

Modélisation des séries temporelles

V. Monbet - 2016-2017

Contents

1	Introduction	4
1.1	Objectifs (et moyens)	4
1.2	Qu'est-ce qu'une série temporelle?	4
2	Lissages exponentiels	7
2.1	Introduction	7
2.1.1	Idées générales	7
2.2	Lissage exponentiel simple	7
2.2.1	Définition	8
2.2.2	Formules récursives de mise à jour	8
2.2.3	Interprétation géométrique	8
2.2.4	Choix de la constante de lissage	9
2.3	Lissage exponentiel double	9
2.3.1	Définition	9
2.3.2	Calcul des coefficients	10
2.3.3	Formules récursives de mises à jour	11
2.4	Méthode de Holt-Winters	11
2.4.1	Introduction	11
2.4.2	Ajout d'une composante saisonnière. Décomposition additive.	12
3	Modélisation de la composante déterministe	13
3.1	Analyse de la tendance	14
3.1.1	Moyenne mobile	14
3.1.2	Différenciation	15
3.1.3	Modèle additif avec tendance paramétrique	16
3.1.4	Comparaison des différentes méthodes	18
3.1.5	Conclusion	19
3.2	Analyse des composantes saisonnières	19
3.2.1	Calcul des moyennes saisonnières	19
3.2.2	Vérification du modèle	20
3.2.3	Différenciation	22
3.2.4	Comparaison des différentes méthodes	23
4	Premiers pas vers la modélisation de la composante aléatoire	24
4.1	Généralités sur les processus stationnaires	24
4.1.1	Définitions	24
4.1.2	Quelques commentaires sur l'autocorrélation	25
4.2	Quelques exemples classiques de processus.	25
4.2.1	Bruit blanc	25

4.2.2	Processus gaussien	26
4.2.3	Processus moyenne mobile	26
4.2.4	Processus autorégressif	27
5	Auto-corrélation partielle	28
5.1	Rappels	28
5.2	Coefficient de corrélation partielle entre deux variables	29
5.3	Autocorrélation partielle d'un processus stationnaire	29
5.3.1	Définition	30
5.3.2	Algorithme de Durbin-Watson	30
6	Modèles Auto-régressifs	32
6.1	Modèles autorégressifs d'ordre 1	32
6.2	Modèles auto-régressifs d'ordre p	33
6.3	Inférence statistique pour les modèles auto-régressifs	35
6.3.1	Estimation des paramètres	35
6.3.2	Estimateurs de Yule-Walker	36
6.3.3	Maximum de vraisemblance conditionnel	38
6.4	Prédiction dans les modèles autorégressifs	38
6.5	Sélection de modèle	39
6.6	Validation de modèle	40
7	Les processus moyenne mobile	41
7.1	Les processus moyenne mobile	41
7.2	Inférence pour les processus MA	41
7.3	Prédiction des processus MA	43
8	Les processus ARIMA	44
8.1	Les processus ARMA	44
8.1.1	Définition et propriétés	44
8.1.2	Inférence	46
8.1.3	Prédiction	46
8.1.4	Validation	47
8.2	Les processus ARIMA et SARIMA	47
8.2.1	Processus non stationnaires	47
8.2.2	Processus ARIMA	47
8.2.3	Test racine unité	47
8.2.4	Modèle SARIMA	48
8.3	Modèles avec co-variables : ARMAX	48
9	Modèles hétéroscédastiques	49
9.1	Introduction	49
9.2	Caractéristiques de la volatilité	49
9.3	Structure du modèle	51
9.3.1	Généralités	51
9.3.2	Modèle ARCH(1)	52
9.3.3	Inconvénients	53
9.4	Inférence	53

Chapter 1

Introduction

Ce cours est une initiation aux méthodes probabilistes de prévision de phénomènes qui évoluent dans le temps et de modélisation de séries temporelles.

1.1 Objectifs (et moyens)

Les principaux objectifs de la modélisation des séries temporelles sont les suivants.

- Decrire. Par exemple,
 - en économétrie, détecter puis analyser les périodes de crises et croissances ;
 - en reconnaissance vocale, reconnaître les mots dans des signaux ;
 - dans le séquençage du génome, détecter les parties de l'ADN qui contiennent de l'information.
- Comparer deux séries temporelles. Par exemple, l'évolution démographique de deux régions ou deux séquences d'ADN.
- Prédire l'évolution future de la série temporelle à partir des valeurs qui ont été observées. Par exemple, pour des raisons socio-économiques on veut prévoir le temps qu'il va faire, l'évolution des ventes d'un produit, la consommation d'électricité, etc.)
Comment prévoir : en s'appuyant sur le passé. Pour prédire les ventes de l'année $j + 1$, on s'appuie sur l'évolution des ventes durant les années $j, j - 1, \dots$ mais on tient compte aussi d'évènement (conjoncture, économique, crise, ...).
Peut-on prévoir parfaitement bien?

Quelles méthodes pour modéliser et/ou prévoir?

1.2 Qu'est-ce qu'une série temporelle?

Définition - Une série temporelle (ou chronologique) est une suite d'observations x_1, x_2, \dots, x_n indexée par le temps. On supposera qu'il s'agit d'une réalisation d'un processus X , c'est à dire d'une suite $\{X_j\}$ de variables aléatoires.

Une série temporelle est généralement constituée de plusieurs éléments.

- Tendance : représente l'évolution à long terme de la série (échelle interannuelle). Exemples : croissance économique, évolution climatologique à long terme (cyclique ou non)

- Saisonnalité : évolution se répétant régulièrement tous les ans ou tous les mois ou toutes les semaines. Exemples :
 - En météorologie, températures plus faibles en hiver qu'en été.
 - En économie, saisonnalité induite par les périodes de vacances, les périodes de fêtes, le climat...
 - Chiffre d'affaire d'un magasin.
- Composante stationnaire (ou résiduelle) : ce qui reste lorsque l'on a enlevé les autres composantes. Décrit l'évolution à court terme de la série (échelle journalière).

Notion de série temporelle stationnaire définie plus précisément dans la suite. Cette hypothèse jouera un rôle fondamental dans la suite, et remplacera l'hypothèse usuelle des v.a i.i.d. (ici, il peut exister une dépendance entre deux valeurs successives prises par la série observée).

