

Table des matières

I	Séries chronologiques	3
I.1	Introduction	3
I.1.1	Motivations et objectifs	3
I.1.2	Exemples de séries temporelles	4
I.1.3	Repères historiques	6
I.1.4	Principaux modèles statistiques pour l'étude des séries temporelles	7
I.2	Processus univariés à temps discret	9
I.2.1	Processus stochastique du second ordre	9
I.2.2	Processus stationnaire	10
I.2.3	Autocovariance et autocorrélations	11
I.2.4	Estimation des moments pour les processus stationnaires	13
I.2.5	Tests de blancheur	14
I.2.6	Densité spectrale	15
I.3	Décomposition saisonnière	17
I.3.1	Principe de la décomposition saisonnière	17
I.3.2	Décomposition saisonnière à l'aide de la régression linéaire	17
I.4	Prévision et processus des innovations	20
I.5	Étude des processus AR	21
I.5.1	Définition	21
I.5.2	Processus AR canonique et écriture $MA(\infty)$	22
I.5.3	Autocorrélations simples d'un processus AR	23
I.5.4	Autocorrélations partielle d'un processus AR	24
I.5.5	Exemples	24
I.6	Processus MA	25
I.6.1	Processus MA canonique et écriture $AR(\infty)$	26
I.6.2	Autocorrélations simples d'un processus MA	26
I.6.3	Autocorrélations partielles d'un processus MA	27
I.7	Processus ARMA	28
I.7.1	Processus ARMA canonique et écritures $MA(\infty)$ et $AR(\infty)$	29
I.7.2	Autocorrélations d'un processus ARMA	30
I.7.3	Densité spectrale d'un processus ARMA	31
I.7.4	Estimation des processus ARMA	31
I.7.5	Choix de modèle	34
I.8	Pratique des modèles SARIMA	35
I.8.1	Méthodologie	35

I.8.2 Exemple 38

Chapitre I

Séries chronologiques

I.1 Introduction

I.1.1 Motivations et objectifs

Une **série temporelle**, ou série chronologique, est un ensemble d'observations qui se distinguent par le rôle important que joue l'ordre dans lequel elles ont été recueillies. Les objectifs du cours sont les suivants :

1. Comprendre les problématiques posées par le lien temporel.
2. Etre capable de mettre en oeuvre des techniques de base (statistiques) sur des séries temporelles.

L'importance de ce domaine est illustrée par les nombreux domaines d'application :

- Economie : prévision d'indices économiques...
- Finance : évolution des cours de la bourse...
- Démographie : analyse de l'évolution d'une population...
- Météorologie : analyse de données climatiques...
- Médecine : analyse d'electrocardiogrammes...
- Géophysique : analyse de données sismiques...
- Théorie du signal : transmission de signaux bruités...
- Traitement d'images : analyse d'images satellites, médicales...
- Energie : prévision de la consommation d'électricité...

Les buts poursuivis sont multiples :

- Description : détermination de composantes...
- Filtrage : transformation de la série dans le but d'éliminer certaines caractéristiques ou des valeurs aberrantes...
- Modélisation : recherche de causalité...
- Prévision.

Il existe évidemment des interactions entre ces différents objectifs. Afin de mener l'étude pratique d'une série temporelle, on ne doit pas négliger la phase descriptive, pour envisager éventuellement un filtrage de la série. La modélisation, qui s'avère souvent plus facile sur une série filtrée, fournit les outils pour effectuer des prévisions.

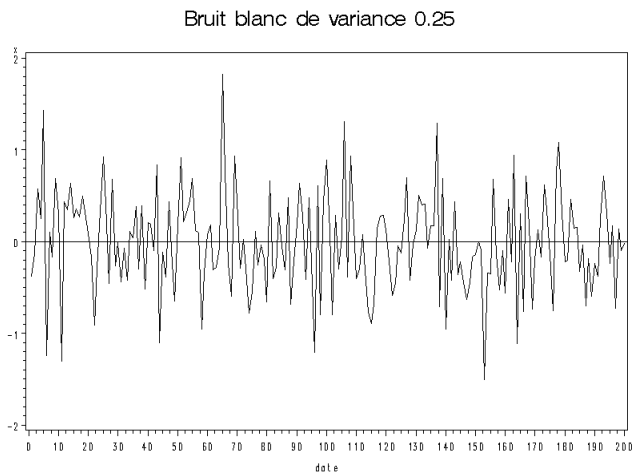
I.1.2 Exemples de séries temporelles

Dans ce document, les exemples, les traitements et donc les graphiques sont obtenus à l'aide du logiciel SAS.

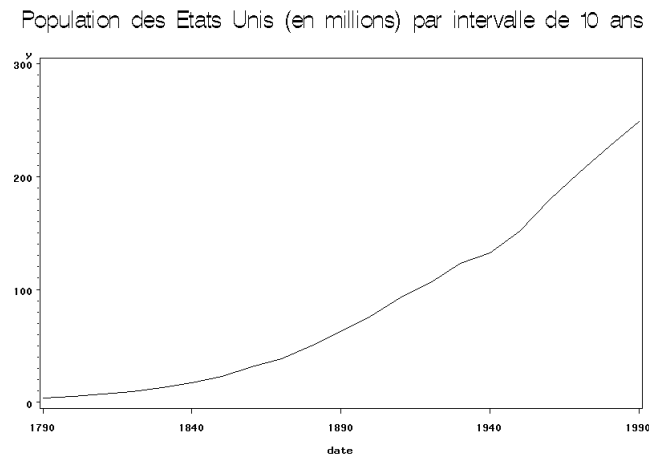
1. La partition de musique est un exemple de série temporelle, même s'il est rare de la modéliser mathématiquement (si on excepte la musique sérielle : Cf. Arnold Schönberg) : modifier l'emplacement d'une note a des conséquences musicales évidentes.



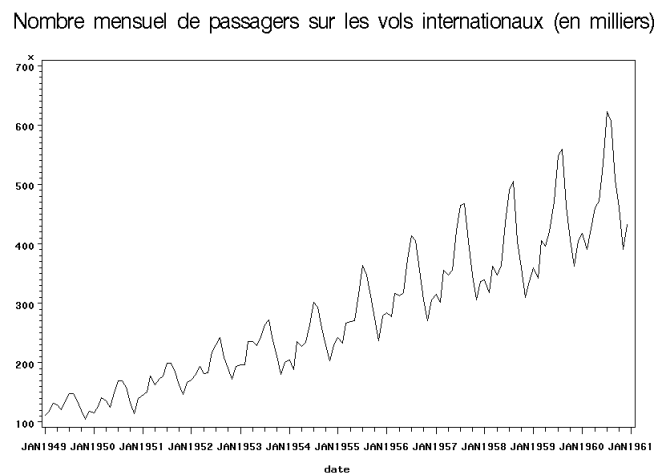
2. Un bruit blanc fort est constitué de variables aléatoires indépendantes et identiquement distribuées (*i.i.d.*), d'espérance nulle. Voici un exemple simulé à partir d'une loi normale $\mathcal{N}(0, 0.25)$.



3. La série suivante représente la population des Etats Unis de 1790 à 1990. On peut observer une tendance de type exponentiel puis linéaire, ainsi qu'un point d'inflexion dont il faudrait déterminer la cause si on veut modéliser le plus correctement possible cette série.

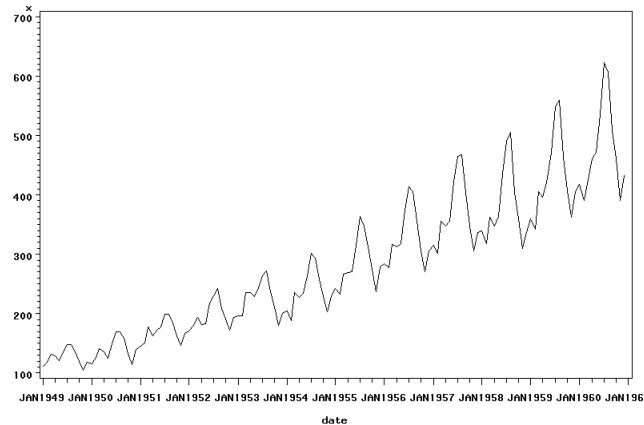


4. Le graphique ci-dessous représente le nombre mensuel de passagers aériens de janvier 1949 à décembre 1960. C'est une des séries les plus utilisées comme exemple d'application et elle figurera abondamment dans ce document sous la dénomination "Nombre de passagers aériens".



On observe une tendance de type exponentiel ainsi qu'une saisonnalité de période 12 (saisonnalité annuelle) qui s'accroît avec le temps. Si on effectue une transformation logarithmique de cette série, on se ramène à une tendance linéaire et à une saisonnalité non volatile :

Nombre mensuel de passagers sur les vols internationaux (en milliers)



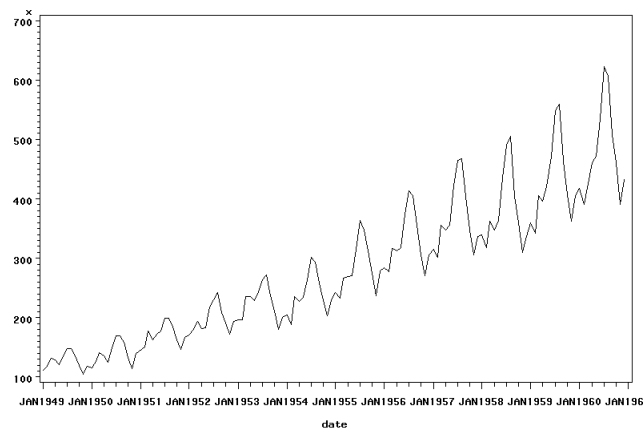
I.1.3 Repères historiques

On peut distinguer trois phases dans l'histoire de l'analyse des séries temporelles :

1. Les séries temporelles apparaissent avec l'utilisation de la représentation graphique, en astronomie.

Le plus ancien diagramme connu figure ci-dessous ; il représente l'inclinaison des planètes en fonction du temps (illustration d'un commentaire du *Songe de Scipion* de Cicéron, extrait des *Saturnales* de Macrobius, en 395 après J.C).

Nombre mensuel de passagers sur les vols internationaux (en milliers)



2. A partir des 18ème et 19ème siècles, on passe de la visualisation graphique aux premières techniques temporelles (déterministes). Citons deux voies très importantes :
 - les travaux de Schuster (1898, 1906) à partir de ceux de Fourier (1807) et Stokes (1879), sur l'analyse fréquentielle d'une série temporelle (un signal est approché par une somme de sinusoïdales) ;
 - les travaux de Pearson (1916) sur la décomposition d'une série temporelle en termes de composantes tendancielle, cyclique, saisonnière et accidentelle.

3. A partir du 20ème siècle, l'aléatoire est pris en compte, notamment à l'aide des travaux de Yule (1927). Ces travaux sont issus de l'observation du mouvement oscillatoire d'un pendule bombardé de petits pois lancés par un enfant !

Il y a ensuite de nombreux contributeurs aux méthodes aléatoires : Cramer (1937, 1951), Wold (1938), Kolmogorov (1939)...

I.1.4 Principaux modèles statistiques pour l'étude des séries temporelles

On présente ici les principales familles de modèles utilisés pour traiter les séries temporelles.

Modèles autorégressifs (“Auto-Regressive”)

Ils ont été introduits par Yule en 1927. On prend en compte une dépendance linéaire du processus à son propre passé :

$$AR(p) : \quad X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + \varepsilon_t,$$

où $p \in \mathbb{N}^*$ est l'ordre du processus, $\varphi_1, \dots, \varphi_p$ sont des constantes réelles et $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc (cf. définition I.3).

Modèles moyennes mobiles (“Moving Average”)

Ils ont également été introduits en 1927, par Slutsky. Un processus moyenne mobile est la somme d'un bruit blanc et des ses retards :

$$MA(q) : \quad X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

où $q \in \mathbb{N}^*$ est fixé et $\theta_1, \dots, \theta_q$ sont des constantes réelles.

Modèles ARMA (“Auto-Regressive Moving Average”)

Développés par Box & Jenkins en 1970, les modèles ARMA sont une combinaison des modèles autorégressif et moyenne mobile :

$$ARMA(p, q) : \quad X_t - \varphi_1 X_{t-1} - \dots - \varphi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}.$$

Les modèles ARIMA (AutoRegressive Integrated Moving Average) et SARIMA (Seasonnal AutoRegressive Integrated Moving Average) (un processus SARIMA est un processus ARMA intégré avec une composante saisonnière) ont ensuite été développés afin de pouvoir modéliser un grand nombre de phénomènes réels qui présentent des tendances et/ou des saisonnalités. On applique en fait des modèles ARMA à des séries dites différenciées ; par exemple, pour un ARIMA d'ordre 1, on suppose que $X_t - X_{t-1}$ est un ARMA (on note en général $\nabla X_t = X_t - X_{t-1}$ ou encore $(I - B)X_t$ en utilisant les opérateurs identité I et rétrograde (“backward”) $BX_t = X_{t-1}$).

Modèles ARCH (“Auto-Regressive Conditional Heteroskedasticity”)

En 1982, Engle a proposé des modèles autorégressifs prenant en compte une “volatilité stochastique” :

$$ARCH(p) : \quad X_t = \varepsilon_t \sqrt{h_t} \text{ avec } h_t = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_p X_{t-p}^2,$$

où $p \geq 1$ est fixé et $\alpha_0, \dots, \alpha_p$ sont des constantes positives.

Modèles à mémoire longue

Les modèles envisagés la plupart du temps sont des modèles dits à mémoire courte : deux instants éloignés du processus n’ont que très peu d’interaction entre eux. Il existe d’autres modélisations pour les processus à mémoire longue, tels que les processus FARIMA (fractional ARIMA) introduits en 1980 par Granger.

Modèles multivariés (“Vector Auto-Regressive”)

On est parfois contraints de modéliser un phénomène multiple, ou plusieurs séries ayant de fortes relations entre elles. Les modèles autorégressifs vectoriels sont un exemple d’une modélisation multivariée :

$$VAR : \quad X_t = AX_{t-1} + E_t \quad \text{où} \quad X_t = (X_t^1, X_t^2),$$

où A est une matrice carrée constante et $(E_t)_{t \in \mathbb{Z}}$ est un bruit blanc multidimensionnel.

