma_\varepsilon^2(X_{2t}|\bar{A}_t \setminus \bar{X}_{1t}) \quad \text{pour tout } t \quad (\text{Causalité en variance})$$

ou, de façon équivalente,

$$E(X_{2t}|\bar{A}_t) = E(X_{2t}|\bar{A}_t \setminus \bar{X}_{1t}) \text{ presque sûrement pour tout } t. \quad (\text{Causalité en moyenne})$$

En d'autres termes,  $X_1$  ne cause pas  $X_2$  si et seulement si la prédiction de  $X_{2t}$  sachant l'historique de  $X_{1t}$  et  $X_{2t}$  reste la même que si on connaît seulement l'historique de  $X_{2t}$ . Pour les démonstrations qui suivent, on utilisera cette définition de causalité utilisant ces prédictions. Dans les sections qui suivent,  $\varepsilon_t$  est un bruit blanc faible à deux dimensions tel que :

1.  $E[\varepsilon_t] = 0$ ;
2.  $E[\varepsilon'_t \varepsilon_s] = \begin{cases} I_2 & \text{si } t = s \\ 0 & \text{si } t \neq s \end{cases}$ ,

où  $I_2$  est la matrice identité d'ordre deux et  $\varepsilon'_t$  est le vecteur  $\varepsilon_t$  transposé.

On regarde maintenant une forme bidimensionnelle des modèles vus précédemment.

### 3.2.2 Processus autorégressif bidimensionnel

**Définition 3.18.** On dit que  $(X_1, X_2)'$  est un processus autorégressif s'il peut s'écrire sous la forme

$$\sum_{j=0}^p \phi_j \begin{pmatrix} X_{1,t-j} \\ X_{2,t-j} \end{pmatrix} = \begin{pmatrix} \phi_{11}(B) & \phi_{12}(B) \\ \phi_{21}(B) & \phi_{22}(B) \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

où  $\phi_i$  sont des matrices de format 2x2 telles que  $\phi_0 = I_2$  et où  $\phi_{11}, \phi_{12}, \phi_{21}$  et  $\phi_{22}$  sont des polynômes en  $B$  définissant les coefficients autorégressifs.

Le théorème suivant nous donne les conditions nécessaires et suffisantes pour que  $X_1$  ne cause pas  $X_2$ .

---

<sup>1</sup>De l'anglais, feedback.



**Théorème 3.19.** *Soit  $(X_1, X_2)'$ , un processus autorégressif, alors  $X_1 \rightarrow X_2$  si et seulement si  $(X_1, X_2)'$  est tel que*

$$\begin{pmatrix} \phi_{11}(B) & \phi_{12}(B) \\ 0 & \phi_{22}(B) \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}.$$

*Démonstration.* On reprend ici une démonstration de [BDR92] pour le cas bidimensionnel. Premièrement, établissons certaines équations qui nous serviront pour la démonstration. On ne démontre que la condition suffisante dans cet article, mais le reste de la démonstration peut se trouver dans [BDR92]. Soit le processus  $(X_1, X_2)'$  autorégressif d'ordre deux défini par :

$$\phi(B) \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \phi_{11}(B) & \phi_{12}(B) \\ \phi_{21}(B) & \phi_{22}(B) \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

où  $\phi$  est une matrice de format  $2 \times 2$ .

La définition du processus bidimensionnel introduite dans 3.18 exige que la matrice  $\phi_0$  soit la matrice identité. Ceci nous garantit que les coefficients de  $X_{1t}$  et  $X_{2t}$  seront toujours un. On peut donc écrire les polynômes  $\phi_{21}(B)$  et  $\phi_{22}(B)$  sous les formes suivantes :

$$\phi_{21}(B) = - \sum_{k=1}^{\infty} \phi_{21(k)} B^k \quad \text{et} \quad \phi_{22}(B) = 1 - \sum_{k=1}^{\infty} \phi_{22(k)} B^k.$$

Si l'on multiplie la deuxième ligne de la matrice  $\phi$  avec la colonne  $\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix}$ , on peut isoler  $X_{2t}$  et l'écrire sous la forme

$$X_{2t} = \sum_{k=1}^{\infty} \phi_{21(k)} X_{1,t-k} + \sum_{k=1}^{\infty} \phi_{22(k)} X_{2,t-k} + \varepsilon_{2t}.$$

En posant  $Z_{it} = \sum_{k=1}^{\infty} \phi_{2i(k)} X_{i,t-k}$ , on peut écrire

$$X_{2t} = Z_{1t} + Z_{2t} + \varepsilon_{2t}.$$

On peut donc dire que

$$\begin{aligned} E(X_{2t} | \bar{X}_{1t}, \bar{X}_{2t}) &= E(Z_{1t} + Z_{2t} + \varepsilon_{2t} | \bar{X}_{1t}, \bar{X}_{2t}) \\ &= E(Z_{1t} | \bar{X}_{1t}, \bar{X}_{2t}) + E(Z_{2t} | \bar{X}_{1t}, \bar{X}_{2t}) + E(\varepsilon_{2t} | \bar{X}_{1t}, \bar{X}_{2t}) \\ &= Z_{1t} + Z_{2t}. \end{aligned}$$

La dernière égalité vient du fait que comme  $Z_{it}$  est composé des valeurs de l'historique de  $X_{it}$ , la prédiction de  $Z_{it}$  restera  $Z_{it}$ . Ensuite, la prédiction de  $\varepsilon_{2t}$  est 0, car celui-ci est non corrélé avec  $\bar{X}_{2t}$ . De même, on peut dire que

$$\begin{aligned} E(X_{2t} | \bar{X}_{2t}) &= E(Z_{1t} + Z_{2t} + \varepsilon_{2t} | \bar{X}_{2t}) \\ &= E(Z_{1t} | \bar{X}_{2t}) + E(Z_{2t} | \bar{X}_{2t}) + E(\varepsilon_{2t} | \bar{X}_{2t}) \\ &= E(Z_{1t} | \bar{X}_{2t}) + Z_{2t}. \end{aligned}$$

Si l'on suppose que  $\phi_{21} = 0$ , il s'ensuit que  $Z_{1t} = 0$ , ce qui implique que  $E(Z_{1t}|\bar{X}_{2t}) = 0$ . Alors, on a que  $E(X_{2t}|\bar{X}_{1t}, \bar{X}_{2t}) = E(X_{2t}|\bar{X}_{2t})$  et par conséquent,  $X_1 \nrightarrow X_2$ .  $\square$

**Exemple 3.20.** Soit le processus  $(X_1, X_2)$  ayant pour modèle :

$$\begin{cases} X_{1t} &= -0,5X_{1,t-1} + X_{1,t-2} + 2X_{2,t-1} + \varepsilon_{1t} \\ X_{2t} &= 0,2X_{2,t-1} + 0,5X_{2,t-2} + \varepsilon_{2t}. \end{cases}$$

On peut réécrire le modèle comme suit :

$$\begin{pmatrix} 1 + 0,5B - B^2 & -2B \\ 0 & 1 - 0,2B - 0,5B^2 \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}.$$

Nous avons que  $\phi_{21}(B) = 0$  donc, en vertu du théorème précédent,  $X_1 \nrightarrow X_2$ .