Le modèle le plus courant consiste à supposer que la série initiale s'écrit sous la forme (modèle additif)

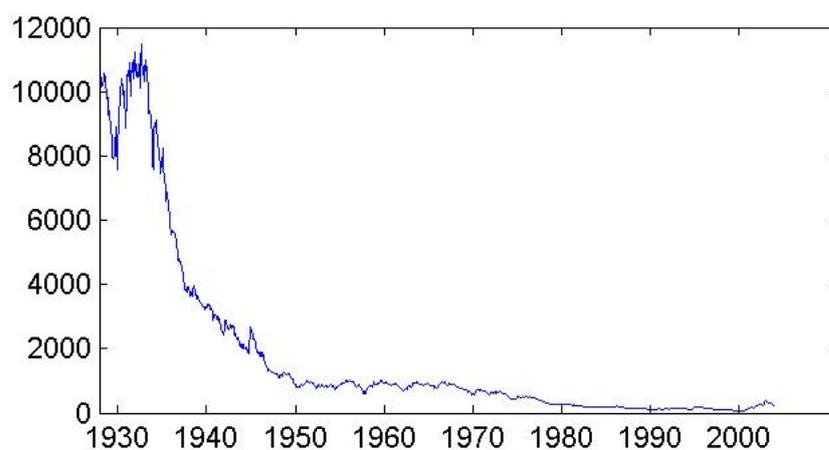
$$X_t = T_t + S_t + Y_t \text{ pour tout } t \in \{1, \dots, n\}$$

avec X_t la tendance, S_t la composante saisonnière (fonction périodique de période un an) et Y_t la composante stationnaire.

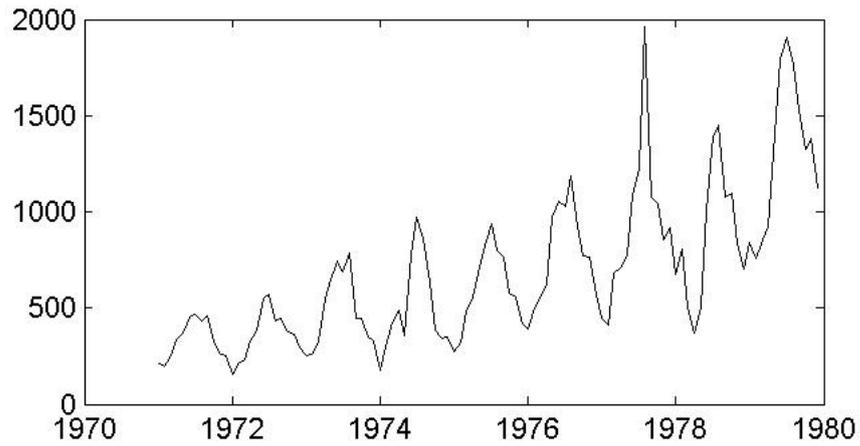
Exemples de séries temporelles

- **Economie**

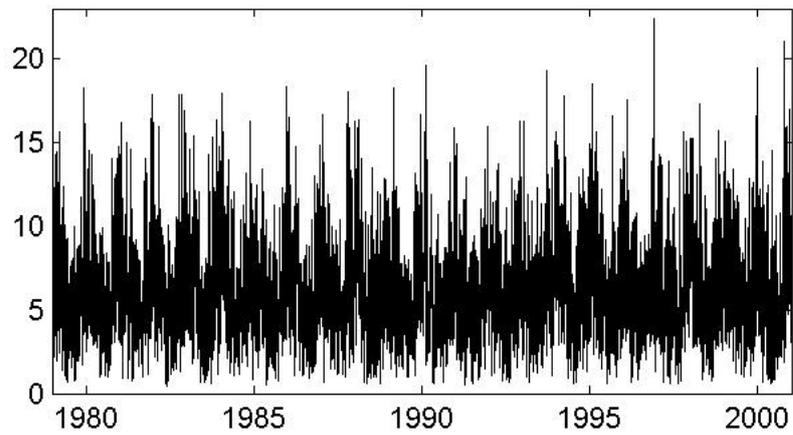
1. évolution du cours du Dow Jones entre 1928 et 2004, données mensuelles



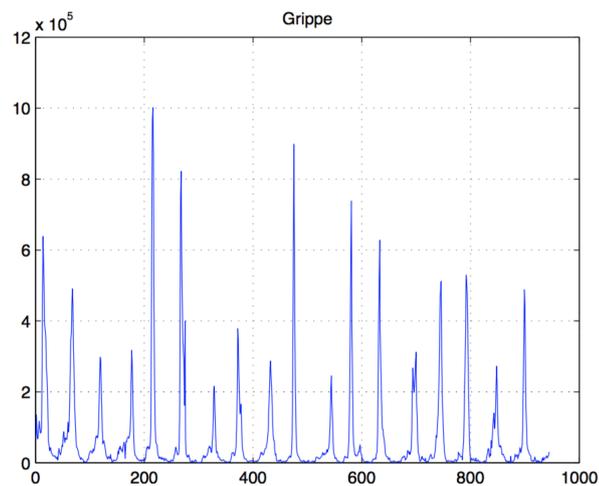
2. production de poissons, en milliers de francs, en Finistère nord (Brest, Morlaix, Paimpol) entre 1971 et 1979, données mensuelles



- **Environnement** - évolution de l'intensité du vent, en m/s, au large de la Bretagne entre 1979 et 2001, données journalières.



- **Santé** - Exemple 4 : Evolution des cas de grippe en France, par semaine, de 1984 (semaine 44) à 2002 (semaine 50), soit 945 valeurs.



Chapter 2

Lissages exponentiels

2.1 Introduction

Ce cours sur le lissage exponentiel est emprunté à M. Cl. Viano
<http://math.univ-lille1.fr/~viano/economcours.pdf>

Dans l'industrie, il est courant que l'on doive effectuer des prévisions de vente de production ou de stock pour une centaine de produits. Ces prévisions sont effectuées à partir de données par exemple mensuelles et l'horizon de prévision est court, un an maximum. Un certain nombre de techniques "autoprojectives" sont regroupées sous la rubrique du lissage exponentiel. Ces approches construisent des prévisions assez précises. Elles sont peu coûteuses et les calculs peuvent être rendus très rapides par l'utilisation de formules récursives de mise à jour. Elles ont beaucoup de succès dans les milieux dans lesquels un grand nombre de chroniques unidimensionnelles doivent être analysées séparément dans un but de prévision. Les succès sont importants malgré l'absence de bases théoriques solides comparables à celles des méthodes ARMA, ARIMA et SARIMA. Nous présentons les bases empiriques des techniques et précisons les conditions d'utilisation optimales. Les procédures automatiques présentent quelques dangers. On risque des mauvaises prévisions si la technique utilisée n'est pas adaptée à la chronique analysée. Le lissage exponentiel est brièvement présenté dans Gourieroux-Monfort ([3]). Une présentation détaillée figure dans Mélard ([4]) page 139-170. Un article de Gardner ([2]) fait un point sur la question.

2.1.1 Idées générales

On dispose d'une série chronologique $x = (x_1, \dots, x_n)$ de longueur n enregistrée aux dates $1, \dots, n$. On se situe à la date n et on souhaite prévoir la valeur x_{n+h} non encore observée à l'horizon h . On note cette prévision : $\hat{x}_{n,h}$. L'entier n est parfois appelé base de la prévision. La plus ancienne des techniques est celle du lissage exponentiel simple. Dans toutes ces techniques il s'agit d'ajuster à la chronique, localement, une fonction simple : une constante dans le lissage exponentiel simple, une droite dans le lissage exponentiel double, des fonctions polynomiales ou périodiques dans les lissages plus généraux.