Modèles non-paramétriques

Tous les modèles envisagés précédemment sont paramétriques : l’estimation d’un ou plusieurs paramètres suffit pour déterminer les relations temporelles d’un processus. On peut cependant considérer que la fonction de lien n’est pas paramétrée ; on cherche alors à déterminer une fonction f dans des classes de fonctions adaptées traduisant la relation temporelle, par exemple :

$$X_t = f(X_{t-1}, \dots, X_{t-p}) + \varepsilon_t.$$

La fonction f peut être estimée à l’aide de la méthode des noyaux, des séries de Fourier, des ondelettes...

Modèles semi-paramétriques

Les modèles non-paramétriques souffrent du “fléau de la dimension”. On utilise alors une modélisation non-paramétrique sur une ou plusieurs combinaison de variables, par exemple

$$X_t = f(\theta_1 X_{t-1} + \dots + \theta_p X_{t-p}) + \varepsilon_t,$$

où $p \geq 1$ est fixé et $\theta_1, \dots, \theta_p$ sont des constantes.

I.2 Processus univariés à temps discret

Cette section pose les bases de la modélisation probabiliste des séries temporelles. Il définit les notions de processus stochastique et de stationnarité, ainsi que des outils d'analyse comme les autocorrélogrammes et le périodogramme.

Soit $(x_t)_{t \in T}$ une famille d'observations d'un phénomène qui peut être physique, économique, biologique... Chaque observation $x_t \in \mathbb{R}^d$ a été enregistrée à un temps spécifique $t \in T$ et on appelle **série temporelle** cet ensemble d'observations.

Si T est dénombrable (en général $T \subset \mathbb{Z}$), on parle de série temporelle **à temps discret**. Si T n'est pas dénombrable (en général un intervalle de \mathbb{R}), on parle de série temporelle **à temps continu**.

On considère des séries temporelles à valeur dans \mathbb{R}^d . Si $d = 1$, on parle de série **univariée**. Si $d > 1$, on parle de série **multivariée**.

On désignera dans la suite par série temporelle une **série temporelle univariée à temps discret**.

On considère en Statistique que l'observation x est la réalisation d'une variable aléatoire X . De manière analogue, une série temporelle $(x_t)_{t \in T}$ est considérée comme la réalisation d'un processus stochastique (d'une suite de variables aléatoires) $(X_t)_{t \in T}$.

I.2.1 Processus stochastique du second ordre

Définition I.1. Soit $X := (X_t)_{t \in T}$ un processus stochastique (i.e. une suite de variables aléatoires). Le processus X est dit du second ordre si pour tout $t \in T$, X_t est une variable aléatoire de carré intégrable i.e. $\mathbb{E}(|X_t|^2) < +\infty$.

Voici deux exemples de processus du second ordre qui sont fondamentaux dans la suite.

Définition I.2. $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un **bruit blanc fort** si :

- $(\varepsilon_t)_{t \in \mathbb{Z}}$ est une suite de variables aléatoires réelles indépendantes et identiquement distribuées (i.i.d.),
- $\forall t \in \mathbb{Z} : \mathbb{E}(\varepsilon_t) = 0$ et $\mathbb{E}(\varepsilon_t^2) = \sigma^2$.

Définition I.3. $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un **bruit blanc faible** si :

- $(\varepsilon_t)_{t \in \mathbb{Z}}$ est une suite de variables aléatoires réelles identiquement distribuées,
- $\forall (t, t') \in \mathbb{Z}^2, t \neq t' : \text{Cov}(\varepsilon_t, \varepsilon_{t'}) = 0$,
- $\forall t \in \mathbb{Z} : \mathbb{E}(\varepsilon_t) = 0$ et $\mathbb{E}(\varepsilon_t^2) = \sigma^2$.

On rappelle maintenant un résultat important pour l'étude de processus du second ordre.

Proposition I.4. Soient $(X_n)_{n \in \mathbb{Z}}$ et $(Y_n)_{n \in \mathbb{Z}}$ deux processus tels que

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left[\left(\sum_{i=n}^{+\infty} X_i \right)^2 + \left(\sum_{i=-\infty}^{-n} X_i \right)^2 \right] = 0 \text{ et } \lim_{n \rightarrow +\infty} \mathbb{E} \left[\left(\sum_{i=n}^{\infty} Y_i \right)^2 + \left(\sum_{i=-\infty}^{-n} Y_i \right)^2 \right] = 0.$$

Alors, on a

$$\text{Cov} \left(\sum_{i=-\infty}^{+\infty} X_i, \sum_{j=-\infty}^{+\infty} Y_j \right) = \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} \text{Cov}(X_i, Y_j)$$

I.2.2 Processus stationnaire

Dans de très nombreux cas, on ne peut pas renouveler la suite de mesures dans des conditions identiques (par exemple le taux de chômage mensuel). Alors pour que le modèle déduit à partir d'une suite d'observations ait un sens, il faut que toute portion de la trajectoire observée fournisse des informations sur la loi du processus et que des portions différentes, mais de même longueur, fournissent les mêmes indications. D'où la notion de stationnarité.

Définition I.5. Un processus $(X_t)_{t \in T}$ est **fortement stationnaire** ou **strictement stationnaire** si, pour tous $k \geq 1$, $(t_1, \dots, t_k) \in T^k$, h tel que $(t_1 + h, \dots, t_k + h) \in T^k$, les vecteurs $(X_{t_1}, \dots, X_{t_k})$ et $(X_{t_1+h}, \dots, X_{t_k+h})$ ont même loi.

Cette propriété très forte est très difficile à vérifier, d'où la notion de stationnarité faible.

Définition I.6. Un processus $(X_t)_{t \in T}$ du second ordre est **faiblement stationnaire** ou **stationnaire à l'ordre 2**, si son espérance $\mathbb{E}(X_t)$ et ses autocovariances $\mathbb{Cov}(X_s, X_t)$ sont invariantes par translation dans le temps :

- $\forall t \in T, \quad \mathbb{E}(X_t) = \mathbb{E}(X_0),$
- $\forall (s, t) \in T^2, \forall h / (s + h, t + h) \in T^2, \quad \mathbb{Cov}(X_s, X_t) = \mathbb{Cov}(X_{s+h}, X_{t+h}).$

Dans la suite, on notera $\mu_X = \mathbb{E}(X_0)$ et $\gamma_X(h) = \mathbb{Cov}(X_0, X_h)$.

Remarque I.7.

- La stationnarité faible est bien plus facile à étudier et à vérifier que la stationnarité stricte.
- Un processus stationnaire n'est pas obligatoirement borné ou "sympathique" (par exemple, un bruit blanc).
- Pour un processus du second ordre, la stationnarité stricte implique la stationnarité faible. La réciproque est fausse (elle n'est vraie que pour les processus gaussiens).

◇

Exemple I.8. 1. Un bruit blanc fort est fortement stationnaire.

2. Un bruit blanc faible est faiblement stationnaire.

3. Un processus présentant une tendance et/ou une saisonnalité n'est pas faiblement stationnaire.

◇

On désignera dans la suite par processus stationnaire, un processus faiblement stationnaire (donc du second ordre). De plus, on supposera que le processus est à temps discret et plus précisément que $T = \mathbb{Z}$.

Proposition I.9. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire (au sens faible). Soit une suite $(a_i)_{i \in \mathbb{Z}}$ telle que $\sum_{i \in \mathbb{Z}} |a_i| < +\infty$. Le processus $(Y_t)_{t \in \mathbb{Z}}$ où $Y_t = \sum_{i \in \mathbb{Z}} a_i X_{t-i}$, est stationnaire et

- $\mu_Y = \mu_X \sum_{i \in \mathbb{Z}} a_i$ où $\mu_X = \mathbb{E}(X)$,
- $\gamma_Y(h) = \sum_{i \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} a_i a_j \gamma_X(h + j - i).$

Cela implique notamment qu'une somme de v.a.r d'un processus stationnaire est stationnaire.

Notons que la somme de deux processus stationnaires n'est pas forcément stationnaire.

Il existe des non-stationnarités particulières dont :

- $(X_t)_{t \in \mathbb{Z}}$ est un processus **non-stationnaire TS** (Trend Stationary) s'il peut s'écrire sous la forme : $X_t = f(t) + Y_t$, où f est une fonction déterministe et $(Y_t)_{t \in \mathbb{Z}}$ un processus stationnaire.
- $(X_t)_{t \in \mathbb{Z}}$ est un processus **non-stationnaire DS** (Difference Stationary) s'il est stationnaire après d différenciations : $\nabla^d X_t = (I - B)^d X_t$ où $BX_t = X_{t-1}$. Si $d = 2$, cela signifie que le processus défini pour tout $t \in \mathbb{Z}$ par $\nabla^2 X_t = X_t - 2X_{t-1} + X_{t-2}$ est stationnaire.

Ces processus peuvent néanmoins être rendus stationnaires par une suite de transformations usuelles (désaisonnalisation, différenciation, transformation non linéaire...). Une méthodologie classique est celle de Box-Jenkins que nous étudierons par la suite et en TP.

I.2.3 Autocovariance et autocorrélations

L'autocovariance (resp. l'autocorrélation) d'un processus stochastique est la covariance (resp. la corrélation) de ce processus avec une version décalé dans le temps de lui même. Ces fonctions sont bien définies pour un processus stationnaire. Dans la suite on considère un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$.

Fonction d'autocovariance

Définition I.10. On appelle **fonction d'autocovariance** du processus X la fonction γ suivante :

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

Proposition I.11. La fonction d'autocovariance vérifie les propriétés suivantes :

- $\gamma(0) \geq 0$,
- $|\gamma(h)| \leq \gamma(0)$,
- γ est une fonction symétrique : pour tout $h \in \mathbb{N}$, $\gamma(-h) = \gamma(h)$,
- γ est une fonction semi-définie positive :

$$\forall n \in \mathbb{N}^*, \forall (a_i)_{i \in \{1, \dots, n\}} \in \mathbb{R}^n : \sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma(i-j) \geq 0.$$

Comme la fonction γ est symétrique, on calculera la fonction d'autocovariance pour $h \in \mathbb{N}$. Réciproquement, si une fonction γ vérifie :

- $\gamma(-h) = \gamma(h)$, $h \in \mathbb{N}^*$,
- $\forall n \in \mathbb{N}^*, \forall (a_i)_{i \in \{1, \dots, n\}} \in \mathbb{R}^n : \sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma(i-j) \geq 0$,

alors c'est une fonction d'autocovariance.

Autocorrélogramme simple

Définition I.12. On appelle **autocorrélogramme simple** du processus X la fonction ρ suivante :

$$\forall h \in \mathbb{Z} : \rho(h) = \text{Corr}(X_t, X_{t-h}) = \frac{\gamma(h)}{\gamma(0)}$$

L'autocorrélogramme simple est donc la fonction d'autocovariance renormalisée. Il vérifie des propriétés similaires.

Proposition I.13. *On a :*

- $\rho(0) = 1$,
- $|\rho(h)| \leq 1$,
- ρ est une fonction symétrique : $\forall h \in \mathbb{N} : \rho(-h) = \rho(h)$,
- ρ est une fonction définie positive :

$$\forall n \geq 0, \forall (a_i)_{i \in \{1, \dots, n\}} : \sum_{i=1}^n \sum_{j=1}^n a_i a_j \rho(i-j) \geq 0.$$

Définition I.14. *On appelle matrice d'autocorrélation de (X_t, \dots, X_{t-h+1}) avec $h \in \mathbb{N}^*$ la matrice suivante :*

$$R_h = \begin{bmatrix} 1 & \rho(1) & \dots & \rho(h-1) \\ \rho(1) & 1 & \dots & \dots \\ \dots & \dots & \dots & \rho(1) \\ \rho(h-1) & \dots & \rho(1) & 1 \end{bmatrix} \quad (\text{I.1})$$

Proposition I.15. *La fonction ρ est une fonction définie positive si et seulement si $\forall h \in \mathbb{N}^* : \det R_h \geq 0$.*

La seconde condition fixe de nombreuses contraintes aux corrélations, par exemple :

$$\det R_3 \geq 0 \Leftrightarrow [1 - \rho(2)] [1 + \rho(2) - 2\rho^2(1)] \geq 0.$$

Autocorrélogramme partiel

Une seconde façon de mesurer l'influence entre le processus et son décalé dans le temps est de calculer la corrélation entre deux instants en enlevant une partie de l'information contenue entre ces deux instants. Plus précisément, on calcule la corrélation entre le terme $X_t - \mathbb{E}\mathbb{L}(X_t|X_{t-1}, \dots, X_{t-h+1})$ et le terme $X_{t-h} - \mathbb{E}\mathbb{L}(X_{t-h}|X_{t-1}, \dots, X_{t-h+1})$, où la quantité $\mathbb{E}\mathbb{L}(X_t|X_{t-1}, \dots, X_{t-h+1})$ (resp. $\mathbb{E}\mathbb{L}(X_{t-h}|X_{t-1}, \dots, X_{t-h+1})$) désigne la régression linéaire de X_t (resp. X_{t-h}) sur $X_{t-1}, \dots, X_{t-h+1}$.