### 3.2.3 Processus à moyenne mobile bidimensionnel

On dit que  $(X_1, X_2)'$  possède une représentation de processus à moyenne mobile si l'on peut écrire :

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \sum_{j=0}^q \theta_j \begin{pmatrix} \varepsilon_{1,t-j} \\ \varepsilon_{2,t-j} \end{pmatrix} = \begin{pmatrix} \theta_{11}(B) & \theta_{12}(B) \\ \theta_{21}(B) & \theta_{22}(B) \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

où, les  $\theta_j$  sont des matrices 2x2 telles que  $\theta_0 = I_2$  et où  $\theta_{11}, \theta_{12}, \theta_{21}$  et  $\theta_{22}$  sont des polynômes en  $B$  définissant les coefficients du processus à moyenne mobile. Le théorème suivant nous donne les conditions nécessaires et suffisantes pour que  $(X_1, X_2)'$  soit inversible.

**Théorème 3.21.** *Si  $(X_1, X_2)'$  est un processus à moyenne mobile bidimensionnel tel que :*

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \theta_{11}(B) & \theta_{12}(B) \\ \theta_{21}(B) & \theta_{22}(B) \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

*alors  $(X_1, X_2)'$  est inversible si et seulement si*

$$\begin{vmatrix} \theta_{11}(z) & \theta_{12}(z) \\ \theta_{21}(z) & \theta_{22}(z) \end{vmatrix} \neq 0 \quad \text{pour tout } z \in \mathbb{C} \text{ tel que } |z| \leq 1.$$

Le théorème qui suit nous donne des conditions nécessaires et suffisantes pour vérifier la causalité entre deux séries possédant une structure MA bidimensionnelle.

**Théorème 3.22.** *Soit  $X = (X_1, X_2)'$  un processus à moyenne mobile bidimensionnel inversible, alors  $X_1 \nrightarrow X_2$  si et seulement si  $(X_1, X_2)'$  est tel que*

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \theta_{11}(B) & \theta_{12}(B) \\ 0 & \theta_{22}(B) \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}.$$

*Démonstration.* Soit le processus à moyenne mobile donné par

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \theta_{11}(B) & \theta_{12}(B) \\ \theta_{21}(B) & \theta_{22}(B) \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}.$$

Comme le processus est inversible, la matrice  $\theta = \begin{pmatrix} \theta_{11}(B) & \theta_{12}(B) \\ \theta_{21}(B) & \theta_{22}(B) \end{pmatrix}$  est inversible et on a

$$\theta^{-1} = \frac{1}{\det(\theta)} \begin{pmatrix} \theta_{22}(B) & -\theta_{12}(B) \\ -\theta_{21}(B) & \theta_{11}(B) \end{pmatrix}.$$

On peut donc multiplier par l'inverse des deux côtés de l'équation. On obtient donc :

$$\frac{1}{\det(\theta)} \begin{pmatrix} \theta_{22}(B) & -\theta_{12}(B) \\ -\theta_{21}(B) & \theta_{11}(B) \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

qui est un processus autorégressif. On peut donc appliquer le théorème précédent et on montre que  $X_1 \nrightarrow X_2$  si et seulement si  $\theta_{21} = 0$ .  $\square$

**Exemple 3.23.** Soit le processus  $(X_1, X_2)$  ayant pour modèle :

$$\begin{cases} X_{1t} = \varepsilon_{1,t} - 0,25\varepsilon_{1,t-2} + \varepsilon_{2,t-1} - 0,25\varepsilon_{2,t-2} \\ X_{2t} = \varepsilon_{2,t} - 0,2\varepsilon_{2,t-2} \end{cases}.$$

On peut réécrire le modèle comme suit :

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} 1 - 0,25B^2 & B - 0,25B^2 \\ 0 & 1 - 0,2B^2 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}.$$

Comme  $\theta_{21}(B) = 0$ , en vertu du théorème précédent,  $X_1 \nrightarrow X_2$ .

Considérons maintenant un processus ARMA bidimensionnel et donnons la forme de ce modèle.

### 3.2.4 Processus ARMA bidimensionnel

Pour la suite du document, nous écrirons  $\theta_{ij} = \theta_{ij}(B)$  et  $\phi_{ij} = \phi_{ij}(B)$ . Le processus  $(X_1, X_2)'$  possède une représentation ARMA s'il peut s'écrire sous la forme :

$$\begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

où  $\phi_{ij}$  et  $\theta_{ij}$  sont des polynômes en  $B$  pour  $i, j = 1, 2$ .

**Théorème 3.24.** Soit  $(X_1, X_2)'$  suivant un modèle ARMA inversible donné par

$$\begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}.$$

a) *Condition suffisante* : Si  $\phi_{21} = 0$  et  $\theta_{12} = \theta_{21} = 0$ , alors  $X_1 \nrightarrow X_2$ .

Cette condition correspond à la représentation suivante :

$$\begin{pmatrix} \phi_{11} & \phi_{12} \\ 0 & \phi_{22} \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \theta_{11} & 0 \\ 0 & \theta_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}.$$

b) *Condition nécessaire et suffisante* :  $X_1 \nrightarrow X_2$  si et seulement si

$$\phi_{11}\theta_{21} - \phi_{21}\theta_{11} = 0.$$

Pour démontrer ces résultats, on utilise le fait que le processus est inversible et on multiplie par  $\theta^{-1}$  comme dans l'autre démonstration. On utilise ensuite la condition de causalité du processus autorégressif. Le résultat suit. Théoriquement, ces conditions de causalité sont plutôt faciles à vérifier lorsque l'on possède déjà un modèle. Cependant, en pratique, nous possédons un échantillon du vecteur  $X_t$ . Nous devons d'abord déterminer quel type de modèle utiliser et l'ordre de celui-ci.