2.2 Lissage exponentiel simple

Dans cette partie la prévision très simple, trop simple, ne dépend pas de l'horizon h de prévision. Nous conservons l'indice pour être cohérent avec la suite.

2.2.1 Définition

La prévision $\hat{x}_{n,h}$ est construite : en prenant en compte toute l'histoire de la chronique, de sorte que, plus on s'éloigne de la base n de la prévision, moins l'influence des observations correspondantes est importante. Cette décroissance de l'influence est de type exponentiel. De là vient le nom de la technique. On se donne α , appelé constante de lissage, avec $0 < \alpha < 1$, et on définit (la prévision $\hat{x}_{n,h}$ ne dépend pas de h) :

$$\hat{x}_{n,h} = (1 - \alpha) \sum_{j=0}^{n-1} \alpha^j x_{n-j}. \quad (2.1)$$

Remarquons que plus α est voisin de 1, plus l'influence des observations éloignées dans le temps de la base n est grande.

2.2.2 Formules récursives de mise à jour

La définition (2.1) vérifie les formules de récurrences suivantes

$$\hat{x}_{n,h} = (1 - \alpha)x_n + \alpha\hat{x}_{n-1,h} \quad (2.2)$$

$$\hat{x}_{n,h} = \hat{x}_{n-1,h} + (1 - \alpha)(x_n - \hat{x}_{n-1,h}) \quad (2.3)$$

Par conséquent, la connaissance de la prévision à l'horizon h sur la base $n - 1$, soit $\hat{x}_{n-1,h}$, et l'observation x_n suffisent pour calculer immédiatement la prévision suivante $\hat{x}_{n,h}$.

Cependant, l'utilisation des formules récursives nécessite d'initialiser la récurrence en prenant par exemple $\hat{x}_{1,h} = x_1$, on remarque que pour n assez grand la valeur initiale a peu d'influence. Remarquons aussi que la formule (2.2) permet d'interpréter $\hat{x}_{n,h}$ comme le barycentre de x_n et de $\hat{x}_{n-1,h}$ affectés respectivement des masses $1 - \alpha$ et α . L'influence de la dernière observation est d'autant plus grande que α est proche de 0.

2.2.3 Interprétation géométrique

Asymptotiquement $\hat{x}_{n,h}$ est la meilleure approximation au sens des moindres carrés (pour des pondérations exponentielles) de x par le vecteur constant $\alpha\delta$ où $\delta = (1, \dots, 1) \in \mathbb{R}^n$. En effet la valeur de a qui minimise

$$\sum_{j=0}^{n-1} \alpha^j (x_{n-j} - a)^2. \quad (2.4)$$

est fournie par

$$\hat{a}(n) = \frac{1 - \alpha}{1 - \alpha^n} \sum_{j=0}^{n-1} \alpha^j x_{n-j}.$$

Si l'on transforme légèrement le modèle en considérant que

$$x = (x_1, \dots, x_n) = (\dots, 0, \dots, 0, x_1, \dots, x_n)$$

c'est à dire que les coordonnées du nouveau vecteur sont nulles pour tous les indices négatifs ou nul. On cherche donc la valeur de a qui minimise

$$\sum_{j=0}^{+\infty} \alpha^j (x_{n-j} - a)^2$$

la solution est fournie par

$$\hat{x}_{n,h} = (1 - \alpha) \sum_{j=0}^{n-1} \alpha^j x_{n-j}.$$

Il est facile de montrer que les deux estimateurs sont asymptotiquement équivalents, on a

$$\lim_{n \rightarrow \infty} \hat{a}(n) = \lim_{n \rightarrow \infty} \hat{x}_{n,h}.$$

Dans la minimisation de (2.4) là aussi, l'influence des observations décroît exponentiellement lorsqu'on s'éloigne de la base. On peut dire que l'ajustement par la constante α est local et se fait au voisinage de n . La taille réelle de la plage des observations influentes dépend de α . Plus α est grand plus la taille est grande.

2.2.4 Choix de la constante de lissage

La méthode n'est pas bonne en présence de tendance, de composante saisonnière ou de fluctuations de hautes fréquences. En pratique le lissage exponentiel simple est assez peu utilisé.

Des valeurs de α comprises entre 0.7 et 0.99 produisent des prévisions qui tiennent compte du passé lointain. Elles sont rigides et peu sensibles aux variations conjoncturelles. Il n'en n'est plus de même si α est compris entre 0.01 et 0.3, la prévision est souple c'est à dire influencée par les observations récentes.

On peut aussi déterminer un α adapté à la prévision d'une série donnée. Pour un horizon h on cherche la valeur de α qui minimise

$$\sum_{t=1}^{n-h} (x_{t+h} - (1 - \alpha) \sum_{j=0}^{t-1} \alpha^j x_{t-j})^2,$$

c'est à dire la valeur qui en moyenne réalise la meilleure prévision à l'horizon h sur les bases $1, 2, \dots, n - h$ en utilisant le critère des moindres carrés. On choisit une grille de valeurs de α pour lesquelles on calcule le critère et on retient la valeur qui le minimise. Cette technique de choix de constante peut être reprise dans les méthodes qui vont suivre. Les calculs sont souvent plus longs car ce n'est, en fait, plus une constante de lissage que nous sommes amenés à déterminer mais souvent deux ou trois.

2.3 Lissage exponentiel double

2.3.1 Définition

On ajuste au voisinage de n une droite d'équation $y_t = a_1 + a_2(t - n)$. Le prédicteur sera :

$$\hat{x}_{n,h} = \hat{a}_1 + \hat{a}_2 h. \quad (2.5)$$

Les coefficients \hat{a}_1, \hat{a}_2 sont solution de

$$\min_{a_1 \in \mathbb{R}, a_2 \in \mathbb{R}} \sum_{j=0}^{n-1} \alpha^j (x_{n-j} - (a_1 - a_2 j))^2. \quad (2.6)$$

En remplaçant, comme dans le lissage simple, la sommation finie par une somme de 0 à $+\infty$, le problème précédent devient

$$\min_{a_1 \in \mathbb{R}, a_2 \in \mathbb{R}} \sum_{j=0}^{+\infty} \alpha^j (x_{n-j} - (a_1 - a_2 j))^2.$$