On rappelle que la régression linéaire d'une variable aléatoire Z sur n variables aléatoires Y_1, \dots, Y_n est la variable aléatoire, notée $\mathbb{E}\mathbb{L}(Z|Y_1, \dots, Y_n)$, est la combinaison linéaire des Y_1, \dots, Y_n qui minimise $\bar{Z} \mapsto \text{Var}(Z - \bar{Z})$. Plus précisément, on a

$$\mathbb{E}\mathbb{L}(Z|Y_1, \dots, Y_n) = \sum_{i=1}^n a_i Y_i \quad \text{avec} \quad (a_1, \dots, a_n) = \underset{(x_1, \dots, x_n) \in \mathbb{R}^n}{\operatorname{argmin}} \quad \text{Var}\left(Z - \sum_{i=1}^n x_i Y_i\right).$$

Définition I.16. *On appelle autocorrélogramme partiel du processus X la fonction r suivante : $r(1) = \rho(1)$ et pour tout $h \in \mathbb{N}^*$,*

$$\begin{aligned} r(h) &= \text{Corr}(X_t, X_{t-h} | X_{t-1}, \dots, X_{t-h+1}) \\ &= \frac{\text{Cov}(X_t - \mathbb{E}\mathbb{L}(X_t|X_{t-1}, \dots, X_{t-h+1}), X_{t-h} - \mathbb{E}\mathbb{L}(X_{t-h}|X_{t-1}, \dots, X_{t-h+1}))}{\text{Var}(X_t - \mathbb{E}\mathbb{L}(X_t|X_{t-1}, \dots, X_{t-h+1}))}. \end{aligned}$$

Théorème I.17. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire. On considère la régression linéaire de X_t sur X_{t-1}, \dots, X_{t-h} :

$$\mathbb{E}(X_t | X_{t-1}, \dots, X_{t-h}) = \sum_{i=1}^h a_i(h) X_{t-i},$$

et $\varepsilon_t = X_t - \mathbb{E}(X_t | X_{t-1}, \dots, X_{t-h})$. Alors on a

- $\mathbb{E}(\varepsilon_t) = 0$ et il existe $\sigma > 0$ tel que pour tout $t \in \mathbb{Z}$, $\mathbb{E}(\varepsilon_t^2) = \sigma^2$
- $\forall i \in \{1, \dots, h\} : \mathbb{E}(\varepsilon_t X_{t-i}) = 0$

Et on a $a_h(h) = r(h)$ ainsi que

$$\begin{pmatrix} \rho(1) \\ \rho(2) \\ \dots \\ \rho(h) \end{pmatrix} = R_h \begin{pmatrix} a_1(h) \\ a_2(h) \\ \dots \\ a_h(h) \end{pmatrix}.$$

D'après cette proposition, si on a une estimation de $(\rho(1), \dots, \rho(h))$ et donc de la matrice d'autocorrélation R_h définie par (I.1), alors on est capable d'estimer $(a_1(h), \dots, a_h(h))$ (en inversant la matrice R_h). On obtient aussi, d'après le théorème précédent, une estimation de l'autocorrélogramme partiel $r(h)$.

Une façon moins coûteuse d'inverser la matrice R_h pour déterminer les autocorrélations partielles est d'utiliser l'algorithme de Durbin-Levinson.

Algorithme I.18. Algorithme de Durbin-Levinson :

$$\begin{aligned} a_1(1) &= \rho(1) \\ \forall h \geq 2, \forall i \in \{1, \dots, h-1\} : a_i(h) &= a_i(h-1) - a_h(h) a_{h-i}(h-1), \\ \forall h \geq 2 : a_h(h) &= \frac{\rho(h) - \sum_{i=1}^{h-1} \rho(h-i) a_i(h-1)}{1 - \sum_{i=1}^{h-1} \rho(i) a_i(h-1)}. \end{aligned} \quad \diamond$$

I.2.4 Estimation des moments pour les processus stationnaires

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire.

Espérance

L'estimateur naturel (sans biais) de $\mathbb{E}(X) = \mu$ à partir de (X_1, \dots, X_T) est la moyenne empirique $\overline{X}_T : \overline{X}_T = \frac{1}{T} \sum_{i=1}^T X_i$. C'est un estimateur convergent vers μ lorsque T tend vers l'infini.

Autocovariance et autocorrélation

Si on dispose de (X_1, \dots, X_T) alors on peut considérer les deux estimateurs suivants de la fonction d'autocorrélation $\gamma(h)$: pour tout $h \in \{1, \dots, T-1\}$

$$\hat{\gamma}(h) = \frac{1}{T} \sum_{t=h+1}^T (X_t - \overline{X}_T) (X_{t-h} - \overline{X}_T),$$

et

$$\bar{\gamma}(h) = \frac{1}{T-h} \sum_{t=h+1}^T (X_t - \bar{X}_T) (X_{t-h} - \bar{X}_T).$$

Ils sont tous les deux convergents et asymptotiquement sans biais mais le biais de $\bar{\gamma}(h)$ est plus petit que celui de $\hat{\gamma}(h)$. Par contre le premier est défini positif tandis que le second ne l'est pas en général.

On en déduit l'estimateur suivant de l'autocorrélogramme simple $\rho(h)$:

$$\forall h \in \{1, \dots, T-1\} : \hat{\rho}(h) = \frac{\sum_{t=h+1}^T (X_t - \bar{X}_T) (X_{t-h} - \bar{X}_T)}{\sum_{t=1}^T (X_t - \bar{X}_T)^2}$$

Remarque I.19. On peut noter que :

- Appliquer ces estimateurs avec $h = T-1$ pose problème.
- On peut faire les calculs même lorsque le processus n'est pas stationnaire !
- Les estimations des autocorrélations partielles se déduisent des estimations des autocorrélations simples grâce à l'algorithme de Durbin-Levinson.

◇

I.2.5 Tests de blancheur

Il existe différents tests de blancheur nécessaires notamment pour valider les modélisations SARIMA (Seasonal ARIMA), notamment le **test de Portmanteau**.

On considère le test suivant (de Portmanteau ou de Ljung-Box) :

$$\begin{cases} H_0 : (X_t)_{t \in \mathbb{Z}} \text{ est un bruit blanc} \\ H_1 : (X_t)_{t \in \mathbb{Z}} \text{ n'est pas un bruit blanc} \end{cases}$$

Si on dispose de (X_1, \dots, X_T) , on considère la statistique suivante (calculée sur les k premières estimations des autocorrélations) :

$$\bar{Q}_k = T \sum_{h=1}^k \hat{\rho}^2(h)$$

Une trop grande valeur de \bar{Q}_k indique que les autocorrélations sont trop importantes pour être celles d'un bruit blanc. On peut également considérer la statistique (qui converge plus vite) :

$$Q_k = T(T+2) \sum_{h=1}^k \frac{\hat{\rho}^2(h)}{T-h}$$

Sous H_1 , Q_k diverge vers l'infini quand T tend vers l'infini. Asymptotiquement, sous H_0 , Q_k suit une loi du χ^2 à k degrés de liberté. On rejette donc l'hypothèse H_0 au niveau α si :

$$Q_k > \chi_k^2(1-\alpha)$$

où $\chi_k^2(1 - \alpha)$ désigne le quantile d'ordre $1 - \alpha$ d'une loi du χ^2 à k degrés de liberté. En général on préfère travailler sur la p -valeur :

$$p\text{-valeur} = \mathbb{P}(\chi_k^2 \geq Q_k)$$

On rejette l'hypothèse H_0 au niveau α si : $p\text{-valeur} < \alpha$.

Exemple I.20. Le logiciel SAS permet d'effectuer le test de blancheur. La sortie suivante correspond au test de blancheur appliquée à un bruit blanc simulé :

Autocorrelation Check for White Noise									
To	Chi-		Pr >	Autocorrelations					
Lag	Square	DF	ChiSq						
6	7.09	6	0.2126	-0.044	0.042	-0.152	0.026	0.066	0.041
12	11.46	12	0.4927	-0.076	-0.105	0.032	-0.006	-0.017	0.032
18	13.96	18	0.7280	-0.027	-0.032	0.032	0.000	-0.093	0.003
24	20.95	24	0.6642	0.022	-0.114	-0.009	0.040	-0.093	-0.044

◇

Dans cet exemple, on a choisit $k = 6, 12, 18$ et 24 . On peut lire par exemple :

$$q_{18} = 13.96 \text{ où } q_{18} \text{ est la réalisation de } Q_{18},$$

$$p\text{-valeur} = 0.7320.$$

On ne rejette donc pas (fort heureusement) l'hypothèse de blancheur de cette série.

I.2.6 Densité spectrale

Définition I.21. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire de fonction d'autocovariance γ .

On appelle **densité spectrale** de X la transformée de Fourier discrète, f , de la suite des autocovariances (lorsqu'elle existe) :

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} \gamma(h) e^{-ih\omega}$$

Cette densité spectrale existe lorsque :

$$\sum_{h=-\infty}^{+\infty} |\gamma(h)| < +\infty$$

Si elle existe elle est continue, positive, paire et 2π -périodique (ce qui permet de l'étudier uniquement sur $[0, \pi]$).

Théorème I.22 (Théorème spectral). Si f est la densité spectrale de X alors :

$$\gamma(h) = \int_{-\pi}^{+\pi} f(\omega) e^{ih\omega} d\omega.$$

Exemple I.23. Soit $(\varepsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc faible de variance σ^2 . On a : $\gamma(h) = \sigma^2$ si $h = 0$ et $\gamma(h) = 0$ sinon. On en déduit donc :

$$f(\omega) = \frac{\sigma^2}{2\pi}$$

Réciproquement, si la densité spectrale est constante, alors le processus correspondant est un bruit blanc faible. ◇

Définition I.24. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire.

Si on dispose de (X_1, \dots, X_T) , on appelle périodogramme la fonction I_T suivante :

$$I_T(\omega) = \frac{1}{T} \left| \sum_{t=1}^T X_t e^{-it\omega} \right|^2$$

Si le processus $(X_t)_{t \in \mathbb{Z}}$ admet une densité spectrale, alors $\frac{1}{2\pi} I_T(\omega)$ est un estimateur sans biais de la densité spectrale (mais non-convergent !). On peut résoudre le problème de convergence en lissant le périodogramme et en considérant l'estimateur suivant :

$$\hat{f}(\omega) = \frac{1}{2\pi} \sum_{|j| \leq m_T} W_T(j) I_T \left(g(T, \omega) + \frac{2\pi j}{T} \right),$$

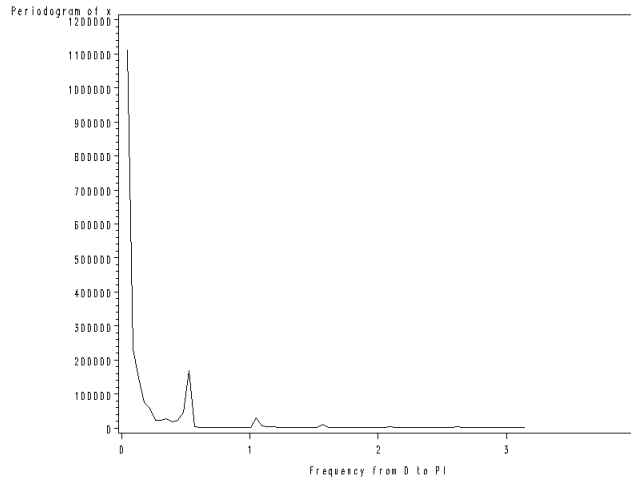
où

- $g(T, \omega)$ est le multiple de $2\pi/T$ le plus proche de ω ,
- $m_T \rightarrow +\infty$ et $\frac{m_T}{T} \rightarrow 0$ quand $T \rightarrow +\infty$,
- $\forall j \in \mathbb{Z}, W_T(j) \geq 0, W_T(-j) = W_T(j)$,
- $\sum_{|j| \leq m_T} W_T(j) = 1$ et $\sum_{|j| \leq m_T} W_T^2(j) \rightarrow 0$.

Par exemple les poids constants $W_T(j) = 1/(2m_T + 1)$ conviennent, mais il existe d'autres choix avec de meilleures propriétés.

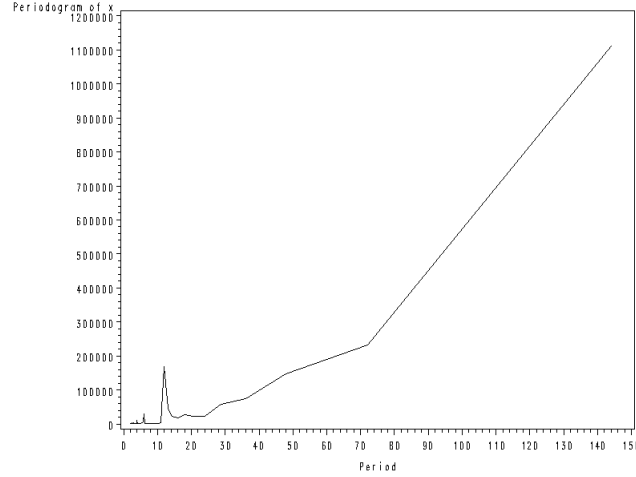
Une grande valeur du périodogramme suggère que la série a une composante saisonnière à la fréquence correspondante.

Exemple I.25. Le graphique ci-dessous représente le périodogramme lissé du nombre de passagers aériens, dans le domaine fréquentiel :



Le graphique ci-dessous représente le périodogramme du nombre de passagers aériens, dans le domaine temporel¹ :

¹On passe du domaine fréquentiel au domaine temporel à l'aide de la relation : $T = \frac{2\pi}{\omega}$.



◇

I.3 Décomposition saisonnière

Cette partie présente la décomposition saisonnière, à l'aide de la régression et des moyennes mobiles.

I.3.1 Principe de la décomposition saisonnière

Il est courant de décomposer un processus $(X_t)_{t \in \{1, \dots, T\}}$ en différents termes :

- Une tendance : T_t
- Une composante saisonnière : S_t (de période p)
- Un résidu aléatoire : ε_t
- Et éventuellement une composante cyclique (de long terme) : C_t

Il existe différentes modèles possibles dont :

- Le modèle additif : $X_t = T_t + S_t + \varepsilon_t$, où $\mathbb{E}(\varepsilon_t) = 0$ et $\mathbb{V}(\varepsilon_t) = \sigma^2$.
- Le modèle multiplicatif : $X_t = T_t S_t (1 + \varepsilon_t)$, où $\mathbb{E}(\varepsilon_t) = 0$ et $\mathbb{V}(\varepsilon_t) = \sigma^2$.