## 4 Estimation du modèle

### 4.1 Autocorrélation d'un processus MA( $q$ )

Au début de cet article, nous avons défini la fonction d'autocorrélation d'un processus à l'aide de l'égalité

$$\rho^X(h) = \frac{r^X(h)}{r^X(0)},$$

où  $r^X(h)$  était la fonction d'autocovariance définie par

$$r^X(h) = \text{Cov}(X_t, X_{t+h}).$$

Rappelons que si  $X_t \sim \text{MA}(q)$ , alors :

$$\rho^X(h) = 0 \quad \text{pour tout } h > q. \quad (4.1)$$

Cependant, on ne connaît généralement pas  $r^X(h)$ . On devra donc l'estimer à partir de l'autocovariance empirique notée  $\hat{r}^X(h)$ . Soit  $T$ , le nombre d'observations de notre échantillon et définissons par  $\bar{X}_T$ , la moyenne empirique. L'autocovariance empirique est alors définie par

$$\hat{r}^X(h) = \begin{cases} \frac{1}{T-h} \sum_{t=1}^{T-h} (X_t - \bar{X}_T)(X_{t+h} - \bar{X}_T) & \text{si } 0 \leq h \leq T \\ 0 & \text{sinon.} \end{cases} \quad (4.2)$$

On aura aussi l'autocorrélation empirique définie par :

$$\hat{\rho}^X(h) = \hat{r}^X(h) / \hat{r}^X(0). \quad (4.3)$$

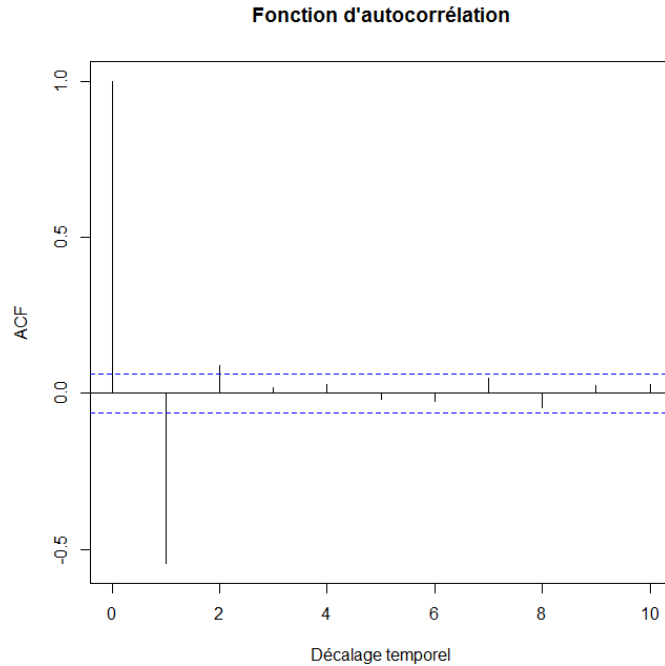


FIGURE 1 – Autocorrélation d'un exemple de série chronologique

On sait que la fonction d'autocorrélation d'un processus à moyenne mobile d'ordre  $q$  s'annule pour tout  $h > q$ . La figure 1 illustre l'autocorrélation empirique d'une série chronologique. Les lignes pointillées délimitent la région où l'autocorrélation est significativement nulle. Grâce à ce graphe, on remarque que l'autocorrélation s'annule à partir de  $h = 3$ . On peut donc conclure que les données semblent représenter un processus à moyenne mobile d'ordre  $q = 2$ . Des logiciels comme R peuvent générer facilement la fonction d'autocorrélation empirique et nous donner une idée du modèle que la série peut suivre. Voyons maintenant si l'on peut appliquer le même procédé pour identifier l'ordre d'un processus autorégressif. Soit le processus AR(1) suivant :

$$X_t = 0,5X_{t-1} + \varepsilon_t \text{ où } \{\varepsilon_t\} \sim \text{WN}(0,1). \quad (4.4)$$

On peut montrer que la fonction d'autocorrélation est donné par  $\rho^X(h) = (0,5)^h$ . La figure 2 représente l'autocorrélation pour mille réalisations de ce processus. On remarque que la fonction prend plus de temps avant d'être significativement nulle et ne s'annule donc pas pour tout  $h > 1$ . De plus, l'équation donnée par le modèle autorégressif indique qu'il y a une corrélation de 0,5 entre les variables  $X_t$  et  $X_{t-1}$ . Or, comme  $X_{t-1}$  satisfait l'équation du modèle, on peut réécrire  $X_t$  sous la forme

$$X_t = 0,5^2 X_{t-2} + 0,5\varepsilon_{t-1} + \varepsilon_t \text{ où } \{\varepsilon_t\} \sim \text{WN}(0,1).$$

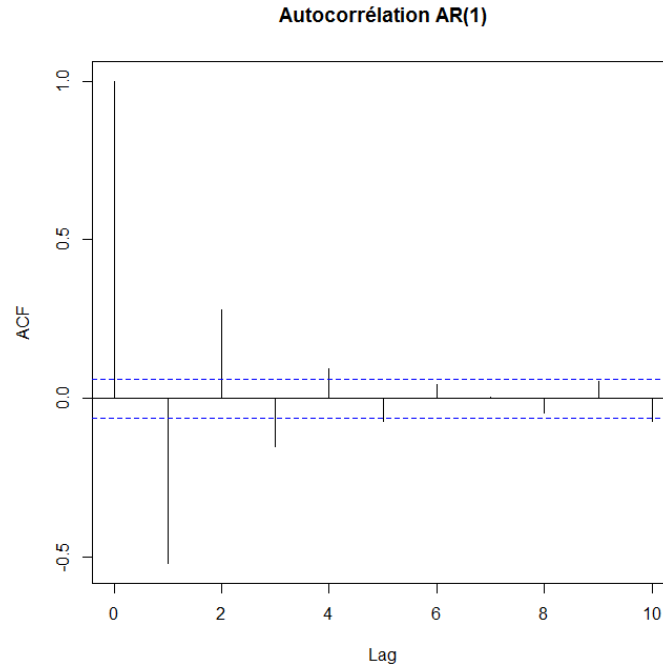


FIGURE 2 – Autocorrélation pour mille réalisations de  $X_t = 0,5X_{t-1} + \varepsilon_t$

On voit donc que la corrélation entre  $X_t$  et  $X_{t-2}$  est de  $0,5^2$ . On peut refaire la même chose pour  $X_{t-h}$  et montrer que l'autocorrélation ne s'annule pas.

## 4.2 Autocorrélation partielle d'un processus AR( $p$ )

L'idée est donc de définir une nouvelle fonction d'autocorrélation (appelée autocorrélation partielle (PACF)) qui calcule la corrélation entre  $X_t$  et  $X_{t+h}$  en ne tenant pas compte de l'influence linéaire des valeurs  $X_{t+1}, X_{t+2}, \dots, X_{t+h-1}$ . On note l'autocorrélation partielle  $\alpha^X(h)$ .

*Remarque 4.1.* Il est possible de démontrer que :

1.  $\alpha^X(0) = 1$  ;
2.  $\alpha^X(1) = \rho^X(1)$ .