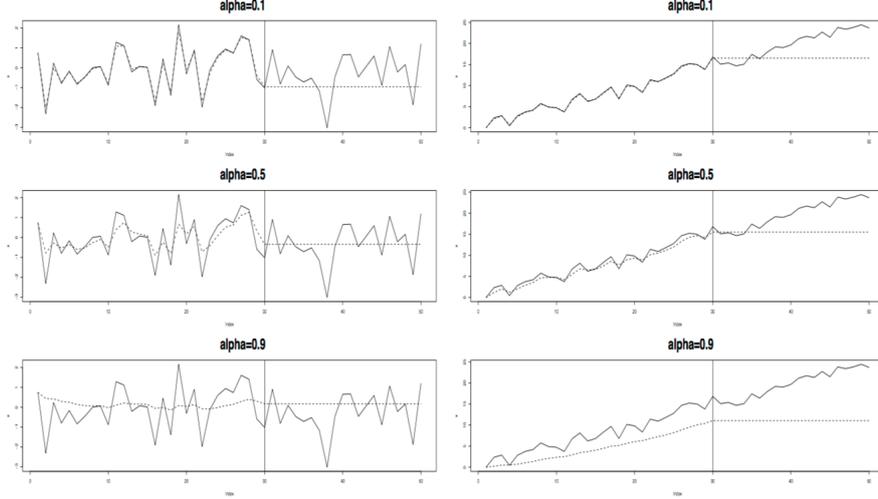


Figure 2.1: Lissage et prévision par la méthode de lissage exponentiel simple d'un bruit blanc [gauche] et d'une tendance linéaire $X(t) = 0.5t + \epsilon_t$, $\epsilon_t \sim \mathcal{N}(0, 1)$ [droite] ($\alpha = .1, .5, .9$)

2.3.2 Calcul des coefficients

On annule les dérivées partielles par rapport à a_1 et a_2 du critère (6) et on vérifie à l'aide des dérivées secondes que l'on a bien un minimum. En notant C la fonction à minimiser

$$\frac{\partial C}{\partial a_1} = -2 \sum_{j=0}^{n-1} \alpha^j (x_{n-j} - (a_1 - a_2 j))$$

$$\frac{\partial C}{\partial a_2} = 2 \sum_{j=0}^{n-1} j \alpha^j (x_{n-j} - (a_1 - a_2 j)),$$

Puisque $\sum_0^\infty \alpha^j = \frac{1}{1-\alpha}$, $\sum_0^\infty j \alpha^j = \frac{\alpha}{(1-\alpha)^2}$ et $\sum_0^\infty j^2 \alpha^j = \frac{\alpha(1+\alpha)}{(1-\alpha)^3}$, en annulant les dérivées, on obtient

$$(1-\alpha) \sum_{j=0}^{n-1} \alpha^j x_{n-j} - a_1 + a_2 \frac{\alpha}{(1-\alpha)} = 0$$

$$(1-\alpha)^2 \sum_{j=0}^{n-1} j \alpha^j x_{n-j} - a_1 \alpha + a_2 \frac{\alpha(1+\alpha)}{(1-\alpha)} = 0$$

Introduisons deux lissages exponentiels simples successifs, celui des x produit L_1 , celui de L_1 produit L_2 :

$$L_1(n) = (1-\alpha) \sum_{j=0}^{n-1} \alpha^j x_{n-j}$$

$$L_2(n) = (1-\alpha) \sum_{j=0}^{n-1} \alpha^j L_1(n-j)$$

On a donc

$$L_2(n) - (1-\alpha)L_1(n) = (1-\alpha)^2 \sum_{j=1}^{n-1} j \alpha^j x_{n-j}$$

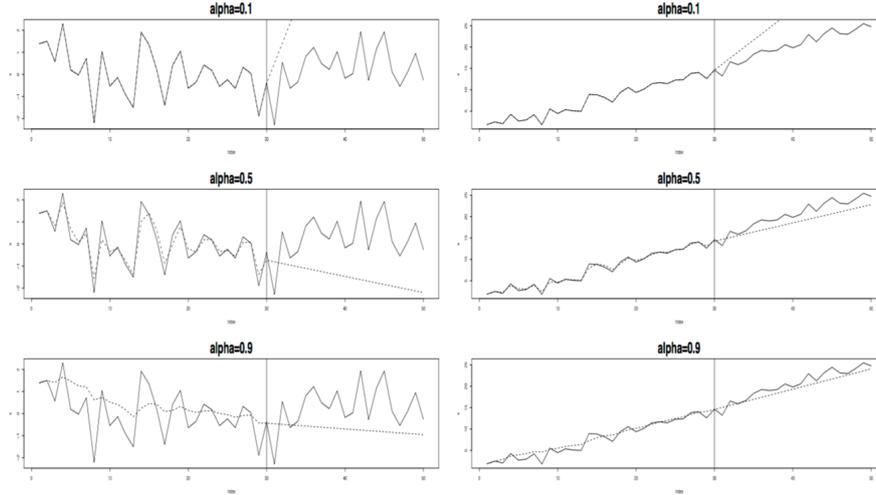


Figure 2.2: Lissage et prévision par la méthode de lissage exponentiel double d'un bruit blanc [gauche] et d'une tendance linéaire $X(t) = 0.5t + \epsilon_t$, $\epsilon_t \sim \mathcal{N}(0, 1)$ [droite]($\alpha = .1, .5, .9$)

et L_1 et L_2 sont liés à a_1 et a_2 par les relations

$$L_1(n) - a_1 + a_2 \frac{\alpha}{1 - \alpha} = 0$$

$$L_2(n) - (1 - \alpha)L_1(n) - a_1\alpha + a_2 \frac{\alpha(1 + \alpha)}{1 - \alpha} = 0.$$

Les solutions recherchées vérifient donc,

$$\hat{a}_1(n) = 2L_1(n) - L_2(n)$$

$$\hat{a}_2(n) = \frac{1 - \alpha}{\alpha} (L_1(n) - L_2(n)).$$

Soit encore

$$L_1(n) = \hat{a}_1(n) - \frac{\alpha}{1 - \alpha} \hat{a}_2(n)$$

$$L_2(n) = \hat{a}_1(n) - \frac{2\alpha}{1 - \alpha} \hat{a}_2(n).$$

2.3.3 Formules récursives de mises à jour

On peut en déduire des formules récursives de mise à jour

$$\hat{a}_1(n) = \hat{a}_1(n - 1) + \hat{a}_2(n - 1) + (1 - \alpha^2)(x_n - \hat{x}_{n-1,1}) \quad (2.7)$$

$$\hat{a}_2(n) = \hat{a}_2(n - 1) + (1 - \alpha)^2(x_n - \hat{x}_{n-1,1}). \quad (2.8)$$

On peut prendre comme valeurs initiales $\hat{a}_1(0) = x_1$ et $\hat{a}_2(0) = x_2 - x_1$.