I.3.2 Décomposition saisonnière à l'aide de la régression linéaire

Principe

On suppose que les composantes tendancielle et saisonnières sont des combinaisons linéaires de fonctions connues dans le temps :

$$T_t = \sum_{i=1}^n \alpha_i T_t^i, \quad S_t = \sum_{j=1}^p \beta_j S_t^j,$$

où :

$$S_t^j = \begin{cases} 1 & \text{si } t = j[p], \\ 0 & \text{sinon.} \end{cases}$$

Exemple I.26. On peut considérer les tendances suivantes :

1. Tendance linéaire : $T_t = \alpha_0 + \alpha_1 t$
2. Tendance quadratique : $T_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2$
3. Tendance exponentielle : $T_t = c\alpha^t$
4. Tendance de Gompertz : $T_t = \exp(c_1 \alpha^t + c_2)$

Les deux premiers cas peuvent se résoudre à l'aide de la régression linéaire. Les deux suivants se résolvent par optimisation (en minimisant par exemple l'erreur quadratique). \diamond

Estimation dans le cas d'un modèle additif

Le modèle considéré est le suivant :

$$\forall t \in \{1, \dots, T\} : X_t = \sum_{i=1}^n \alpha_i T_t^i + \sum_{j=1}^p \beta_j S_t^j + \varepsilon_t,$$

où $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc. On cherche à minimiser l'erreur quadratique

$$\sum_{t=1}^T (X_t - \sum_{i=1}^n \alpha_i T_t^i - \sum_{j=1}^p \beta_j S_t^j)^2$$

en les paramètres $\alpha_1, \dots, \alpha_n$ et β_1, \dots, β_p , où n et p sont fixés. On utilise les notations suivantes :

$$D = \begin{bmatrix} T_1^1 & \dots & T_1^n \\ \dots & \dots & \dots \\ T_T^1 & \dots & T_T^n \end{bmatrix} \quad S = \begin{bmatrix} S_1^1 & \dots & S_1^p \\ \dots & \dots & \dots \\ S_T^1 & \dots & S_T^p \end{bmatrix}.$$

Avec ces notations, le système s'écrit sous la forme vectorielle suivante :

$$X = D\alpha + S\beta + \varepsilon = Yb + \varepsilon,$$

où :

$$X = \begin{pmatrix} X_1 \\ \dots \\ X_T \end{pmatrix}; \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_T \end{pmatrix}; \alpha = \begin{pmatrix} \alpha_1 \\ \dots \\ \alpha_n \end{pmatrix}; \beta = \begin{pmatrix} \beta_1 \\ \dots \\ \beta_p \end{pmatrix}; Y = [D, S]; b = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}.$$

Un estimateur de b dans $X = Yb + \varepsilon$ par les moindres carrés ordinaires (MCO) est :

$$\hat{b} = (Y'Y)^{-1} Y'X.$$

Ainsi on obtient les estimateurs de α et β :

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{bmatrix} D'D & D'S \\ S'D & S'S \end{bmatrix}^{-1} \begin{pmatrix} D'X \\ S'X \end{pmatrix}$$

De plus, ce sont des estimateurs sans biais

$$\mathbb{E} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix},$$

et de variance

$$\mathbb{V} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = s^2 \begin{bmatrix} D'D & D'S \\ S'D & S'S \end{bmatrix}^{-1}$$

où $s^2 = \frac{1}{T-n-p} \sum_{t=1}^T \varepsilon_t^2$ est un estimateur convergent (lorsque T tend vers l'infini) et sans biais de la variance de ε_t .

Résultats

Une fois les estimations effectuées, on peut obtenir la série ajustée \hat{X}_t et la série corrigée des variations saisonnières X_t^{CVS} :

$$\begin{aligned} - \hat{X}_t &= \sum_{i=1}^n \hat{\alpha}_i T_t^i + \sum_{j=1}^p \hat{\beta}_j S_t^j \\ - X_t^{CVS} &= X_t - \hat{X}_t \end{aligned}$$

Prévision

Si on désire prévoir X_{T+h} , on peut supposer que le modèle est toujours valable à cet horizon, ce qui donne :

$$X_{T+h} = \sum_{i=1}^n \alpha_i T_{T+h}^i + \sum_{j=1}^p \beta_j S_{T+h}^j + \varepsilon_{T+h}.$$

La meilleure prévision au sens de l'erreur quadratique moyenne est l'espérance de X_{T+h} notée $X_T(h)$ avec

$$X_T(h) = \mathbb{E}(X_{T+h}) = \sum_{i=1}^n \alpha_i T_{T+h}^i + \sum_{j=1}^p \beta_j S_{T+h}^j,$$

dont l'estimateur $\hat{X}_T(h)$ est

$$\hat{X}_T(h) = \hat{X}_{T+h} = \sum_{i=1}^n \hat{\alpha}_i T_{T+h}^i + \sum_{j=1}^p \hat{\beta}_j S_{T+h}^j.$$

La notation $\hat{X}_T(h)$ est utilisé pour signifier que l'on prédit à l'horizon h connaissant les données jusqu'à T . On peut également estimer des intervalles de confiance (asymptotiques).

Exemple

Considérons le nombre de passagers aériens. La croissance du nombre de passagers aérien est exponentielle. On effectue donc une transformation logarithmique sur les données. Et on considère le modèle suivant :

$$\forall t \in \{1, \dots, 144\} : X_t = at + b + \sum_{j=1}^{12} c_j S_t^j + \varepsilon_t$$

où on a tenu compte d'une saisonnalité annuelle :

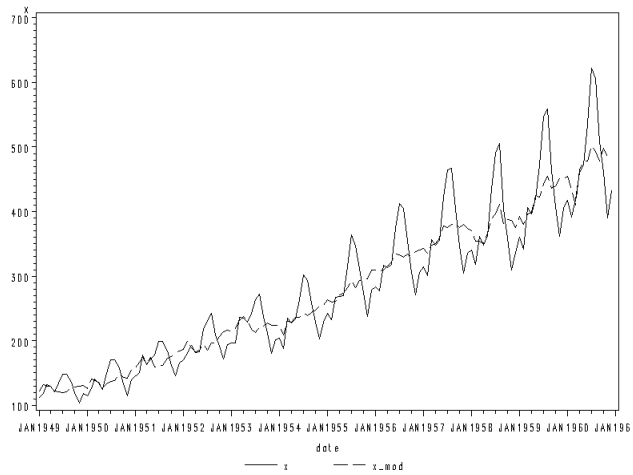
$$S_t^j = \begin{cases} 1 & \text{si } t = j [12], \\ 0 & \text{sinon.} \end{cases}$$

Ce modèle est l'analogue du premier modèle historique de Buys-Ballot (1847).

Comme $\sum_{j=1}^{12} S_t^j = 1$, le modèle indéterminé. Pour résoudre ce problème on peut soit poser

$c_{12} = 0$ soit contraindre $\sum_{j=1}^{12} c_j = 0$.

Le graphique suivant représente les résultats obtenus : série brute et série corrigée des variations saisonnières (on a effectué une transformation logarithmique aux données avant traitement, puis une transformation exponentielle après traitement).



Remarque I.27. Il est aussi possible d'utiliser des moyennes mobiles (ou glissantes) pour effectuer les décompositions saisonnières. Une moyenne mobile est une combinaison linéaire d'opérateurs retard :

$$M = \sum_{i=-m_1}^{m_2} \theta_i B^{-i}$$

où $(m_1, m_2) \in \mathbb{N}^2$, $(\theta_{-m_1}, \theta_{m_2}) \in \mathbb{R}^{*2}$ et B est l'opérateur retard : $BX_t = X_{t-1}$ et plus généralement $B^i X_t = X_{t-i}$. Un algorithme classique de désaisonnalisation utilisant les moyennes mobiles est l'algorithme X11 mis au point par le "Bureau of Census". \diamond

I.4 Prédiction et processus des innovations

On considère un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$. Soit t fixé. La prédiction linéaire optimale de X_t sachant son passé est définie comme la régression linéaire de X_t sur l'espace fermé engendré par les combinaisons linéaires des $(X_i)_{i \leq t-1}$. On la note \hat{X}_t ,

$$\hat{X}_t = c_0 + \sum_{i=-\infty}^{t-1} a_i X_i, \quad (\text{I.2})$$

où $(c_0, (a_i)_{1 \leq i \leq t-1}) = \operatorname{argmin}_{c'_0, (a'_i)_{1 \leq i \leq t-1}} \mathbb{E}[(X_t - c'_0 + \sum_{i=-\infty}^{t-1} a'_i X_i)^2]$. Les erreurs de prévision successives $\varepsilon_t = X_t - \hat{X}_t$ forment un processus appelé **processus des innovations**.

Proposition I.28. *Le processus des innovations, ou plus simplement innovations, est un bruit blanc.*

Remarque I.29. On utilise souvent des algorithmes itératifs afin de déterminer les prévisions basées sur (X_1, \dots, X_{T+1}) , à partir de celles basées sur (X_1, \dots, X_T) . L'algorithme de Durbin-Levinson et l'algorithme des innovations sont couramment utilisés. \diamond

I.5 Étude des processus AR

Cette partie ainsi que les suivantes présentent les modèles proposés par Box et Jenkins. Ils sont utilisés pour traiter de nombreuses séries temporelles.

Sauf mention contraire, on considère dans ce chapitre des processus centrés. Si le processus $(X_t)_{t \in \mathbb{Z}}$ n'est pas centré, on obtient les mêmes résultats mais sur le processus $(Y_t)_{t \in \mathbb{Z}}$ tel que $Y_t = X_t - \mu_X$.

I.5.1 Définition

Définition I.30. Soit $(\varepsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc (faible) de variance σ^2 et $p \geq 1$.

On dit qu'un processus $(X_t)_{t \in \mathbb{Z}}$ est un processus autorégressif ou encore **processus AR** (AutoRegressive) d'ordre p , noté $AR(p)$, si :

- $(X_t)_{t \in \mathbb{Z}}$ est stationnaire,
- $\forall t \in \mathbb{Z} : X_t = \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t$ où $(\varphi_1, \dots, \varphi_p) \in \mathbb{R}^p$ sont des constantes et $\varphi_p \neq 0$.

On utilise généralement la notation suivante :

$$\Phi(B) X_t = \varepsilon_t, \tag{I.3}$$

$$\text{où } \Phi(B) = I - \sum_{i=1}^p \varphi_i B^i.$$

Exemple I.31. Soit $(\varepsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc faible de variance σ^2 . Soit le processus $(X_t)_{t \in \mathbb{Z}}$ suivant :

$$X_t = X_{t-1} + \varepsilon_t.$$

On a, pour $h \in \mathbb{N}^*$: $X_t - X_{t-h} = \varepsilon_t + \dots + \varepsilon_{t-h+1}$, d'où :

$$\mathbb{E}[(X_t - X_{t-h})^2] = h\sigma^2$$

Si le processus $(X_t)_{t \in \mathbb{Z}}$ était stationnaire, on aurait :

$$\begin{aligned} \mathbb{E}[(X_t - X_{t-h})^2] &= \mathbb{E}(X_t^2) + \mathbb{E}(X_{t-h}^2) - 2\mathbb{E}(X_t X_{t-h}) \\ &= \mathbb{E}(X_t^2) + \mathbb{E}(X_{t-h}^2) + \mathbb{E}(X_t^2) + \mathbb{E}(X_{t-h}^2) - \mathbb{E}[(X_t + X_{t-h})^2] \\ &= 4\mathbb{V}(X_t) - \mathbb{E}[(X_t + X_{t-h})^2] \\ &\leq 4\mathbb{V}(X_t). \end{aligned}$$

Or il est impossible d'avoir, pour tout $h \in \mathbb{N}^* : h\sigma^2 \leq 4\mathbb{V}(X_t)$. Le processus $(X_t)_{t \in \mathbb{Z}}$ n'est donc pas stationnaire. \diamond

Dans l'exemple précédent, on a $\Phi(x) = 1 - x$, et Φ possède une racine (ici égale à 1) de module égal à 1. Plus généralement, on peut montrer que si $\Phi(B)$ a une racine de module égal à 1 (sur le cercle unité) alors le processus $(X_t)_{t \in \mathbb{Z}}$ n'est pas stationnaire, et il ne peut donc pas s'agir d'un processus AR.

I.5.2 Processus AR canonique et écriture $MA(\infty)$

Définition I.32. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus $AR(p)$ associé à Φ par I.3. Si les racines de Φ sont toutes à l'extérieur du cercle unité alors on a sa représentation canonique.

Proposition I.33. Si X est sous sa forme canonique, alors le bruit blanc associé est le processus des innovations : $\varepsilon_t = X_t - \hat{X}_t$ où \hat{X}_t est donné par (I.2).

Si le polynôme $\Phi(B)$, que l'on suppose inversible, a des racines à l'intérieur du cercle unité, alors le processus $(\varepsilon_t)_{t \in \mathbb{Z}}$ n'est pas le processus des innovations. Pour se ramener à une écriture canonique (c'est à dire une écriture où les résidus sont l'innovation) on transforme l'opérateur $\Phi(B)$.

Soient $z_i = \frac{1}{\lambda_i}$, $i \in \{1, \dots, p\}$, les racines de $\Phi(z)$.

Supposons que les $(z_i)_{i \in \{1, \dots, r\}}$ soient toutes de module inférieur à 1 (à l'extérieur du cercle unité) et que les $(z_i)_{i \in \{r+1, \dots, p\}}$ soient toutes de module supérieur à 1 (à l'intérieur du cercle unité). On peut écrire :

$$\Phi(B) = \prod_{i=1}^p (I - \lambda_i B)$$

On construit alors un nouveau polynôme $\Phi^*(B)$ qui a toutes ses racines à l'extérieur du cercle unité :

$$\Phi^*(B) = \prod_{i=1}^r \left(I - \frac{1}{\lambda_i} B \right) \prod_{i=r+1}^p (I - \lambda_i B),$$

et on obtient l'écriture canonique du processus.

Proposition I.34. On définit le processus $(\eta_t)_{t \in \mathbb{Z}}$ par $\eta_t = \Phi^*(B)X_t$. Alors $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc faible et c'est le processus des innovations du processus $AR(X_t)_{t \in \mathbb{Z}}$.

On préfère travailler avec la représentation canonique car la représentation inverse ne fera apparaître que les instants passés du processus. Dans la suite, on considère des processus AR canoniques.

Proposition I.35. Soit $(\varepsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc faible de variance σ^2 .