Dans l'exemple précédent,  $\alpha^X(-2)$  est l'autocorrélation entre  $X_t$  et  $X_{t-2}$  sans l'influence de  $X_{t-1}$ . Cela revient à dire que :

$$\alpha^X(-2) = \text{Corr}(X_t - 0,5X_{t-1}, X_{t+2} - 0,5X_{t-1}).$$

On remarque que cette nouvelle autocorrélation est plus difficile à calculer que l'ancienne. On peut définir l'autocorrélation partielle de façon plus formelle grâce à l'algorithme de Durbin-Levinson. Le livre de Brockwell [BD09] fournit

de bonnes explications sur cet algorithme. Avec cette nouvelle formulation, on obtient une façon, analogue à (4.1), de tester si notre processus est autorégressif :

$$\text{si } \{X_t\} \sim \text{AR}(p) \text{ , alors } \alpha^X(h) = 0 \quad \text{pour tout } h > p. \quad (4.5)$$

On peut maintenant regarder dans la figure 3 l'autocorrélation partielle(PACF) pour ces mille observations.

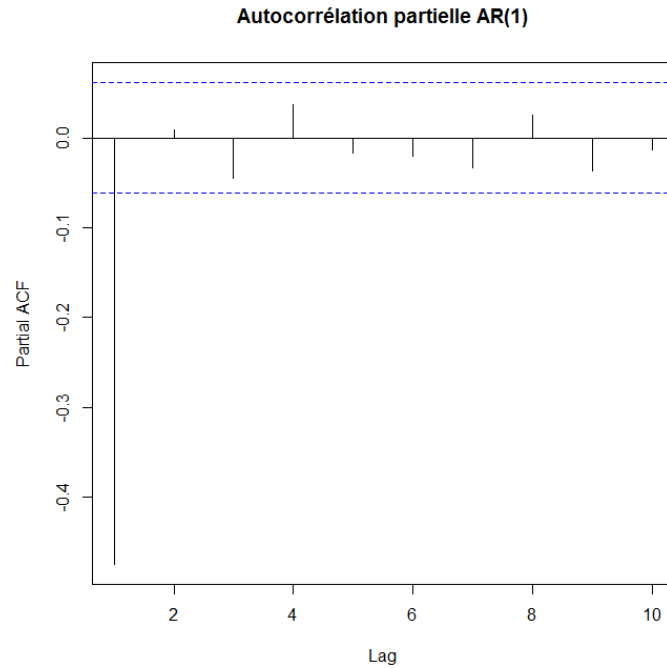


FIGURE 3 – Autocorrélation partielle du processus AR(1)

Maintenant, on a un graphe qui est significativement nul lorsque  $h > 1$ . L'ACF et le PACF sont des outils faciles à utiliser qui permettent de reconnaître les modèles autorégressifs et à moyenne mobile. Dans certains cas, il est possible que ses fonctions décroissent très lentement. À ce moment, il faudra songer à utiliser d'autres types de modèles, par exemple, le modèle ARMA.

### 4.3 Estimation des paramètres

Dans cette section, nous introduisons les équations de Yule-Walker qui permettent d'estimer les paramètres d'un modèle autorégressif. Par la suite, on introduit la méthode du maximum de vraisemblance avec des critères nous permettant de choisir l'ordre d'un modèle AR, MA ou ARMA.

### 4.3.1 Équations de Yule-Walker

Soit  $X_t \sim \text{AR}(p)$ , un processus autorégressif d'ordre  $p$ , stationnaire et causal, de moyenne nulle défini par

$$X_t = \sum_{k=1}^p \phi_k X_{t-k} + \varepsilon_t \text{ où } \{\varepsilon_t\} \sim \text{WN}(0, \sigma_\varepsilon^2). \quad (4.6)$$

L'idée de l'estimateur de Yule-Walker est de trouver des équations reliant les paramètres  $\phi_k$  et l'autocovariance utilisant l'hypothèse de processus causal. Commençons d'abord par multiplier la première équation par  $X_{t-h}$  et prenons l'espérance des 2 côtés de l'équation. Nous obtenons

$$\begin{aligned} r^X(h) &= E \left[ \sum_{k=1}^p \phi_k X_{t-k} X_{t-h} + \varepsilon_t X_{t-h} \right] \\ &= \sum_{k=1}^p \phi_k E[X_{t-k} X_{t-h}] + E[\varepsilon_t X_{t-h}] \\ &= \sum_{k=1}^p \phi_k r^X(k-h) + E[\varepsilon_t X_{t-h}]. \end{aligned} \quad (4.7)$$

Regardons plus attentivement le terme  $E[\varepsilon_t X_{t-h}]$  de (4.7). Comme nous savons que  $\{X_t\}$  est causal, on peut utiliser la relation (3.10) et obtenir :

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

On a donc que

$$\begin{aligned} E[\varepsilon_t X_{t-h}] &= E \left[ \varepsilon_t \sum_{k=0}^{\infty} a_k \varepsilon_{t-h-k} \right] \\ &= E \left[ \sum_{k=0}^{\infty} a_k \varepsilon_t \varepsilon_{t-h-k} \right] \\ &= \sum_{k=0}^{\infty} a_k E[\varepsilon_t \varepsilon_{t-h-k}]. \end{aligned} \quad (4.8)$$

Pour  $h = 0$ , on a

$$E[\varepsilon_t \varepsilon_{t-k}] = \begin{cases} \sigma_\varepsilon^2 & \text{si } k = 0 \\ 0 & \text{sinon.} \end{cases}$$

Pour  $h > 0$ , on aura toujours que  $E[\varepsilon_t \varepsilon_{t-h-k}] = 0$ , car  $\{\varepsilon_t\} \sim \text{WN}(0, \sigma_\varepsilon^2)$ . On obtient donc le système d'équations suivant :

$$r^X(h) = \begin{cases} \sum_{k=0}^p r^X(k-h) \phi_k & \text{si } h = 1, \dots, p \\ \sum_{k=0}^p r^X(k-h) \phi_k + \sigma_\varepsilon^2 & \text{si } h = 0. \end{cases} \quad (4.9)$$



On a un système d'équations à résoudre avec  $p$  équations. Si l'on note par  $r^X(p)$ , le vecteur  $(r^X(1), r^X(2), \dots, r^X(p))'$  et par  $\phi_p$ , le vecteur  $(\phi_1, \phi_2, \dots, \phi_p)'$ , on obtient, en utilisant la matrice suivante,

$$R_p^X = \begin{pmatrix} r^X(0) & r^X(1) & \dots & r^X(p-1) \\ r^X(1) & r^X(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & r^X(1) \\ r^X(p-1) & \dots & r^X(1) & r^X(0) \end{pmatrix},$$

le système

$$\begin{cases} r^X(p) = R_p^X \phi_p \\ r^X(0) = \phi_p' r^X(p) + \sigma_\varepsilon^2. \end{cases}$$