2.4 Méthode de Holt-Winters

2.4.1 Introduction

Cette approche a pour but d'améliorer et de généraliser le lissage exponentiel simple. Nous étudions plusieurs cas particuliers de cette méthode :

- ajustement d'une droite affine (sans saisonnalité)
- ajustement d'une droite affine + une composante saisonnière
- ajustement d'une constante + une composante saisonnière

le dernier cas étant déduit du second. On ajuste au voisinage de l'instant n une droite

$$x_t = a_1 + (t - n)a_2$$

La variante proposée par Holt-Winters porte sur les formules de mise à jour dans l'estimation de a_1 et a_2 par les formules (2.7) et (2.8). La formule (2.9) s'interprète comme le barycentre de l'observation x_n et de la prévision à l'horizon 1 à la date $n - 1$. La formule (2.10) s'interprète comme le barycentre de la pente estimée à l'instant $n - 1$ et la différence $\hat{a}_1(n) - \hat{a}_1(n - 1)$ entre les ordonnées à l'origine aux dates n et $n - 1$.

Soit $0 < \beta < 1$ et $0 < \gamma < 1$ deux constantes fixées et les formules de mise à jour

$$\hat{a}_1(n) = \beta x_n + (1 - \beta)(\hat{a}_1(n - 1) + \hat{a}_2(n - 1)) \quad (2.9)$$

$$\hat{a}_2(n) = \gamma(\hat{a}_1(n) - \hat{a}_1(n - 1)) + (1 - \gamma)\hat{a}_2(n - 1). \quad (2.10)$$

La prévision prend la forme

$$\hat{x}_{n,h} = \hat{a}_1(n) + h\hat{a}_2(n).$$

Cette méthode est plus souple que le lissage exponentiel amélioré, dans la mesure où elle fait intervenir deux constantes β et γ au lieu d'une α . Les équations (2.7) et (2.8), qui sont un cas particulier des formules (2.9) et (2.10) avec $\beta = 1 - \alpha^2$ et $\gamma = \frac{1 - \alpha}{1 + \alpha}$. On remarque que si β et γ sont petits, le lissage est fort puisqu'alors α est grand et que l'on tient compte du passé lointain.

2.4.2 Ajout d'une composante saisonnière. Décomposition additive.

On ajuste au voisinage de n , la chronique

$$x_t = a_1 + (t - n)a_2 + s_t$$

où s_t est une composante périodique de période T . Les formules récursives de mise à jour sont : soit $0 < \beta < 1$,

$$\hat{a}_1(n) = \beta(x_n - \hat{s}_{n-T}) + (1 - \beta)(\hat{a}_1(n - 1) + \hat{a}_2(n - 1))$$

$$\hat{a}_2(n) = \gamma(\hat{a}_1(n) - \hat{a}_1(n - 1)) + (1 - \gamma)\hat{a}_2(n - 1)$$

$$\hat{s}_n = \delta(x_n - \hat{a}_1(n)) + (1 - \delta)\hat{s}_{n-T}.$$

La prévision prend la forme

$$\hat{x}_{n,h} = \hat{a}_1(n) + (t - n)\hat{a}_2(n) + \hat{s}_{n+h-T}, \quad 1 \leq h \leq T$$

$$\hat{x}_{n,h} = \hat{a}_1(n) + (t - n)\hat{a}_2(n) + \hat{s}_{n+h-2T}, \quad T + 1 \leq h \leq 2T.$$

et ainsi de suite pour $h > 2T$. Dans cette procédure la difficulté de choisir les constantes de lissage est encore plus grande que pour les autres situations.

Les valeurs initiales suivantes permettent (mais on pourrait en proposer d'autres) de commencer le calcul récursif des prévisions à partir de l'instant $T + 2$:

$$\hat{a}_1(T + 1) = x_{T+1}$$

$$\hat{a}_2(T + 1) = \frac{x_{T+1} - x_1}{T}$$

$$\hat{s}_j = x_j - (x_1 + (T - 1)\hat{a}_2(T + 1)) \quad \text{pour } j \in \{1, \dots, T\}.$$

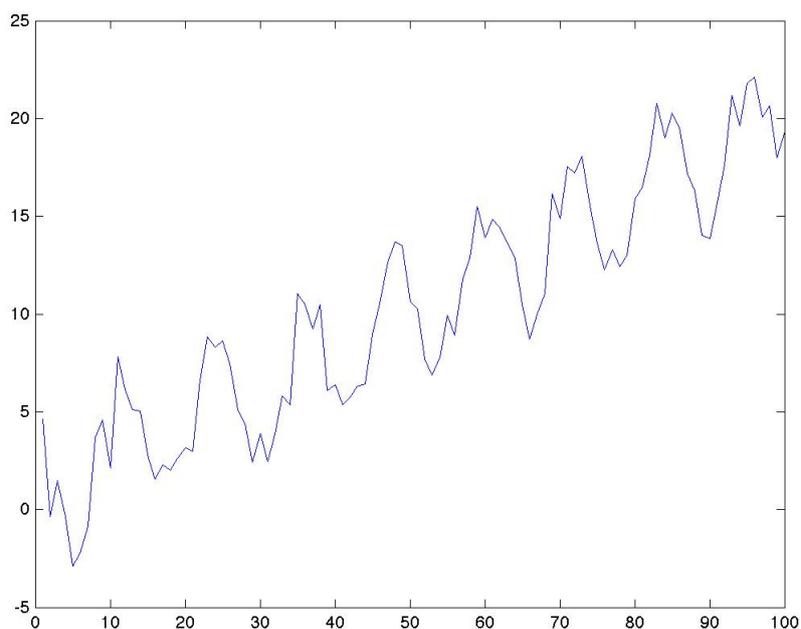
Chapter 3

Modélisation de la composante déterministe

Le modèle le plus usuel consiste à supposer que la série initiale s'écrit sous la forme (modèle additif)

$$X_t = T_t + S_t + Y_t \text{ pour tout } t \in \{1, \dots, n\}$$

avec X_t la tendance, S_t la composante saisonnière (fonction périodique de période un an) et Y_t la composante stationnaire.



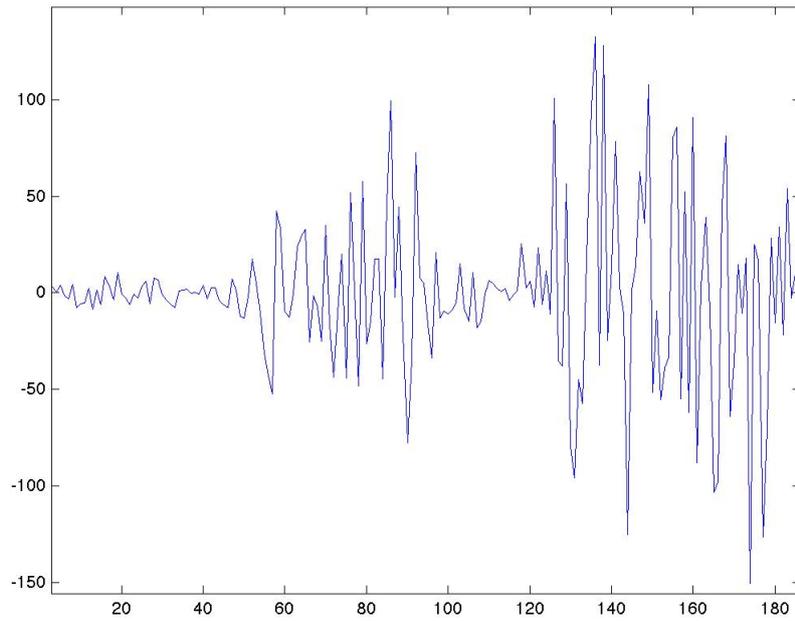
Quand le modèle s'écrit comme une somme de plusieurs composantes comme ci-dessus on parle de **modèle additif**.