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus $AR(p)$ canonique : $\Phi(B)X_t = \varepsilon_t$. Il admet alors une écriture $MA(\infty)$:

$$X_t = \Phi^{-1}(B)\varepsilon_t = \varepsilon_t + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t-i},$$

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

I.5.3 Autocorrélations simples d'un processus AR

Soit $(\varepsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc faible de variance σ^2 et $(X_t)_{t \in \mathbb{Z}}$ un processus AR(p) canonique :
 $X_t = \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t$ pour $t \in \mathbb{Z}$ et $\varphi = (\varphi_1, \dots, \varphi_p) \in \mathbb{R}^p$.

Lemma I.36. On a $\gamma(0) = \frac{\sigma^2}{1 - \sum_{i=1}^p \varphi_i \rho(i)}$ et $\rho(h) = \sum_{i=1}^p \varphi_i \rho(h-i)$ pour tout $h \in \mathbb{N}^*$.

Démonstration. On a :

$$\begin{aligned}
 \gamma(0) &= \mathbb{Cov}(X_t, X_t) \\
 &= \mathbb{Cov}\left(X_t, \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t\right) \\
 &= \sum_{i=1}^p \varphi_i \mathbb{Cov}(X_t, X_{t-i}) + \mathbb{Cov}(X_t, \varepsilon_t) \\
 &= \sum_{i=1}^p \varphi_i \gamma(i) + \mathbb{Cov}\left(\sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t, \varepsilon_t\right) \\
 &= \sum_{i=1}^p \varphi_i \gamma(i) + \mathbb{Cov}(\varepsilon_t, \varepsilon_t) \text{ car } \varepsilon_t \perp \mathcal{H}_{-\infty}^{t-1}(X) \\
 &= \sum_{i=1}^p \varphi_i \gamma(0) \rho(i) + \sigma^2.
 \end{aligned}$$

On a de même :

$$\begin{aligned}
 \forall h \in \mathbb{N}^* : \gamma(h) &= \mathbb{Cov}(X_t, X_{t-h}) \\
 &= \mathbb{Cov}\left(\sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t, X_{t-h}\right) \\
 &= \sum_{i=1}^p \varphi_i \mathbb{Cov}(X_{t-i}, X_{t-h}) + \mathbb{Cov}(\varepsilon_t, X_{t-h}) \\
 &= \sum_{i=1}^p \varphi_i \gamma(h-i) \text{ car } \varepsilon_t \perp \mathcal{H}_{-\infty}^{t-1}(X).
 \end{aligned}$$

□

On obtient finalement le système suivant, appelé équations de Yule-Walker :

$$\left\{ \begin{array}{l} \gamma(0) = \frac{\sigma^2}{1 - \sum_{i=1}^p \varphi_i \rho(i)} \\ \begin{pmatrix} \rho(1) \\ \rho(2) \\ \dots \\ \rho(p) \end{pmatrix} = \begin{bmatrix} 1 & \rho(1) & \dots & \rho(p-1) \\ \rho(1) & 1 & \dots & \dots \\ \dots & \dots & \dots & \rho(1) \dots \\ \rho(p-1) & \dots & \rho(1) & 1 \end{bmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \dots \\ \varphi_p \end{pmatrix} \end{array} \right. = R_p \varphi.$$

Les autocorrélations simples sont solution d'une équation de récurrence linéaire simple d'ordre p .

Remarque I.37. Si les racines de $\Phi(z) : z_i = \frac{1}{\lambda_i}$, $i \in \{1, \dots, p\}$, sont réelles et distinctes alors on obtient :

$$\rho(h) = \sum_{i=1}^p c_i \lambda_i^h$$

Comme les racines sont de modules strictement plus grand que 1, les autocorrélations simples décroissent exponentiellement vite vers 0.

Dans le cas général, on obtient une décroissance des autocorrélations simples vers 0, de type exponentiel ou sinusoidal amorti. \diamond

I.5.4 Autocorrélations partielle d'un processus AR

Si $(X_t)_{t \in \mathbb{Z}}$ est un processus $AR(p)$ alors ses autocorrélations partielles s'annulent à partir du rang $p+1$: $r(p) \neq 0$ et pour $h \geq p+1$ on a $r(h) = 0$.

Réciproquement, il s'agit d'une condition nécessaire et suffisante pour qu'un processus $(X_t)_{t \in \mathbb{Z}}$ soit un $AR(p)$.

De plus, si le processus est sous sa forme canonique, on a $r(p) = \varphi_p$. On ne peut rien dire a priori sur $r(h)$ pour $h \in \{1, \dots, p-1\}$ (on sait simplement que $r(1) = \rho(1)$, ce qui est vrai pour tous les processus stationnaires).

I.5.5 Exemples

On présente dans les figures I.1, I.2 et I.3 les autocorrélogrammes de trois processus $AR(1)$ pour lesquels on a simulé 5000 valeurs.

		Size of Working Series		-0.0312
		Standard Deviation		1.04332
		Number of Observations		500
Autocorrelations				
Lag	Covariance	Correlation	-1 0 1 2 3 4 5 6 7 8 9	Std Error
0	1.00004	1.00000		0
1	0.31871	0.28959		0.01594
2	0.10432	0.09762		0.01280
3	-0.010826	-0.00985		0.01043
4	-0.00799	-0.0092		0.01043
5	-0.001948	-0.0023		0.01033
6	-0.018714	-0.0193		0.01034
7	0.011328	0.01059		0.01035
8	0.02401	0.02465		0.01032
9	0.027611	0.02844		0.01032
10	0.020131	0.02074		0.01027
11	-0.015114	-0.01445		0.01027
12	-0.04557	-0.0459		0.01027
13	-0.01284	-0.0118		0.01027
14	-0.01570	-0.01525		0.01038
15	-0.02021	-0.0222		0.01061
16	-0.027475	-0.0289		0.01041
17	0.010386	0.00945		0.01018
18	-0.001280	-0.0014		0.01020
19	0.009716	0.00834		0.01020
20	-0.01281	-0.0025		0.01021
21	-0.008720	-0.00789		0.01022
22	0.004484	0.00437		0.01022
23	0.001897	0.00192		0.01022
24	-0.007717	-0.00707		0.01022
Partial Autocorrelations				
Lag	Correlation	-1 0 1 2 3 4 5 6 7 8 9		
1	0.28959			
2	-0.01797			
3	-0.02057			
4	-0.00825			
5	0.00125			
6	-0.01621			
7	0.02094			
8	0.01359			
9	0.01452			
10	-0.01019			
11	-0.01243			
12	-0.02654			
13	0.01732			
14	-0.01057			
15	-0.02395			
16	-0.01359			
17	0.02440			
18	-0.01352			
19	0.00952			
20	-0.01218			
21	-0.00119			
22	0.00795			
23	0.01970			
24	-0.02209			

FIG. I.1 – $X_t = 0.3X_{t-1} + \varepsilon_t$

		Len of Working Series		-0.01492										
		Standard Deviation		1.90122										
		Number of Observations		5000										
Autocorrelations														
Lag	Covariance	Correlation	-1	9	8	7	6	5	4	3	2	1	0	Std Error
0	1.240885	1.00000												
1	-0.075026	-0.01502												0.01442
2	0.024444	0.04820												0.01742
3	-0.052303	-0.12850												0.01829
4	0.007596	0.04955												0.01842
5	-0.017144	-0.02449												0.01870
6	-0.017719	-0.0224												0.01872
7	0.020523	0.01825												0.01874
8	-0.010542	-0.0815												0.01876
9	0.022422	0.04170												0.01877
10	-0.025860	-0.0321												0.01883
11	0.011305	0.02265												0.01887
12	-0.061652	-0.0451												0.01893
13	0.030490	0.02602												0.01897
14	-0.024525	-0.01920												0.01925
15	0.011354	0.00844												0.01929
16	-0.042359	-0.0214												0.01930
17	0.045817	0.03009												0.01941
18	-0.024735	-0.02559												0.01955
19	0.024407	0.02560												0.01954
20	-0.026102	-0.01445												0.019571
21	0.010274	0.00762												0.019576
22	-0.010467	-0.01445												0.019576
23	0.020212	0.02445												0.019577
24	-0.010216	-0.00765												0.019583

*** marks two standard errors

		Partial Autocorrelations									
Lag	Correlation	-1	9	8	7	6	5	4	3	2	1
1	-0.01502										
2	-0.04949										
3	-0.01752										
4	-0.02012										
5	0.00809										
6	-0.01999										
7	-0.00212										
8	0.00431										
9	0.02655										
10	0.00224										
11	0.01470										
12	-0.03235										
13	-0.01051										
14	-0.00110										
15	-0.00421										
16	-0.00465										
17	0.00755										
18	-0.00425										
19	0.00559										
20	-0.00165										
21	-0.00481										
22	-0.01082										
23	0.01911										
24	0.01459										

FIG. I.2 – $X_t = -0.5X_{t-1} + \varepsilon_t$

		Len of Working Series		-0.11152										
		Standard Deviation		1.61606										
		Number of Observations		5000										
		Autocorrelations												
Lag	Covariance	Correlation	-1	9	8	7	6	5	4	3	2	1	0	Std Error
0	2.059751	1.00000												0
1	2.045776	0.70359												0.01442
2	1.059658	0.45811												0.02114
3	1.227388	0.43020												0.024076
4	0.361324	0.20235												0.025128
5	0.354065	0.23078												0.027447
6	0.020941	0.23002												0.027771
7	0.497694	0.19070												0.028152
8	0.420516	0.15462												0.028610
9	0.212416	0.13009												0.028777
10	0.211209	0.09527												0.028776
11	0.020247	0.04608												0.028724
12	0.040889	0.01657												0.028729
13	0.010277	0.00611												0.028740
14	-0.015285	-0.00586												0.028741
15	-0.040287	-0.01544												0.028741
16	-0.042711	-0.01237												0.028742
17	-0.017827	-0.00609												0.028744
18	-0.017360	-0.00655												0.028745
19	-0.010780	-0.00260												0.028745
20	-0.017076	-0.00654												0.028745
21	-0.011880	-0.00465												0.028746
22	0.02057493	0.00100												0.028746
23	0.010542	0.00115												0.028746
24	-0.0080125	-0.00031												0.028746

*** marks two standard errors

		Partial Autocorrelations									
Lag	Correlation	-1	9	8	7	6	5	4	3	2	1
1	0.70359										
2	-0.01521										
3	-0.00755										
4	0.01085										
5	0.01070										
6	-0.00494										
7	0.02851										
8	-0.00650										
9	-0.01225										
10	-0.03095										
11	-0.01751										
12	-0.01825										
13	0.02811										
14	-0.01701										
15	-0.00705										
16	0.01127										
17	0.02547										
18	-0.01722										
19	0.00996										
20	-0.01005										
21	0.00829										
22	0.00825										
23	0.00450										
24	-0.02850										

FIG. I.3 – $X_t = 0.8X_{t-1} + \varepsilon_t$

I.6 Processus MA

Définition I.38. Soit $(\varepsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc faible de variance σ^2 .

On dit qu'un processus $(X_t)_{t \in \mathbb{Z}}$ est un processus **processus MA** (Moving Average) ou

encore à moyenne mobile d'ordre q , noté $MA(q)$, si :

$$\forall t \in \mathbb{Z} : X_t = \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}$$

où $(\theta_1, \dots, \theta_q) \in \mathbb{R}^q$ et $\theta_q \neq 0$.

On utilise généralement la notation suivante : $X_t = \Theta(B) \varepsilon_t$, où $\Theta(B) = I + \sum_{i=1}^q \theta_i B^i$.

Un processus MA est toujours stationnaire.

I.6.1 Processus MA canonique et écriture $AR(\infty)$

Si les racines de Θ sont toutes à l'extérieur du cercle unité alors on a sa représentation **canonique**. Le bruit blanc associé est alors l'innovation.

On préfère travailler avec la représentation canonique car la représentation inverse ne fera apparaître que les instants passés du bruit blanc. On peut passer d'une représentation non canonique à une représentation canonique en inversant les racines qui sont à l'intérieur du cercle unité (cf. la transformation similaire effectuée sur les processus AR). Dans la suite, on considère des processus MA canoniques.

Proposition I.39. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus $MA(q)$. Si X est sous forme canonique, alors il admet alors une écriture $AR(\infty)$:

$$\varepsilon_t = \Theta^{-1}(B) X_t = X_t + \sum_{i=1}^{+\infty} \pi_i X_{t-i} \quad \text{c'est-à-dire} \quad X_t = -\sum_{i=1}^{+\infty} \pi_i X_{t-i} + \varepsilon_t,$$

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

I.6.2 Autocorrélations simples d'un processus MA

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus $MA(q)$ canonique : pour tout $t \in \mathbb{Z}$, $X_t = \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}$.

Comme $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc, on a :

$$\gamma(0) = \mathbb{V} \left(\varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \right) = \sigma^2 \left(1 + \sum_{i=1}^q \theta_i^2 \right).$$

On a de même :

$$\begin{aligned} \forall h \in \mathbb{N}^* : \gamma(h) &= \text{Cov}(X_t, X_{t-h}) \\ &= \text{Cov} \left(\varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}, \varepsilon_{t-h} + \sum_{j=1}^q \theta_j \varepsilon_{t-h-j} \right) \\ &= \begin{cases} \theta_h + \sum_{i=h+1}^q \theta_i \theta_{i-h} & \text{si } h \in \{1, \dots, q\}, \\ 0 & \text{si } h > q. \end{cases} \end{aligned}$$

Proposition I.40. Si $(X_t)_{t \in \mathbb{Z}}$ est un processus $MA(q)$ alors ses autocorrélations simples s'annulent à partir du rang $q + 1$:

$$\begin{cases} \rho(q) \neq 0, \\ \forall h \geq q + 1 : \rho(h) = 0. \end{cases}$$

Il s'agit d'une condition nécessaire et suffisante pour qu'un processus $(X_t)_{t \in \mathbb{Z}}$ soit un $MA(q)$.

I.6.3 Autocorrélations partielles d'un processus MA

Les autocorrélations partielles sont solution d'une équation de récurrence linéaire simple d'ordre q . Elles décroissent vers 0 de manière exponentielle ou sinusoïdale amortie.