Ces équations sont les *équations de Yule-Walker*. Elles créent un lien entre les paramètres  $\phi_p$  du modèle et la fonction d'autocovariance. Afin d'estimer le vecteur de paramètres  $\phi_p$ , nous devons utiliser la fonction d'autocovariance empirique  $\hat{r}^X(p)$  définie par (4.3) comme estimateur de  $r^X(p)$ . On aura donc un système d'équations qui permettra d'estimer le vecteur  $\phi_p$ . Le système se réécrit de la façon suivante :

$$\begin{cases} \hat{r}^X(p) = \hat{R}_p^X \hat{\phi}_p \\ \hat{r}^X(0) = \hat{\phi}_p' \hat{r}^X(p) + \hat{\sigma}_\varepsilon^2. \end{cases}$$

*Remarque 4.2.*

1. On a que  $\sqrt{T}(\hat{\phi}_p - \phi_p) \xrightarrow{\mathcal{L}} N_p(0, \sigma_\varepsilon^2 R_p^{-1})$  lorsque  $T \rightarrow \infty$  (*Normalité asymptotique*);
2. Si, à partir d'un ARMA( $p, q$ ), on essaie de recréer les équations de Yule-Walker, on se retrouve avec un système non-linéaire qui peut être assez difficile à résoudre.

### 4.3.2 Maximum de vraisemblance

En statistique, l'une des méthodes les plus utilisées pour l'estimation de paramètres est celle du maximum de vraisemblance. C'est Ronald Fisher qui, en 1912 [Fis12], a publié sa méthode initialement appelée « Le critère absolu ». À l'époque, l'une des méthodes couramment utilisées était celle des moindres carrés. Soit l'échantillon  $\{X_1, X_2, \dots, X_T\}$ , la vraisemblance quantifie la probabilité que ces observations proviennent d'une loi paramétrique, par exemple une loi normale. La vraisemblance correspond donc au produit des fonctions de densité des  $X_i$ . On peut montrer que la vraisemblance, sous l'hypothèse de normalité, est donnée par l'expression :

$$\mathcal{L}(R_T^X) = (2\pi)^{-T/2} \cdot |\det(R_T^X)|^{\frac{1}{2}} \cdot e^{-\frac{1}{2} X_T' (R_T^X)^{-1} X_T}, \quad (4.10)$$

où  $R_T^X$  est la matrice de covariance dont l'élément  $(i,j)$  est donné par  $r^X(|i-j|)$ . On écrit la vraisemblance en fonction des paramètres  $\phi_i, \theta_i$  et  $\sigma_\varepsilon^2$  et on maximise la vraisemblance par rapport à ces paramètres. 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.

#### 4.4 Critères de sélection

Nous avons vu jusqu'à maintenant que les fonctions d'autocorrélation et d'autocorrélation partielle nous permettent de déterminer l'ordre d'un modèle autorégressif ou à moyenne mobile. Maintenant, l'idée est de créer des critères statistiques qui choisiront l'ordre du modèle.

##### 4.4.1 Critère AIC et BIC pour processus autorégressif

L'idée du critère AIC est de créer une fonction qui nous permettra de calculer la qualité de l'ajustement. On sait que si le nombre de paramètres augmente, la variance  $\hat{\sigma}_\varepsilon^2$  diminue. Dans le but de ne pas se retrouver avec une surparamétrisation du modèle, on ajoute un facteur qui permettra de faire un compromis entre le nombre de paramètres et la variance minimale. Dans les paragraphes qui suivent, on considère un modèle  $AR(p)$  et on calcule  $\hat{\sigma}_\varepsilon^2$  à l'aide du maximum de vraisemblance pour plusieurs valeurs positives de  $p$ . On pourrait aussi utiliser cette méthode pour un modèle  $MA(q)$ . Le critère AIC consiste à calculer

$$AIC(p) = \log(\hat{\sigma}_\varepsilon^2) + 2\frac{p}{T}.$$

En utilisant ce critère, on remarque que si  $\hat{p}$  est le paramètre obtenu de la minimisation et que  $p$  est le paramètre du vrai modèle, on a la propriété suivante :

$$P(\hat{p} \geq p) \longrightarrow 1 \text{ lorsque } T \longrightarrow \infty.$$

Le critère a donc tendance à choisir un nombre de paramètres plus grand que celui du vrai modèle, ce qui nous conduit à un plus petit terme d'erreur  $\hat{\sigma}_\varepsilon^2$ . Si l'on désire avoir un meilleur choix de l'ordre  $p$ , il existe le critère BIC qui utilise une plus forte pénalité. Le critère BIC sélectionne le paramètre  $p$  qui minimise la quantité suivante :

$$BIC(p) = \log(\hat{\sigma}_\varepsilon^2) + \frac{p}{T} \log(T).$$

##### 4.4.2 Critère AIC et BIC pour ARMA

Il est plus coûteux en calculs de déduire l'ordre  $p$  et  $q$  pour un processus  $ARMA(p,q)$ , car on doit maintenant minimiser une fonction à deux variables. Les

critères AIC et BIC pour un processus ARMA( $p,q$ ) prennent la forme suivante :

$$\text{AIC}(p,q) = \log(\hat{\sigma}_\varepsilon^2) + 2\frac{(p+q)}{T}$$

et

$$\text{BIC}(p,q) = \log(\hat{\sigma}_\varepsilon^2) + \frac{p+q}{T}\log(T).$$

Afin de minimiser ces fonctions, une méthode consiste à faire deux boucles itératives sur  $p$  et  $q$  pour tester tous les couples  $(p,q)$  jusqu'à certaines bornes  $p < P$  et  $q < Q$ . À l'intérieur de ces boucles, on calcule d'abord les estimateurs  $\hat{\phi}$ ,  $\hat{\theta}$  et  $\hat{\sigma}_\varepsilon^2$  utilisant par exemple les moindres carrés ou le maximum de vraisemblance, on calcule les critères AIC et BIC pour ces différents ordres et on trouve le minimum de ces quantités. On a donc les valeurs  $\hat{p}$  et  $\hat{q}$  qui minimisent l'AIC ou le BIC. Ensuite, on calcule des estimateurs efficaces des paramètres du modèle ARMA( $\hat{p},\hat{q}$ ) utilisant la méthode du maximum de vraisemblance.