Certaines séries temporelles se comportent selon un **modèle multiplicatif**. On a alors

$$X_t = T_t S_t Y_t \text{ pour tout } t \in \{1, \dots, n\}$$

On remarque qu'on se ramène naturellement à un modèle additif par passage au log :

$$\log(X_t) = \log(T_t) + \log(S_t) + \log(Y_t) \text{ pour tout } t \in \{1, \dots, n\}$$



3.1 Analyse de la tendance

exemple : poissons

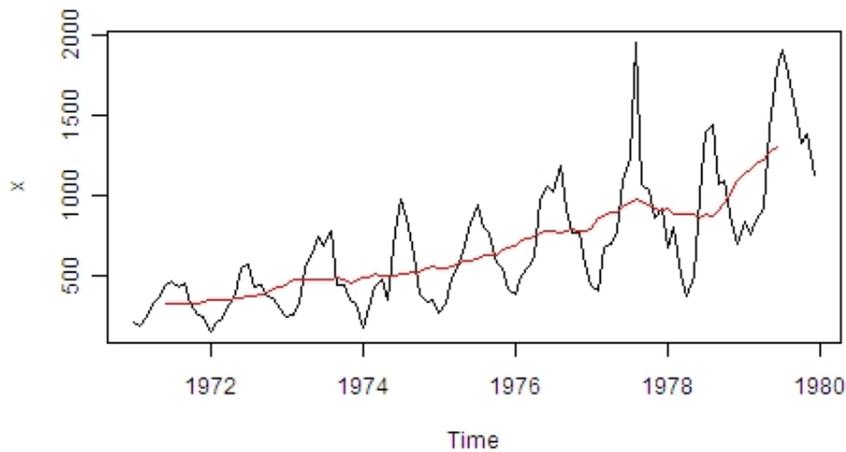
3.1.1 Moyenne mobile

Si $\{x_i\}_{i \in \{1, \dots, n\}}$ est une série temporelle, alors la moyenne mobile d'ordre p associée est la série temporelle définie pour $t \in \{p + 1, \dots, n - p\}$ par

$$\hat{x}_t = \sum_{i=t-p}^p x_i$$

avec $2p + 1$ la largeur de la fenêtre.

Exemple : Poissons



3.1.2 Différenciation

On peut éliminer la tendance par différenciation.

Définition - Notons ∇ l'opérateur de différenciation, c'est à dire l'opérateur défini par

$$\nabla x_t = x_t - x_{t-1}$$

pour tout $t \geq 2$.

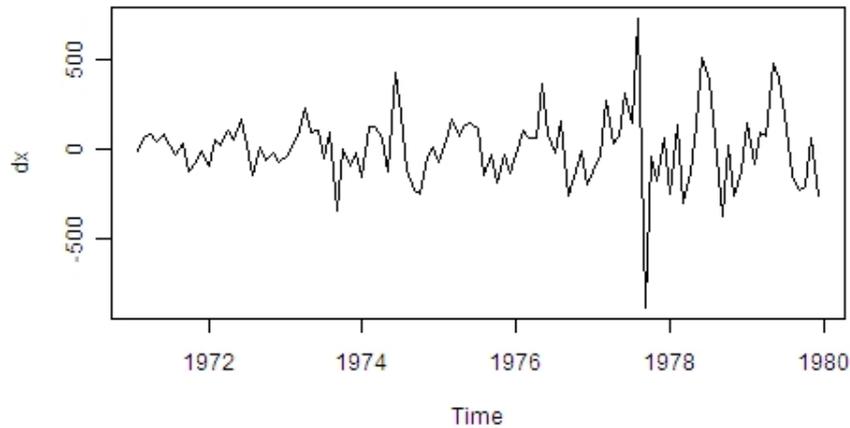
Définition - On définit l'opérateur de différenciation d'ordre k par la formule de récurrence

$$\nabla^{(k)} x_t = \nabla(\nabla^{(k-1)} x_t)$$

Propriétés -

1. Soient u_t et v_t deux séries temporelles, $\nabla(u_t + v_t) = \nabla u_t + \nabla v_t$.
2. Soit x_t une série temporelle et $\lambda \in \mathbb{R}$, $\nabla(\lambda x_t) = \lambda \nabla x_t$.
3. Soit $y_t = a_0 + a_1 t + \dots + a_k t^k$, alors $\nabla^{(k)} y_t = k! a_k$ et $\nabla^{(k+1)} y_t = 0$.

La 3ème propriété implique que si $X_t = T_t + z_t$ avec T_t une tendance polynomiale, alors on peut supprimer la tendance en appliquant successivement plusieurs fois l'opérateur ∇ .



3.1.3 Modèle additif avec tendance paramétrique

Définition On dit qu'une série temporelle suit un modèle de tendance additif lorsqu'elle peut se décomposer sous la forme

$$X_t = T_t + Z_t$$

avec Z_t une série temporelle sans tendance.

On dit qu'une série temporelle a une tendance linéaire lorsque

$$T_t = a \times t + b$$

On dit qu'une série temporelle a une tendance polynomiale lorsque

$$T_t = a_0 + a_1t + a_2t^2 + \dots$$

Ajustement du modèle

On suppose que les observations $\{x_1, \dots, x_n\}$ suivent un modèle additif avec une tendance paramétrique représentée par une fonction f de paramètres θ , c'est à dire vérifie

$$x_t = f(t; \theta) + z_t$$

avec z_t une série temporelle sans tendance.

On cherche alors à estimer les paramètres inconnus θ . On utilise généralement la méthode des moindres carrés.