Exemples

On présente dans les figures I.4, I.5 et I.6 les autocorrélogrammes de trois processus $MA(1)$ pour lesquels on a simulé 5000 valeurs.

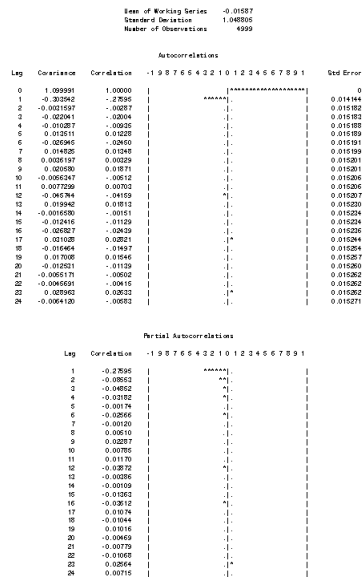


FIG. I.4 – $X_t = \varepsilon_t - 0.3\varepsilon_{t-1}$

			Size of Working Series	-0.02351	
			Standard Deviation	1.119701	
			Number of Observations	4999	
			Autocorrelations		
Lag	Covariance	Correlation	-1	0	Std Error
0	1.253700	1.00000		*****	0
1	0.491909	0.39209		*****	0.014144
2	-0.001783	-0.00145		*****	0.016179
3	-0.042896	-0.03211		*****	0.016182
4	-0.022946	-0.01793		*****	0.016201
5	-0.016422	-0.01211		*****	0.016210
6	-0.019561	-0.01557		*****	0.016212
7	0.010339	0.00825		*****	0.016215
8	0.022778	0.02014		*****	0.016216
9	0.041168	0.03720		*****	0.016224
10	0.008960	0.00707		*****	0.016232
11	-0.005448	-0.00110		*****	0.016234
12	-0.011670	-0.01112		*****	0.016239
13	-0.017166	-0.01304		*****	0.016250
14	-0.013308	-0.01112		*****	0.016262
15	-0.009460	-0.00744		*****	0.016264
16	-0.003224	-0.00111		*****	0.016276
17	0.004479	0.00712		*****	0.016282
18	0.0172912	0.00552		*****	0.016294
19	0.001672	0.00412		*****	0.016294
20	-0.010521	-0.00605		*****	0.016294
21	-0.014946	-0.01160		*****	0.016296
22	0.0002520	0.00791		*****	0.016277
23	0.029123	0.02244		*****	0.016299
24	-0.005522	-0.0052		*****	0.016294

*** means two standard errors

FIG. I.5 - $X_t = \varepsilon_t + 0.5\varepsilon_{t-1}$

		Size of Working Series	-0.00496	
		Standard Deviation	1.207164	
		Number of Observations	4999	
		Autocorrelations		
Lag	Covariance	Correlation	-1 0.975 6 6 4 2 1 0 1 2 2 4 6 6 7 8 9 1	Std Error
0	1.03285	1.00000		0
1	-0.02465	-0.00451	*****	0.014144
2	0.0087175	0.00465	*****	0.017945
3	-0.02280	-0.01399	*****	0.017945
4	-0.0063284	-0.00321	*****	0.017948
5	0.023965	0.02017	*****	0.017948
6	-0.00152	-0.00122	*****	0.017952
7	0.00156	0.00119	*****	0.017951
8	-0.001118	-0.00079	*****	0.017964
9	0.020719	0.01811	*****	0.017964
10	-0.014627	-0.00895	*****	0.017967
11	0.002205	0.00101	*****	0.017969
12	-0.004651	-0.00190	*****	0.017171
13	0.04680	0.02815	*****	0.017931
14	0.002962	0.00195	*****	0.017991
15	-0.004968	-0.00512	*****	0.017201
16	-0.00281	-0.00229	*****	0.017201
17	0.077960	0.04665	*****	0.017206
18	-0.025403	-0.01325	*****	0.017222
19	0.011204	0.01037	*****	0.017227
20	-0.019999	-0.01195	*****	0.017221
21	-0.0049375	-0.00200	*****	0.017233
22	-0.012943	-0.00762	*****	0.017232
23	0.046817	0.02760	*****	0.017234
24	-0.001982	-0.00552	*****	0.017242

*** zero bin standard errors

FIG. I.6 - $X_t = \varepsilon_t - 0.8\varepsilon_{t-1}$

I.7 Processus ARMA

Soit $(\varepsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc faible de variance σ^2 .

Définition I.41. On dit qu'un processus $(X_t)_{t \in \mathbb{Z}}$ est un **processus ARMA** (AutoRegressive Moving Average) d'ordre (p, q) , noté $ARMA(p, q)$, si :

- $(X_t)_{t \in \mathbb{Z}}$ est stationnaire.
- il existe $\varphi = (\varphi_1, \dots, \varphi_p) \in \mathbb{R}^p$ avec $\varphi_p \neq 0$ et $\theta = (\theta_1, \dots, \theta_q) \in \mathbb{R}^q$ avec $\theta_q \neq 0$ tel que pour tout $t \in \mathbb{Z}$,

$$X_t - \sum_{i=1}^p \varphi_i X_{t-i} = \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}.$$

On utilise généralement la notation suivante :

$$\Phi(B) X_t = \Theta(B) \varepsilon_t$$

où $\Phi(B) = I - \sum_{i=1}^p \varphi_i B^i$ et $\Theta(B) = I + \sum_{i=1}^q \theta_i B^i$.

Un processus $AR(p)$ est un processus $ARMA(p, 0)$; un processus $MA(q)$ est un processus $ARMA(0, q)$.

I.7.1 Processus ARMA canonique et écritures $MA(\infty)$ et $AR(\infty)$

Définition I.42. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus $ARMA(p, q)$:

$$\Phi(B) X_t = \Theta(B) \varepsilon_t$$

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**² si Θ a toutes ses racines à l'extérieur du cercle unité.
- **canonique** si la représentation est causale et inversible.

Proposition I.43. Le bruit blanc associé à un processus $ARMA$ sous sa forme canonique est le processus d'innovation.

Proposition I.44. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus $ARMA(p, q)$ canonique : $\Phi(B) X_t = \Theta(B) \varepsilon_t$. Si de plus il est minimal alors :

1. Il admet une **écriture** $MA(\infty)$:

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

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 :

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

²Le terme "inversible" est un point de terminologie ici. Pour un processus $ARMA$, Φ est toujours inversible (au sens stationnaire) et Θ est inversible (au sens stationnaire) si aucune de ses racines n'est sur le cercle unité.

2. Il admet une écriture AR(∞) :

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

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

En posant $\pi_i = 0$ pour $i < 0$, $\varphi_0 = -1$ et $\varphi_i = 0$ pour $i > p$, on a :

$$\forall i \in \mathbb{N} : \pi_i + \sum_{j=1}^q \theta_j \pi_{i-j} = -\varphi_i$$

I.7.2 Autocorrélations d'un processus ARMA

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

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

$$\begin{aligned} \gamma(h) &= \text{Cov}(X_t, X_{t-h}) \\ &= \text{Cov}\left(\varepsilon_t + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t-i}, \varepsilon_{t-h} + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t-h-i}\right) \\ &= \sigma^2 \sum_{i=0}^{+\infty} \psi_i \psi_{i+h} \text{ où } \psi_0 = 1 \end{aligned}$$

– On a :

$$X_t - \varphi_1 X_{t-1} - \dots - \varphi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

D'où :

$$\begin{aligned} \gamma(h) - \varphi_1 \gamma(h-1) - \dots - \varphi_p \gamma(h-p) &= \text{Cov}(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, X_{t-h}) \\ &= \text{Cov}\left(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \sum_{i=0}^{+\infty} \psi_i \varepsilon_{t-h-i}\right) \\ &= \begin{cases} \sigma^2 \sum_{i=0}^{+\infty} \theta_{h+i} \psi_i & \text{si } h \in \{0, \dots, q\} \\ 0 & \text{si } h > q \end{cases} \end{aligned}$$

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.

Il existe des propriétés similaires sur les autocorrélations partielles.

Méthode du coin

Il n'existe pas de caractérisation simple des processus ARMA, basée sur les autocorrélations simples ou partielles. La méthode du coin repose sur des propriétés des matrices d'autocorrélation.

Soient les matrices suivantes et leurs déterminants respectifs :

$$\Omega_{i,j} = \begin{bmatrix} \rho(i) & \rho(i-1) & \dots & \rho(i-j+1) \\ \rho(i-1) & \rho(i) & \dots & \dots \\ \dots & \dots & \dots & \rho(i-1) \\ \rho(i-j+1) & \dots & \rho(i-1) & \rho(i) \end{bmatrix}$$

$$\Delta_{i,j} = \det(\Omega_{i,j})$$

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus $ARMA(p, q)$ canonique minimal :

$$\Phi(B) X_t = \Theta(B) \varepsilon_t$$

On a alors :

- $\forall (i, j) / i > q, j > p : \Delta_{i,j} = 0$
- $\forall (i, j) / i \leq q : \Delta_{i,p} \neq 0$
- $\forall (i, j) / j \leq p : \Delta_{q,j} \neq 0$

On peut visualiser ce résultat sous forme matricielle en représentant la matrice $M = (\Delta_{i,j})_{(i,j) \in \{1, \dots, k\}^2}$ (pour k assez grand) et faire ainsi apparaître un coin :

$$M = \begin{bmatrix} \Delta_{1,1} & \dots & \Delta_{1,p} & \Delta_{1,p+1} & \dots & \Delta_{1,k} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \Delta_{q,1} & \dots & \Delta_{q,p} & \Delta_{q,p+1} & \dots & \Delta_{q,k} \\ \Delta_{q+1,1} & \dots & \Delta_{q+1,p} & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \Delta_{k,1} & \dots & \Delta_{k,p} & \dots & \dots & \dots \end{bmatrix}$$

0

I.7.3 Densité spectrale d'un processus ARMA

Proposition I.45. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus $ARMA(p, q)$ (pas forcément sous sa forme canonique), $\Phi(B) X_t = \Theta(B) \varepsilon_t$. Sa densité spectrale vaut :

$$f(\omega) = \frac{\sigma^2 |\Theta(e^{-i\omega})|^2}{2\pi |\Phi(e^{-i\omega})|^2}$$

I.7.4 Estimation des processus ARMA

Soit $(\varepsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc faible de variance σ^2 et $(X_t)_{t \in \mathbb{Z}}$ un processus $ARMA(p, q)$ canonique minimal : $\Phi(B) X_t = \Theta(B) \varepsilon_t$.

Principe

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 autocorrélations. En général, ces estimateurs ne sont pas efficaces³. C'est pourquoi, on utilise des estimations préliminaires comme première étape dans des méthodes itératives, du type maximum de vraisemblance ou moindres carrés.

Estimation préliminaire

Cas des processus AR

³Un estimateur efficace est un estimateur sans biais atteignant la borne FDCR (donc de variance minimale).

Dans le cas d'un processus $AR(p)$ canonique, on peut utiliser les équations de Yule-Walker :

$$\left\{ \begin{array}{l} \gamma(0) = \frac{\sigma^2}{1 - \sum_{i=1}^p \varphi_i \rho(i)} \\ \begin{pmatrix} \rho(1) \\ \rho(2) \\ \dots \\ \rho(p) \end{pmatrix} = \begin{bmatrix} 1 & \rho(1) & \dots & \rho(p-1) \\ \rho(1) & 1 & \dots & \dots \\ \dots & \dots & \dots & \rho(1) \dots \\ \rho(p-1) & \dots & \rho(1) & 1 \end{bmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \dots \\ \varphi_p \end{pmatrix} \end{array} \right. = R_p \varphi.$$

On en déduit :

$$\left\{ \begin{array}{l} \hat{\sigma}^2 = \hat{\gamma}(0) \left(1 - \sum_{i=1}^p \hat{\varphi}_i \hat{\rho}(i) \right) \\ \begin{pmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 \\ \dots \\ \hat{\varphi}_p \end{pmatrix} = \begin{bmatrix} 1 & \hat{\rho}(1) & \dots & \hat{\rho}(p-1) \\ \hat{\rho}(1) & 1 & \dots & \dots \\ \dots & \dots & \dots & \hat{\rho}(1) \dots \\ \hat{\rho}(p-1) & \dots & \hat{\rho}(1) & 1 \end{bmatrix}^{-1} \begin{pmatrix} \hat{\rho}(1) \\ \hat{\rho}(2) \\ \dots \\ \hat{\rho}(p) \end{pmatrix} \end{array} \right.$$

En notant $\varphi = (\varphi_1, \dots, \varphi_p)$ et $\hat{\varphi} = (\hat{\varphi}_1, \dots, \hat{\varphi}_p)$, on a les résultats suivants :

$$\begin{aligned} \sqrt{n}(\hat{\varphi} - \varphi) &\xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2 \Sigma_p^{-1}) \\ \hat{\sigma}^2 &\xrightarrow{\mathbb{P}} \sigma^2 \end{aligned}$$

où :

$$\Sigma_p = \begin{bmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \dots & \dots \\ \dots & \dots & \dots & \gamma(1) \dots \\ \gamma(p-1) & \dots & \gamma(1) & \gamma(0) \end{bmatrix} = \gamma(0) R_h.$$

Cas des processus MA et ARMA

Si on a une partie moyenne mobile, on utilise plutôt 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 :

$$X_t = \varepsilon_t + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t-i}$$

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 :

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

On utilise ces relations pour obtenir une première estimation de $(\varphi_1, \dots, \varphi_p)$ et $(\theta_1, \dots, \theta_q)$ à partir de $(\hat{\psi}_1(n), \dots, \hat{\psi}_{p+q}(n))$.

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 $\mathcal{N}(0, \sigma^2)$, 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_T) , la vraisemblance s'écrit :

$$V(x_1, \dots, x_T; \varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \sigma^2) = \frac{1}{(2\pi)^{\frac{T}{2}}} \frac{1}{\sqrt{\det \Sigma_T}} \exp\left(-\frac{1}{2}x' \Sigma_T^{-1} x\right)$$

où $x = (x_1, \dots, x_T)$ et Σ_T est la matrice de variance-covariance de (X_1, \dots, X_T) .