## 5 Quelques exemples

**Exemple 5.1.** La figure 4 illustre 250 observations issues du modèle AR(2) suivant :

$$X_t = 0,4X_{t-1} - 0,2X_{t-2} + \varepsilon_t, \quad \text{où } \{\varepsilon_t\} \sim \text{N}(0,1). \quad (5.1)$$

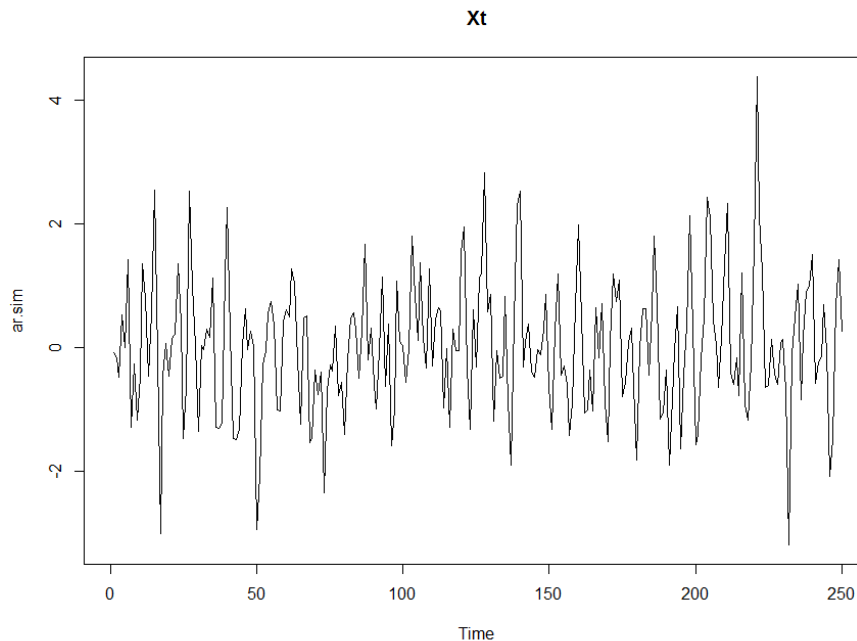


FIGURE 4 – Série  $\{X_t\}$  de l'exemple 5.1 pour 250 observations

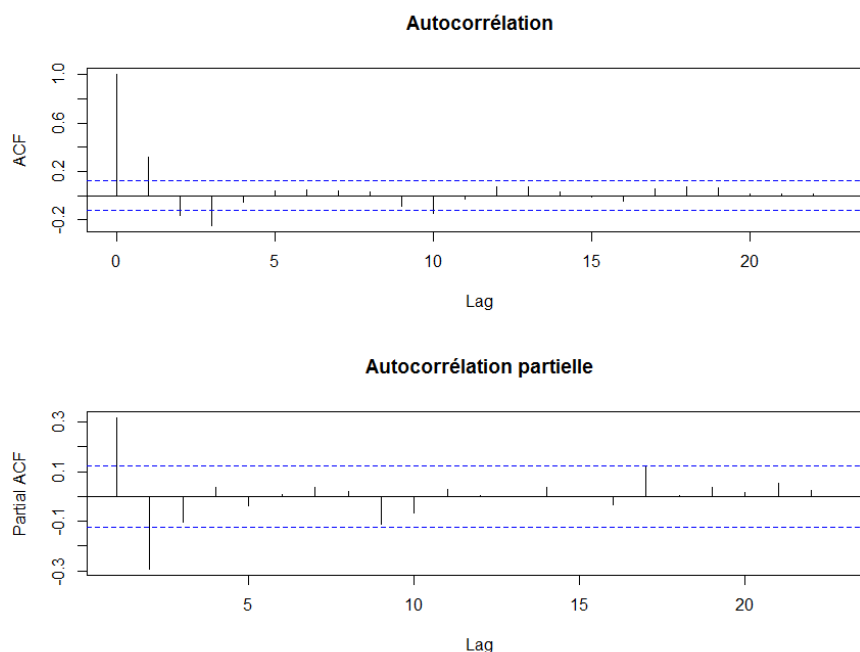


FIGURE 5 – ACF et PACF de l'exemple 5.1

À l'aide de la figure 5, on remarque que l'autocorrélation est significativement nulle à partir de l'horizon  $h = 11$  (lag = 11 sur le graphe). On ne peut pas considérer un modèle MA(10) parce qu'on aurait à estimer 10 paramètres. On se penche donc sur l'autocorrélation partielle. Elle semble être significativement nulle à partir de  $h = 3$  (lag = 3). Le meilleur choix serait donc un modèle AR(2). Regardons maintenant, à l'aide de la table 5.1, une alternative utilisant les critères de sélection automatique.

p\Critère	Aic	Bic
1	0,008335788	0,022421632
2	<b>-0,086972656</b>	<b>-0,058800969</b>
3	-0,075429370	-0,033171839
4	-0,068705027	-0,012361652
5	-0,060676024	0,009753195
6	-0,048572258	0,035942804
7	-0,052288097	0,046312809
8	-0,042556736	0,070130013

TABLE 1 – Critères AIC et BIC pour la série de l'exemple 5.1 AR(2)

Les valeurs en gras sont les valeurs minimales pour chaque critère. On voit bien que chaque critère nous conseille de choisir l'ordre  $\hat{p} = 2$ . Ceci confirme notre intuition obtenue du graphe de l'autocorrélation partielle. Maintenant, si

l'on calcule les estimateurs du maximum de vraisemblance avec  $\hat{p} = 2$ , on obtient les valeurs suivantes :

$$\hat{\phi}_1 = 0,4047 \quad \hat{\phi}_2 = -0,3213 \quad \hat{\sigma}_\varepsilon^2 = 0,8885.$$

On remarque que ces valeurs sont très proches de celles du modèle. Dans certains cas, il arrive que les critères automatiques ne nous donnent pas le même ordre. Par exemple, l'AIC pourrait nous conseiller de prendre l'ordre  $\hat{p} = 3$ , alors que le BIC nous conseillerait  $\hat{p} = 2$ . Il s'agit donc d'ajuster un modèle en considérant l'ordre le plus élevé et de tester si chaque paramètre est significatif.

**Exemple 5.2.** Prenons maintenant la série contenant 600 observations issues du modèle ARMA(2,1) suivant :

$$X_t - 0,4X_{t-1} - 0,2X_{t-2} = \varepsilon_t - 0,2\varepsilon_{t-1}, \text{ où } \{\varepsilon_t\} \sim N(0,1).$$

La figure 6 illustre les deux graphes d'autocorrélation de cette série.