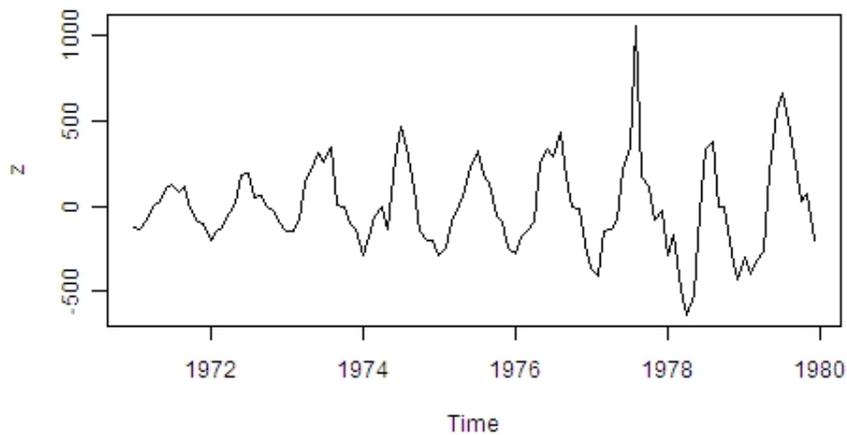
Vérification du modèle et autres modèles paramétriques

Afin de vérifier l'adéquation du modèle aux observations, on calcule les estimations $\hat{\theta}$ puis on trace le graphique

$$t \mapsto x_t - f(t; \hat{\theta})$$

Si le modèle est valable, il ne reste plus de tendance sur cette nouvelle série temporelle. Sinon on cherche un modèle plus adapté.

Série résiduelle après avoir retiré la tendance estimée (polynôme d'ordre 2)



Il reste une tendance visible. Sur la série initiale, il semble que les fluctuations saisonnières sont proportionnelles à la tendance. Dans ce cas, un modèle additif n'est pas adapté. On peut alors tester un modèle multiplicatif.

Définition - On dit qu'une série temporelle suit un modèle de tendance multiplicatif lorsqu'elle peut se décomposer sous la forme

$$X_t = T_t Z_t$$

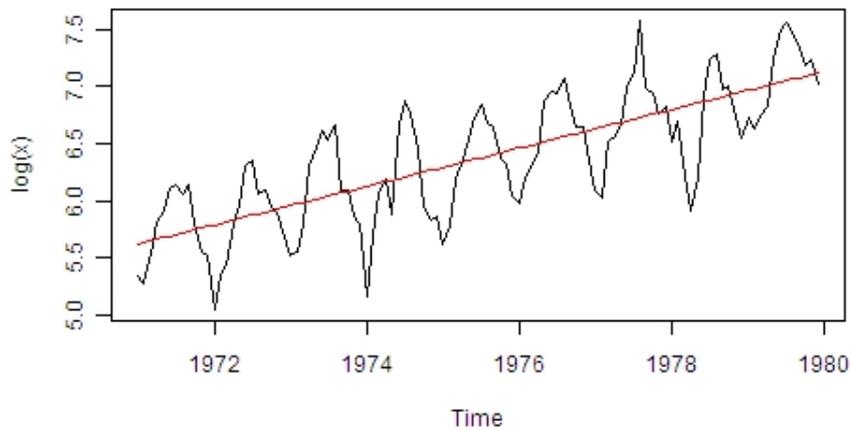
avec Z_t une série temporelle sans tendance.

En pratique, on se ramène à un modèle additif par passage au logarithme, puisqu'on a alors

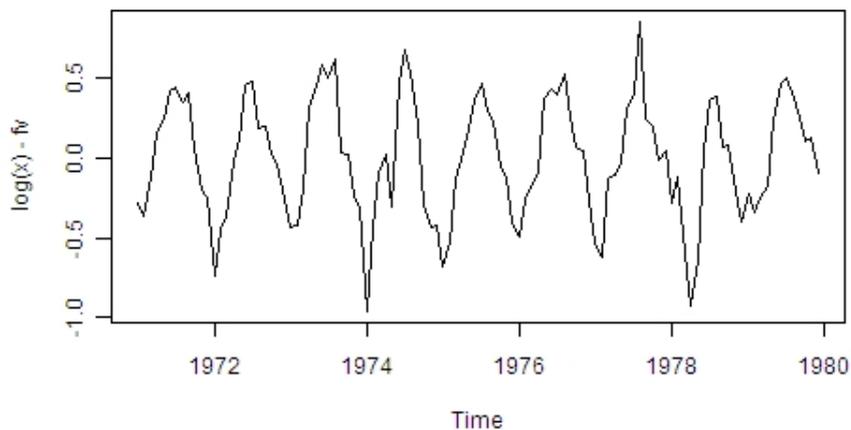
$$\log(X_t) = \log(T_t) + \log(Z_t)$$

Retour sur l'exemple de la production de poisson en finistère nord entre 1971 et 1979.

Logarithme de la série initiale et tendance linéaire ajustée



Série résiduelle après avoir retirée la tendance estimée



3.1.4 Comparaison des différentes méthodes

Les principaux avantages des méthodes paramétriques sur les deux autres méthodes sont

- qu'elles fournissent un modèle facilement interprétable et qui permet, par exemple, de comparer plusieurs séries temporelles entre elles.
- qu'elles peuvent être utilisées en prévision, ce qui n'est pas le cas des autres méthodes.

L'avantage principal des deux premières méthodes (moyenne mobile et différenciation) est qu'elles s'ajustent à de nombreuses séries temporelles sans modifications, alors que pour les modèles paramétriques il peut être difficile de choisir le bon modèle.

3.1.5 Conclusion

Pour modéliser la tendance dans une série temporelle, les différentes étapes sont :

1. Tracer la série temporelle
2. Utiliser ce graphique pour identifier le modèle en répondant aux questions suivantes :
3. Modèle additif ou multiplicatif? Pour cela, on regarde si les composantes saisonnières et stationnaires sont proportionnelles à la tendance ou non. Lorsque le modèle est additif, on travaille sur la série $\log(x_t)$ afin de se ramener à un modèle additif.
4. Méthode paramétrique ou non-paramétrique? Pour cela, on regarde si il existe un modèle simple permettant de décrire la tendance observée (modèle linéaire, polynomiale, exponentiel...). Si on ne trouve pas de tel modèle, on peut utiliser une méthode non paramétrique (MM ou différentiation).
5. Vérifier le modèle en traçant la série "résiduelle", c'est à dire la série dans laquelle on a enlevé la tendance en utilisant la méthode choisie.

3.2 Analyse des composantes saisonnières

Dans ce paragraphe, X_t désigne une série temporelle sans tendance. En pratique, lorsqu'il existe une tendance sur la série initiale X_t , on commence par l'enlever en utilisant l'une des méthodes décrites dans la section précédente.

On supposera dans la suite de cette section, sauf mention contraire, que la série X_t suit un modèle de saisonnalité additif, c'est à dire que

$$X_t = S_t + Z_t$$

avec S_t une fonction périodique de période τ , avec τ le nombre de données par année, c'est à dire vérifiant $S_t = S_{t+\tau}$ et Z_t une série temporelle stationnaire.