Remarquons que (X_1, \dots, X_T) est un vecteur gaussien, en tant que transformation linéaire du vecteur gaussien $(\varepsilon_1, \dots, \varepsilon_T)$.

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

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

On peut utiliser l'algorithme des innovations pour calculer les erreurs de prévision au pas 1, $\varepsilon_i = X_i - \hat{X}_i$, ainsi que leur variance v_i . On évite ainsi le calcul direct de Σ_T^{-1} et de $\det \Sigma_T$. On a :

$$\begin{pmatrix} X_1 \\ X_2 \\ \dots \\ X_T \end{pmatrix} = C_T \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_T \end{pmatrix}$$

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

$$C_T = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \\ \theta_1(1) & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \theta_{T-1}(T-1) & \theta_{T-2}(T-1) & \dots & \theta_1(T-1) & 1 \end{bmatrix}$$

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

$$V_T = \begin{bmatrix} v_0 & 0 & \dots & 0 \\ 0 & v_1 & \dots & \dots \\ \dots & \dots & \dots & 0 \\ 0 & \dots & 0 & v_{T-1} \end{bmatrix}$$

On a $\Sigma_T = C_T V_T C_T'$ et $\det \Sigma_T = (\det C_T)^2 \det V_T = v_0 \dots v_{T-1}$, et donc

$$x' \Sigma_T^{-1} x = \sum_{i=1}^T \frac{(x_i - \hat{x}_i)^2}{v_{i-1}},$$

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

La vraisemblance s'écrit donc :

$$V(x_1, \dots, x_T; \varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \sigma^2) = \frac{1}{(2\pi)^{\frac{T}{2}}} \frac{1}{\sqrt{v_0 \dots v_{T-1}}} \exp \left(-\frac{1}{2} \sum_{i=1}^T \frac{(x_i - \hat{x}_i)^2}{v_{i-1}} \right)$$

L'algorithme des innovations nous indique que :

$$\hat{X}_{i+1} = \begin{cases} \sum_{j=1}^i \theta_j(i) (X_{i+1-j} - \hat{X}_{i+1-j}) & \text{si } 1 \leq i \leq \max(p, q) \\ \varphi_1 X_i + \dots + \varphi_p X_{i+1-p} + \sum_{j=1}^q \theta_j(i) (X_{i+1-j} - \hat{X}_{i+1-j}) & \text{si } i > \max(p, q). \end{cases}$$

On définit $r_i = v_i / \sigma^2$. On peut réécrire la vraisemblance comme suit :

$$V(x_1, \dots, x_T; \varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{T}{2}}} \frac{1}{\sqrt{r_0 \dots r_{T-1}}} \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^T \frac{(x_i - \hat{x}_i)^2}{r_{i-1}} \right)$$

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

$$\begin{aligned} -(\hat{\varphi}_1, \dots, \hat{\varphi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q) &= \arg \min_{(\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q)} \left\{ \log \left[\frac{1}{T} S(\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q) \right] + \frac{1}{T} \sum_{i=1}^T \log r_{i-1} \right\} \\ -\hat{\sigma}^2 &= \frac{1}{T} S(\hat{\varphi}_1, \dots, \hat{\varphi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q) \text{ où } S(\hat{\varphi}_1, \dots, \hat{\varphi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q) = \sum_{i=1}^T \frac{(x_i - \hat{x}_i)^2}{r_{i-1}} \end{aligned}$$

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

Estimation par la méthode des moindres carrés

On cherche cette fois à minimiser la quantité suivante :

$$S(\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q) = \sum_{i=1}^T \frac{(x_i - \hat{x}_i)^2}{r_{i-1}}$$

L'estimateur des moindres carrés vérifie :

$$\begin{aligned} -(\hat{\varphi}_1, \dots, \hat{\varphi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q) &= \arg \min_{(\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q)} S(\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q) \\ -\hat{\sigma}^2 &= \frac{1}{T-p-q} S(\hat{\varphi}_1, \dots, \hat{\varphi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q) \end{aligned}$$

Les estimateurs obtenus sont efficaces.

Remarque I.46. 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. \diamond

I.7.5 Choix de modèle

Les critères de choix de modèles sont un compromis entre l'ajustement de la série modélisée (on cherche à minimiser la variance résiduelle non expliquée : $\hat{\sigma}^2$), et la règle de parcimonie (on cherche à minimiser le nombre de paramètre $p+q$), ces deux critères étant antagonistes. Ces critères sont établis en calculant une distance, appelée distance de Kullback, entre la loi inconnue et l'ensemble des lois proposées par le modèle.

- Critère d’Akaike. La fonction que l’on désire minimiser est

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

- Critère de Schwarz. La fonction que l’on désire minimiser est

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

I.8 Pratique des modèles SARIMA

Cette partie présente la mise en oeuvre pratique des modèles SARIMA.

I.8.1 Méthodologie

Synthèse

La démarche adoptée par Box et Jenkins est la suivante :

1. Stationnarisation
2. Identification a priori de modèles potentiels
3. Estimation des modèles potentiels
4. Vérification des modèles potentiels
5. Choix définitif d’un modèle
6. Prévision à l’aide du modèle choisi
7. Analyse a posteriori

Stationnarisation

Méthodes

La plupart des séries temporelles présentent une tendance et/ou une saisonnalité, et ne sont donc pas modélisables par un processus stationnaire. Afin de se ramener à un processus ARMA, il faut stationnariser la série⁴ et différentes méthodes sont envisageables :

- Décomposition saisonnière

Cette méthode permet d’éliminer tendance et saisonnalité, sources évidentes de non-stationnarité ; il se peut néanmoins que la série résultant de la décomposition ne soit toujours pas stationnaire.

- Différenciation

C’est la méthode employée par les modèles ARIMA et SARIMA. On ne modélise pas la série brute mais la série différenciée, en “tendance” à l’aide de $\nabla^d = (I - B)^d$, ou la série différenciée en “saisonnalité” à l’aide de $\nabla_s^D = (I - B^s)^D$. De manière générale, ∇^d permet de stationnariser des séries possédant une tendance polynomiale de degré d , et ∇_s^D des séries possédant une composante saisonnière de période s .

Si les processus stationnaires peuvent être approchés par des modèles ARMA, il n’est pas certain que la différenciation permette de stationnariser tous les processus.

⁴En toute rigueur, on devrait parler de stationnariser un processus et non une série temporelle.

– Méthode de Box-Cox

Elle permet une stationnarisation en “variance” (ou encore de stationnariser des séries présentant une tendance exponentielle). On utilise la transformation suivante : $\frac{X_t^\lambda - 1}{\lambda}$ avec $\lambda \in \mathbb{R}$. Il existe des méthodes alternatives si X_t n’est pas une série positive. Remarquons que :

$$\frac{X_t^\lambda - 1}{\lambda} \xrightarrow{\lambda \rightarrow 0} \log(X_t)$$

Il existe des tests de stationnarité, mais aucun n’est “universel”. Citons tout de même le test de Dickey-Fuller.

Pratique de la différenciation dans les modèles SARIMA

On utilise très souvent une méthode empirique basée sur l’autocorrélogramme simple.

- On effectue une différenciation en “tendance” si :
 - Les autocorrélations $\hat{\rho}(h)^5$ sont proches de 1 pour un grand nombre de retards.
 - Les premières autocorrélations $\hat{\rho}(h)$ sont proches les unes des autres (même si elles ne sont pas forcément proches de 1).
- On parle souvent de décroissance lente des autocorrélations simples.
- On effectue une différenciation en “saisonnalité” si des comportements similaires sont observés de manière périodique. Par exemple, si $\hat{\rho}(12)$, $\hat{\rho}(24)$, ... sont proches de 1, on utilise une différenciation en “saisonnalité” avec $s = 12$.

Remarques

1. Quelle que soit la méthode, on procède de manière itérative : on effectue une première différenciation ; si celle-ci n’est pas suffisante, on en effectue une seconde...
2. En pratique, on a souvent $d \leq 2$ et $D \leq 2$.

On travaille dorénavant sur une suite stationnarisée.

Identification a priori de modèles potentiels (Ordre de grandeur de p et q)

Une fois la stationnarisation effectuée, on peut se consacrer aux choix potentiels des polynômes AR et MA .

Il existe différentes méthodes pour identifier un modèle $ARMA(p, q)$:

– Méthode de Box et Jenkins

Il s’agit d’une méthode heuristique pour majorer p et q .

- Pour un processus $AR(p)$:

On peut montrer que :

$$\forall h > p : \sqrt{n}\hat{r}(h) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

On peut définir un intervalle de confiance à 95% et on recherche à partir de quelle valeur, 95% des $\hat{r}(h)$ sont dans l’intervalle $\left[-\frac{1.96}{\sqrt{n}}, \frac{1.96}{\sqrt{n}}\right]$.

⁵Il s’agit ici d’un abus de langage car si la sortie nommée “autocorrélations simples” nous permet de douter de la stationnarité du processus sous-jacent, nous ne devrions pas parler alors d’autocorrélations simples.

- Pour un processus $MA(q)$:
On peut montrer que :

$$\forall h > q : \sqrt{n}\hat{\rho}(h) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, 1 + 2\sum_{k=1}^q \rho^2(k)\right)$$

On peut définir un intervalle de confiance à 95% et on recherche à partir de quelle valeur, 95% des $\hat{\rho}(h)$ sont dans l'intervalle suivant :

$$\left[-\frac{1.96}{\sqrt{n}} \left(1 + 2\sum_{k=1}^q \hat{\rho}^2(k) \right)^{\frac{1}{2}}, \frac{1.96}{\sqrt{n}} \left(1 + 2\sum_{k=1}^q \hat{\rho}^2(k) \right)^{\frac{1}{2}} \right]$$

- Méthode du coin

On utilise la méthode du coin avec les estimations de $\hat{\rho}(h)$. Cette méthode ne permet pas toujours d'aboutir, surtout si on a des effets saisonniers.

- Méthode empirique

En pratique (surtout pour les modèles SARIMA), on essaye d'identifier les autocorrélations simples et partielles "significatives" pour caler ensuite des polynômes AR et MA qui reflètent ces liens temporels.

Afin d'obtenir des modèles potentiels, l'idéal est de regarder l'autocorrélogramme partiel afin d'émettre une hypothèse sur la partie autorégressive (simple et saisonnière), la tester puis regarder l'autocorrélogramme simple (et partiel) du résidu afin d'identifier complètement un modèle. Cette démarche par étape permet en général d'obtenir plusieurs modèles potentiels.

Estimation des modèles potentiels

On estime les modèles potentiels à l'aide des méthodes classiques : maximum de vraisemblance ou moindres carrés.

Vérification des modèles potentiels

Afin de vérifier la validité des modèles estimés, on doit vérifier :

- Significativité des paramètres

Par exemple, pour le coefficient AR d'ordre p , on effectue le test suivant :

$$\begin{cases} H_0 : \text{le processus est un } ARMA(p, q) \\ H_1 : \text{le processus est un } ARMA(p-1, q) \end{cases}$$

On utilise pour cela la statistique de test suivante :

$$t = \frac{|\hat{\varphi}_p|}{\sqrt{\mathbb{V}(\hat{\varphi}_p)}},$$

où $\mathbb{V}(\hat{\varphi}_p)$ est la variance (que nous ne précisons pas ici) de $\hat{\varphi}_p$.

Le test de Student permet de rejeter H_0 au niveau 5% si $|t|$ est supérieur à 1.96.

Il existe des résultats similaires pour les coefficients MA.

- Blancheur du résidu

On vérifie que le résidu est bien un bruit blanc, à l'aide du test de Portmanteau par exemple.

Remarque I.47. Les logiciels fournissent généralement la valeur de la statistique de test, ainsi que la p -valeur. On rejette H_0 si la p -valeur est inférieure au niveau du test α . \diamond

Choix définitif d'un modèle

Ce choix s'opère entre les modèles potentiels retenus. Il y a plusieurs critères possibles :

- Des critères d'information basés sur l'information de Kullback (par exemple, les critères d'Akaike et de Schwartz).
- Des critères basés sur le pouvoir prédictif.

Une fois ce choix effectué, le modèle retenu est utilisé à des fins de prévision.

Prévision à l'aide du modèle choisi

La fonction de prévision s'obtient assez facilement à partir des écritures autorégressive ou moyenne mobile.

Analyse a posteriori

L'analyse a posteriori permet de voir les écarts entre les prévisions et les réalisations, en tronquant la série d'un certain nombre de points ; le modèle doit être correctement estimé sur la série tronquée, et les écarts entre prévisions et réalisations doivent être faibles. On utilise des critères d'erreur comme l'erreur quadratique moyenne (Root Mean Square Error : RMSE) ou l'erreur relative absolue moyenne (Mean Average Percentage Error : MAPE) :

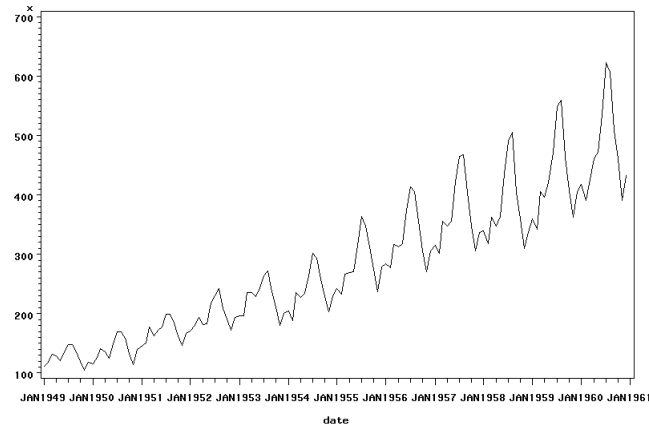
$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_i)^2}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{X_i - \hat{X}_i}{X_i} \right|$$

I.8.2 Exemple

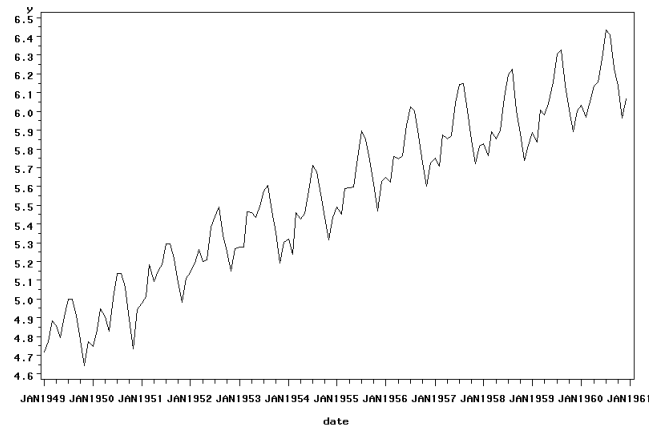
On considère le nombre de passagers aériens qu'on cherche à modéliser à l'aide la méthode de Box et Jenkins.