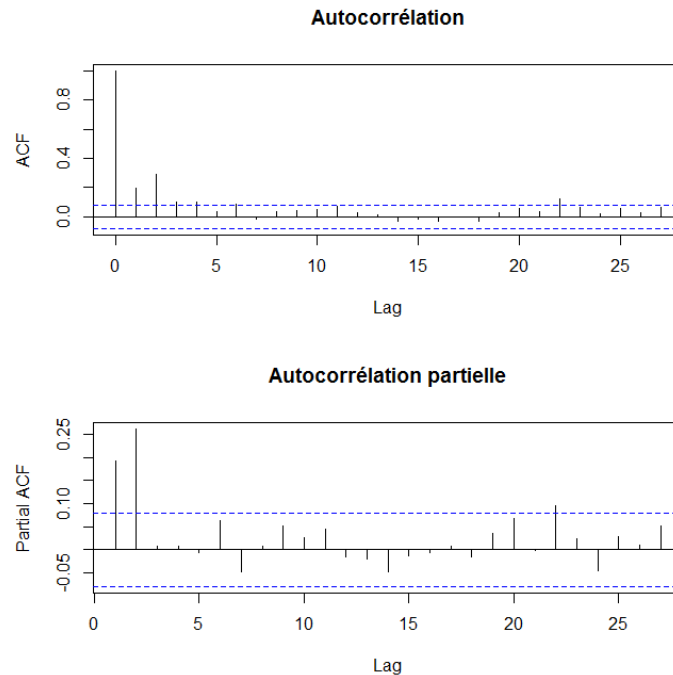


FIGURE 6 – ACF et PACF pour la série de l'exemple 5.2

On voit que l'autocorrélation et l'autocorrélation partielle décroissent assez lentement et prennent du temps avant d'être significativement nulles. Il n'est pas judicieux d'utiliser un modèle autorégressif ou à moyenne mobile. On calcule donc les critères AIC et BIC pour un modèle ARMA( $p,q$ ) pour  $p = 1, \dots, 5$  et  $q = 1, \dots, 5$ . On regarde pour chaque critère le  $\hat{p}$  et  $\hat{q}$  qui minimisent celui-ci. Les tables 2 et 3 résument ces valeurs.

$p \backslash q$	1	2	3	4	5
1	0,03722776	0,01826715	0,01750795	0,02080303	0,02402692
2	0,01488940	0,01812976	0,02038486	0,02284583	0,02478530
3	0,01728771	0,02055846	0,02377663	0,02717506	0,02675940
4	0,02020769	0,02309737	0,02630324	0,02860790	<b>0,01086441</b>
5	0,02271617	0,02622639	0,01608050	0,01867693	0,02988239

TABLE 2 – Critère AIC de la série de l'exemple 5.2 pour  $p = 1, \dots, 5$  et  $q = 1, \dots, 5$

$p \backslash q$	1	2	3	4	5
1	0,05188419	0,04025180	0,04682082	0,05744411	0,06799621
2	<b>0,03687404</b>	0,04744263	0,05702594	0,06681513	0,07608282
3	0,04660058	0,05719954	0,06774592	0,07847257	0,08538513
4	0,05684877	0,06706667	0,07760075	0,08723363	0,07681835
5	0,06668546	0,07752391	0,07470623	0,08463088	0,10316455

TABLE 3 – Critère BIC de la série de l'exemple 5.2 pour  $p, q = 1, \dots, 5$

On remarque que le critère AIC semble proposer un modèle ARMA(4,5). Ensuite, le deuxième choix qui aurait le plus bas AIC serait un modèle ARMA(2,1). Nous avons donc le choix d'estimer neuf paramètres, ou bien d'en estimer trois. Le fait de choisir un modèle à neuf paramètres maximise la vraisemblance du modèle. Ce modèle ajustera très bien notre échantillon, mais pourrait être complètement erroné pour faire des prédictions. On choisira donc d'estimer trois paramètres au lieu de neuf. Une façon de vérifier si ce choix ajuste bien les données seraient de comparer les deux estimateurs de la variance des résidus. Le critère BIC nous donne un ARMA(2,1). On garderait donc le modèle ARMA(2,1) pour ajuster nos données.

**Exemple 5.3.** Étudions maintenant une série réelle assez connue dans des logiciels de statistique, la série AirPassengers. Celle-ci représente le nombre total de passagers ayant voyagé en avion à chaque mois de 1949 à 1960. La figure 7 illustre cette série.

En regardant ce graphique, on se rend compte que certaines portions de la série semblent se répéter. Certains phénomènes reviennent chaque année, on retrouve donc la même variabilité selon la saison. Cependant, malgré le fait que les phénomènes saisonniers se répètent, le nombre de passagers semble augmenter au fil des années selon une fonction linéaire croissante. On utilisera l'opérateur de différentiation  $\Delta$  pour enlever ces phénomènes pour seulement capturer le comportement aléatoire de la série.

On commence d'abord par prendre le logarithme de la série pour stabiliser la variance puisque celle-ci, au début de la série, semble plus faible comparée à celle à la fin de la série. On se retrouvera donc à analyser la série  $\log(X_t)$ . Ensuite, on enlève la saisonnalité. Comme les variations semblent se répéter chaque année et que la variable temporelle est le nombre de mois, il est logique de vouloir appliquer l'opérateur de différentiation 12 fois (12 mois = 1 année). On obtient la nouvelle série  $\Delta^{12} \log(X_t)$ . Cette nouvelle série est illustrée à la figure 8.

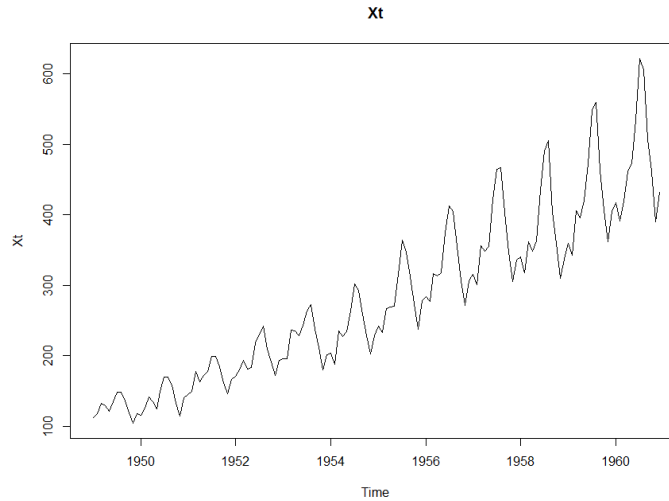


FIGURE 7 – Série AirPassengers

Il nous reste à enlever la tendance (croissance linéaire). Pour ce faire, on différentie une dernière fois pour obtenir  $\{Y_t = \Delta^{13}(\log(X_t))\}$  illustrée à la figure 9.

Cette série est donc plus stationnaire que la série originale. Sa tendance est nulle et sa variance semble stable. Il est possible de tester la stationnarité de la série avec le logiciel R. Voici les critères de sélection pour avoir une idée de l'ordre  $p$  et  $q$  à choisir (tables 4 et 5).

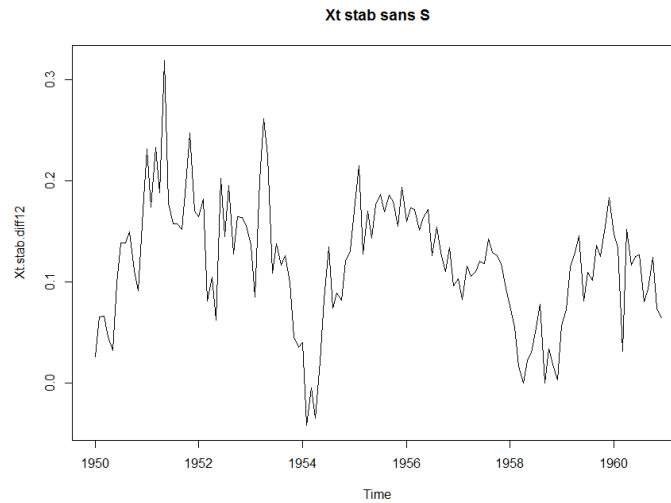
$p \backslash q$	1	2	3	4	5
1	-6,276387	-6,317883	-6,284103	-6,281090	-6,282666
2	-6,326984	-6,321155	-6,417854	-6,412256	-6,334037
3	-6,284579	-6,413744	-6,415536	-6,401048	-6,389769
4	-6,275683	-6,412515	-6,401761	-6,386653	-6,421482
5	-6,303476	-6,290408	-6,440643	<b>-6,484464</b>	-6,347840

TABLE 4 – Critère AIC pour  $\{Y_t = \Delta^{13} \log(X_t)\}$ 

$p \backslash q$	1	2	3	4	5
1	-6,232491	-6,252039	-6,196311	-6,171350	-6,150978
2	-6,261140	-6,233363	<b>-6,308114</b>	-6,280568	-6,180401
3	-6,196787	-6,304003	-6,283848	-6,247411	-6,214185
4	-6,165943	-6,280827	-6,248124	-6,211069	-6,223949
5	-6,171787	-6,136771	-6,265058	-6,286932	-6,128360

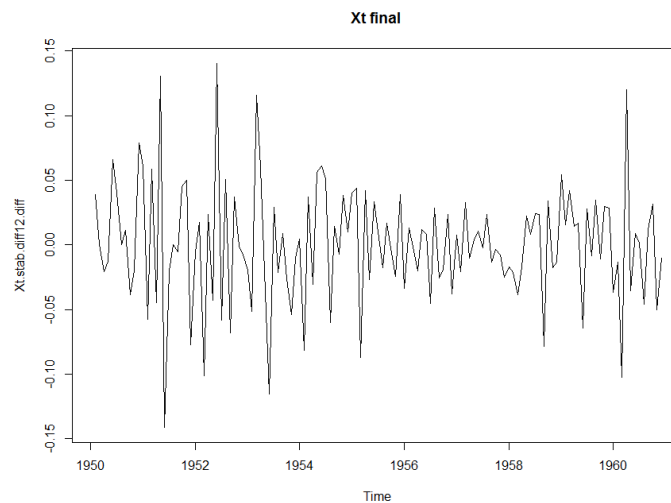
TABLE 5 – Critère BIC pour  $\{Y_t = \Delta^{13} \log(X_t)\}$ 

On remarque que l'AIC semble choisir un modèle ARMA(5,4), car ces ordres minimisent l'AIC. Cependant, en ajustant un modèle ARMA(5,4), on aura à estimer dix paramètres en tout, ce qui semble beaucoup. Il faut se rappeler que

FIGURE 8 – Série  $\{Y_t = \Delta^{12} \log(X_t)\}$ 

l'idée du modèle ARMA était entre autres, de réduire le nombre de paramètres à estimer. De plus, nous savons que l'AIC a tendance à choisir un ordre plus grand que l'ordre réel. On utilise donc le critère BIC. Celui-ci propose de choisir un modèle ARMA(2,3), ce qui est mieux. On va donc ajuster un modèle ARMA(2,3) aux données. Nous obtenons les paramètres estimés suivants :

$$\begin{aligned} \hat{\phi}_1 &= 0,2488 & \hat{\phi}_2 &= -0,8339 & \hat{\theta}_1 &= -0,5904 \\ \hat{\theta}_2 &= 1,0364 & \hat{\theta}_3 &= -0,5205 & \hat{\sigma}_\varepsilon^2 &= 0,001512 \end{aligned}$$

FIGURE 9 – Série finale  $\{Y_t = \Delta^{13} \log(X_t)\}$



$$X_t - 0,2488X_{t-1} + 0,8339X_{t-2} = \varepsilon_t + 0,5904\varepsilon_{t-1} - 1,0364\varepsilon_{t-2} + 0,5205\varepsilon_{t-3},$$

où

$$\{\varepsilon_t\} \sim \text{WN}(0; 0,001512).$$

La dernière étape consiste à vérifier si les termes d'erreur se comportent vraiment comme un bruit blanc. Il faudrait donc vérifier l'espérance et l'auto-corrélation de ces erreurs. Pour ce faire, certaines procédures sont déjà fournies dans le logiciel R. Le test de Ljung-Box peut être utilisé pour vérifier l'autocorrélation. En utilisant ces procédures, on trouve que les erreurs issues de notre série possède bien les caractéristiques d'un bruit blanc. Notre modèle peut donc être un bon choix pour faire des prédictions.

## Références

- [BD09] Peter J. BROCKWELL et Richard A. DAVIS : *Time Series : Theory and Methods*. Springer, 2009.
- [BDR92] H. BOUDJELLABA, J-M. DUFOUR et R. ROY : Testing causality between two vectors in multivariate autoregressive moving average models. *Journal of the American Statistical Association*, 0:1088–1090, 1992.
- [BJ70] G. BOX et G. JENKINS : *Time series analysis : Forecasting and Diagnostic control*. Holden Day, 1970.
- [Fis12] R. A. FISHER : On an absolute criterion for fitting frequency curves. *Gonville and Caius College*, 1912.
- [Gra69] C. W. J. GRANGER : Investigating causal relations by econometric models and cross- spectral methods. *Econometrica*, 37:424–459, 1969.
- [Sch06] A. SCHUSTER : *On the Periodicity of Sun-spots*. Royal Society of London, 1906.
- [Slu27] E. SLUTZKY : The summation of random causes as the source of cyclic processes. *Econometrica*, 5:105–146, 1927.
- [Wol38] H. WOLD : *A study in the Analysis of Stationary Time Series*. Almqvist and Wiksell, 1938.
- [Yul27] U. YULE : On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society*, 226, 1927.

STEVEN FORTIER

DÉPARTEMENT DE MATHÉMATIQUES, UNIVERSITÉ DE SHERBROOKE

Courriel: Steven.Fortier@USherbrooke.ca