Nombre mensuel de passagers sur les vols internationaux (en milliers)



Le graphique de la série brute, figure I.8.2, montre une série avec une tendance (de type parabolique ou exponentielle), ainsi qu'une saisonnalité (de période 12). On constate également un accroissement de la variabilité, ce qui explique la transformation logarithmique opérée par la suite.

Nombre mensuel de passagers sur les vols internationaux (en milliers) : Log



On voit que la série transformée, figure I.8.2, présente une tendance (quasiment linéaire) et conserve une saisonnalité (de période 12) ; c'est cette dernière série qui est modélisée par la suite (pour revenir à la série de base, il suffit d'effectuer une transformation exponentielle).

Period(s) of Differencing		1, 12
Time of Working Series		0.000291
Standard Deviation		0.049572
Number of Observations		121
Observation(s) deleted by differencing		12

Autocorrelations										
Lag	Covariance	Correlation	-1	0	1	2	3	4	5	Std Error
0	0.000000	1.00000								0
1	-0.0007116	-0.0112		*****						0.08720
2	0.00012192	0.0005			***					0.09206
3	-0.0004217	-0.0014				***				0.09370
4	0.00004465	0.0002					***			0.10007
5	0.00011610	0.0005						***		0.10042
6	0.00004465	0.0002							***	0.10175
7	-0.00011610	-0.0005								0.10187
8	-1.5387E-6	-0.0000								0.10179
9	0.00007391	0.0000								0.10179
10	-0.00011610	-0.0005								0.10291
11	0.00012192	0.0005								0.10418
12	-0.00004465	-0.0002								0.10421
13	0.00011610	0.0005								0.11011
14	-0.00012192	-0.0005								0.11026
15	0.00012192	0.0005								0.11074
16	-0.00004465	-0.0002								0.11017
17	0.00011610	0.0005								0.11048
18	-0.00004465	-0.0002								0.11075
19	-0.0000221	-0.0001								0.11070
20	-0.00004465	-0.0002								0.11077
21	0.00000000	0.0000								0.12043
22	-0.00004465	-0.0002								0.12077
23	0.00004465	0.0002								0.12123
24	-0.00004465	-0.0002								0.12052

Partial Autocorrelations										
Lag	Correlation	-1	0	1	2	3	4	5	6	Std Error
1	-0.0112									
2	0.0005			*****						
3	-0.0014				***					
4	0.0002					***				
5	0.0005						***			
6	0.0002							***		
7	-0.0005								***	
8	-0.0002									***
9	0.0005									***
10	0.0002									***
11	0.0005									***
12	-0.0002									***
13	-0.0005									***
14	-0.0002									***
15	0.0005									***
16	-0.0002									***
17	0.0005									***
18	0.0002									***
19	-0.0005									***
20	-0.0002									***
21	0.0000									***
22	-0.0002									***
23	0.0005									***
24	-0.0002									***

L'autocorrélogramme simple, figure I.8.2, de la série doublement différenciée ne semble pas poser de problème de stationnarité. La série sur laquelle on travaille est donc :

$$Y_t = (I - B)(I - B^{12}) \log(X_t)$$

On constate que certaines autocorrélations simples et partielles de cette série sont significativement différentes de 0 ; voici trois modèles qui sont testés, et les résultats obtenus pour chacun de ces modèles, voir figures I.8.2, I.8.2 et I.8.2.

Conditional Least Squares Estimation						
Parameter	Estimate	Standard Error	t Value	Pr > t	Approx	Lag
BA1,1	0.04017	0.12529	0.30	0.7672		1
BA1,2	0.02476	0.12995	0.19	0.8502		12
AR1,1	-0.25680	0.12072	-2.15	0.0338		1
AR1,2	-0.06195	0.15928	-0.39	0.6981		12
Variance Estimates						
Std Error Estimate		0.00940				
Std Error Residual		0.028011				
AIC		-461.506				
BIC		-462.505				
Number of Residuals		121				
* AIC and BIC do not include log transformation.						
Correlations of Parameter Estimates						
Parameter	BA1,1	BA1,2	AR1,1	AR1,2		
BA1,1	1.000	-0.494	0.762	-0.471		
BA1,2	-0.494	1.000	-0.284	0.514		
AR1,1	0.762	-0.284	1.000	-0.261		
AR1,2	-0.471	0.514	-0.261	1.000		
Autocorrelation Check of Residuals						
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations-----		
6	7.30	2	0.0248	-0.012	-0.070	-0.116
12	10.80	8	0.2252	-0.072	-0.024	0.116
18	16.32	14	0.2311	-0.001	0.027	0.055
24	26.30	20	0.1405	-0.102	-0.094	-0.029
30	29.94	26	0.2744	-0.027	0.059	-0.000
36	37.95	32	0.2298	-0.048	0.141	-0.124
Autoregressive Factors						
Factor 1: 1 + 0.2568 B ¹² + 0.06195 B ¹² (12)						
Moving Average Factors						
Factor 1: 1 - 0.04017 B ¹² (1) - 0.02476 B ¹² (12)						

FIG. I.7 – Modèle 1 : $(I - \varphi_1 B - \varphi_{12} B^{12}) Y_t = (I + \theta_1 B + \theta_{12} B^{12}) \varepsilon_t$

```

Conditional Least Squares Estimation

Parameter      Estimation      Standard      t-Value      Approx      Lag
                Error              P > |t|
BA1,1           0.58042          0.07840         7.40        <.0001      12
AR1,1          -0.32109          0.08267        -2.95        <.0001       1

Variance Estimate      0.001624
Std Error Estimate     0.037735
AIC                    -484.872
SBC                    -479.123
Number of Residuals    121
* AIC and SBC do not include log determinant.

Correlations of Parameter
Estimates

Parameter      BA1,1      AR1,1
BA1,1           1.000      0.118
AR1,1           0.118      1.000

Autocorrelation Check of Residuals

To      Q1:      P >      -----Autocorrelations-----
Lag      Square      IF      Q1/Lq
6       7.52      4      0.1107      -0.016      -0.077      -0.128      -0.114      0.102      0.092
12      10.76     10      0.0178      -0.075      -0.022      0.116      -0.044      0.016      -0.016
18      15.09     16      0.4028      0.0011     0.062      0.003      -0.145      0.064      0.041
24      27.12     22      0.2062      -0.089     -0.091      -0.043      0.019      0.224      0.005
30      30.16     28      0.3058      -0.066     0.065      -0.004      0.025      -0.022      -0.087
36      27.42     34      0.3149      -0.042     0.127      -0.122      0.097      -0.064      -0.037

Model for variable y
Period(s) of Differencing      1,12
No season time in this model.

Autoregressive Factors
Factor 1: 1 + 0.32109 B**(1)

Moving Average Factors
Factor 1: 1 - 0.58042 B**(12)

```

FIG. I.8 – Modèle 2 : $(I - \varphi_1 B) Y_t = (I + \theta_{12} B^{12}) \varepsilon_t$

```

Conditional Least Squares Estimation

Parameter      Estimation      Standard      t-Value      Approx      Lag
                Error              P > |t|
BA1,1           0.37727          0.08196         4.60        <.0001      1
BA2,1           0.57236          0.07862         7.34        <.0001     12

Variance Estimate      0.00161
Std Error Estimate     0.03754
AIC                    -485.123
SBC                    -480.383
Number of Residuals    121
* AIC and SBC do not include log determinant.

Correlations of Parameter
Estimates

Parameter      BA1,1      BA2,1
BA1,1           1.000      -0.091
BA2,1          -0.091      1.000

Autocorrelation Check of Residuals

To      Q1:      P >      -----Autocorrelations-----
Lag      Square      IF      Q1/Lq
6       5.16      4      0.2722      0.010      0.028      -0.119      -0.100      0.081      0.077
12      7.09      10      0.6490      -0.049     -0.023      0.114      -0.045      0.026      -0.023
18      11.99     16      0.7942      0.012      0.026      0.064      -0.126      0.065      0.011
24      22.06     22      0.4272      -0.099     -0.096      -0.031      -0.021      0.214      0.012

Model for variable y
Period(s) of Differencing      1,12
No season time in this model.

Moving Average Factors
Factor 1: 1 - 0.37727 B**(1)
Factor 2: 1 - 0.57236 B**(12)

```

FIG. I.9 – Modèle 3 : $Y_t = (I + \theta_1 B) (I + \theta_{12} B^{12}) \varepsilon_t$

Afin de lire quel est le modèle testé sous SAS, il faut regarder la fin du listing dans lequel apparaissent les polynômes moyenne mobile et autorégressif.

On constate que seuls les modèles 2 et 3 conviennent. En effet les coefficients estimés dans le modèle 1 ne sont pas tous significatifs ($MA1,1 = -\theta_1$ et $AR1,2 = -\varphi_{12}$). On peut également remarquer que le test de Portmanteau n'est pas validé sur les 6 premières autocorrélations, ce qui n'est pas forcément trop grave car les résultats sont toujours plus incertains sur six autocorrélations. Ces lectures de tests sont effectuées avec un niveau de test de 5%.

Le modèle 3 présente un AIC plus faible que le modèle 2 ; on choisit donc le modèle 3 pour effectuer la prévision. On pourrait également prendre comme critères le BIC ou encore l'écart-type du résidu.

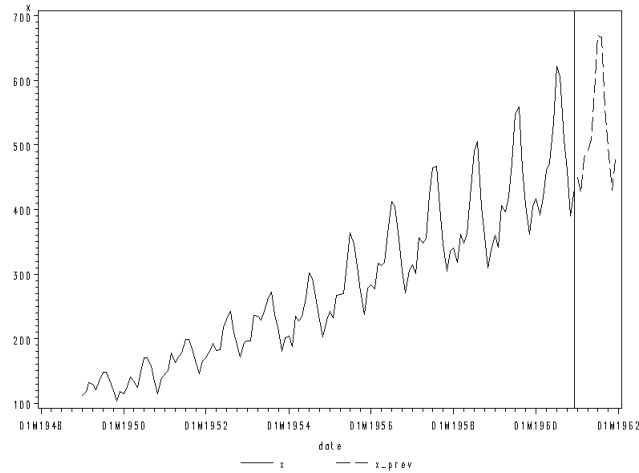


FIG. I.10 – Prédiction par le modèle 3

La prévision effectuée, cf. figure I.8.2, par le modèle 3 semble raisonnable vu le passé de la série.

On effectue une analyse a posteriori, cf. figure I.8.2 et I.8.2, en :

- tronquant la série de 12 points ;
- estimant le modèle 3 sur la série tronquée (on constate que le modèle est correctement estimé) ;
- prévoyant les douze points manquants à l'aide du modèle SARIMA ainsi estimé.

Conditional Least Squares Estimation						
Parameter	Estimation	Standard Error	t Value	Pr > t	Approx	Lag
UA1,1	0.28972	0.08765	2.72	0.0002	1	
UA2,1	0.57785	0.07929	7.29	<.0001	12	
		Variance Estimate	0.001270			
		Std Error Estimate	0.023722			
		AIC	-464.172			
		SBC	-425.514			
		Number of Residuals	119			
* AIC and SBC do not include log determinant.						
Correlations of Parameter Estimation						
Parameter	UA1,1	UA2,1				
UA1,1	1.000	-0.068				
UA2,1	-0.068	1.000				
Autocorrelation Check of Residuals						
To Lag	Chi-Square	DF	Pr > ChiSq	-----Autocorrelations-----		
6	4.96	4	0.2827	0.001	0.050	-0.149 -0.082 0.066 0.060
12	6.36	10	0.7952	-0.008	-0.027	0.002 -0.016 0.020 -0.013
18	10.26	16	0.8520	0.019	0.032	0.028 -0.129 0.073 -0.021
24	17.82	22	0.7222	-0.024	-0.102	-0.047 -0.022 0.187 -0.004
Model for variable y						
Period(s) of Differencing 1,12						
No zero term in this model.						
Using Average Factors						
Factor 1: 1 - 0.28972 B ¹² (1)						
Factor 2: 1 - 0.57785 B ¹² (12)						

FIG. I.11 – Analyse a posteriori

On obtient les résultats suivants :

$$RMSE = 18,5$$

$$MAPE = 2,9$$

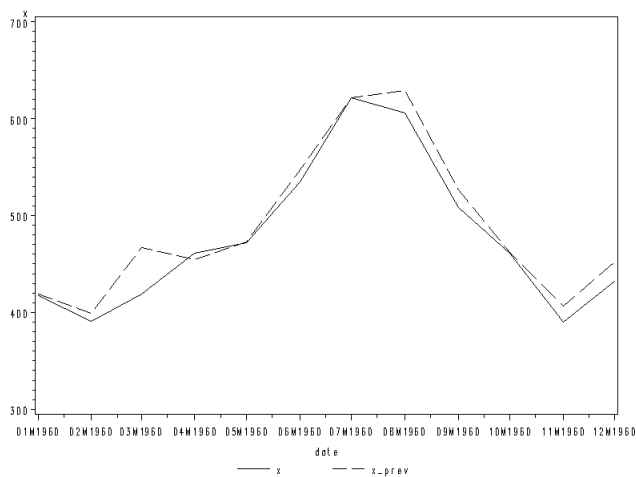


FIG. I.12 – L'interprétation des critères d'erreur dépend de la série et de la qualité de prévision exigée. Dans le cas présent, un MAPE de 2.9% semble satisfaisant a priori.