De nombreuses séries temporelles observées dans la nature suivent un modèle multiplicatif, c'est à dire vérifient

$$X_t = S_t Z_t$$

Dans ce cas, les fluctuations de la série autour de la composante saisonnière sont proportionnelles à celle-ci. On se ramène alors à un modèle additif par passage au logarithme.

Dans le cadre des modèles additifs, plusieurs méthodes peuvent être utilisées pour estimer la fonction $t \mapsto S_t$. C'est l'objet des paragraphes ci-dessous.

3.2.1 Calcul des moyennes saisonnières

On renumérote la série X_t sous la forme X_{jk} , $j = 1, \dots, T_{an}$, $k = 1, \dots, N_{an}$ avec T_{an} la longueur de l'année et N_{an} le nombre d'années de mesure. Le modèle se réécrit alors sous la forme :

$$X_{jk} = S_k + Z_{jk}$$

Un estimateur naturel de S_k est donné par $\frac{1}{T} \sum_{j=1}^{N_{an}} Z_{jk}$, c'est à dire par la moyenne des observations disponibles à la date k pour chaque année. On appelle généralement S_k les "moyennes saisonnières".

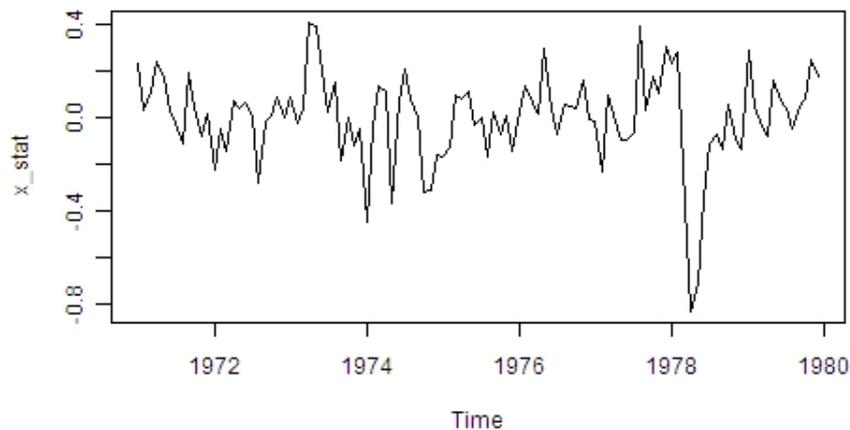
3.2.2 Vérification du modèle

Afin de vérifier l'adéquation du modèle, on trace la série résiduelle Z_t . Si le modèle convient, il ne reste plus de saisonnalité apparente sur cette série résiduelle, généralement appelée "série corrigée des variations saisonnières".

Retour sur l'exemple de la production de poissons.

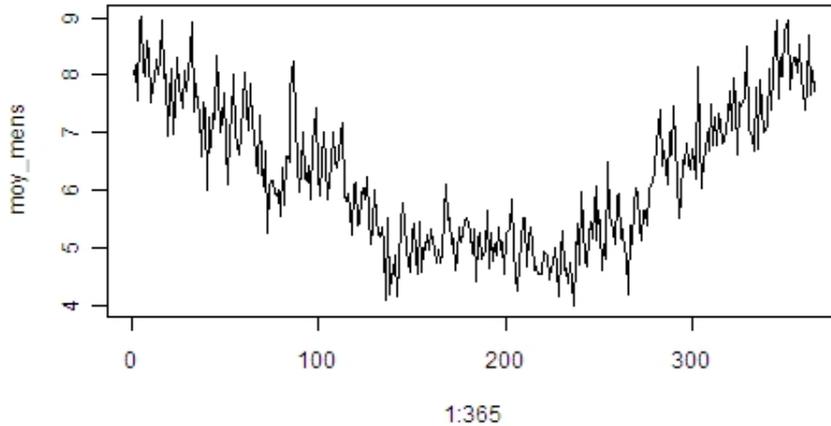
Première étape : traitement de la tendance (cf section 2). On suppose que la série suit un modèle multiplicatif, i.e $X_t = T_t Z_t$. On suppose en outre que $\log(T_t) = \alpha t + \beta_t$. On note a et b les estimateurs des moindres carrés de α et β .

Deuxième étape : traitement de la saisonnalité. On suppose que la série $\log(Z_t)$ suit un modèle de saisonnalité additif, c'est à dire que $\log(Z_t) = S_t + Y_t$. On estime S_t en calculant les moyennes saisonnières associés à la série $\log(Z_t)$. On a $T_{an} = 12$, données mensuelles.



Retour sur l'exemple des données de vent.

On suppose qu'il n'y a pas de tendance dans cette série temporelle. On a $N = 22 \times 365$ (22 ans de données) et $T_{an} = 365$ (données journalières). La séquence S_k représente alors la moyenne associée au jour k de l'année calculée sur les 22 ans de données.



Sur cet exemple, la fonction obtenue n'est pas un bon estimateur de la fonction S_k : les fluctuations rapides observées sont dues au fait qu'on a seulement 22 données pour estimer la valeur de \hat{s}_k pour un k donné. Dans ce cas, on peut lisser la série obtenue en utilisant les moyennes mobiles introduite précédemment. Remarque : ici on cherche à estimer une fonction périodique supposée être périodique à partir d'une observation "bruitée" x_t .

Modèle paramétrique

Lorsque l'on veut décrire la composante saisonnière grâce à un modèle paramétrique simple, on utilise généralement un polynôme trigonométrique. On suppose alors que :

$$X_t = \mu + \alpha_c \cos(\omega t) + \alpha_s \sin(\omega t) + Z_t$$

Pour estimer α_c et α_s , on peut utiliser la méthode des moindres carrés.

Proposition - Les estimateurs des moindres carrés sont solutions du système linéaire :

$$\begin{vmatrix} T & \sum_{t=1}^T \cos(\omega t) & \sum_{t=1}^T \sin(\omega t) \\ \sum_{t=1}^T \cos(\omega t) & \sum_{t=1}^T (\cos(\omega t))^2 & \sum_{t=1}^T \cos(\omega t) \sin(\omega t) \\ \sum_{t=1}^T \sin(\omega t) & \sum_{t=1}^T \sin(\omega t) \cos(\omega t) & \sum_{t=1}^T (\sin(\omega t))^2 \end{vmatrix} \begin{bmatrix} \mu \\ \alpha_c \\ \alpha_s \end{bmatrix} = \begin{vmatrix} x(t) \\ \sum_{t=1}^T x(t) \cos(\omega t) \\ \sum_{t=1}^T x(t) \sin(\omega t) \end{vmatrix}$$

Preuve : en exercice. Il suffit d'écrire le coût aux moindres carrés et de dériver.

Remarque : la solution analytique de ce système est relativement complexe (résolution numérique sous Matlab). Par contre, lorsque T est grand, une solution approchée est donnée par

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T x(t)$$

$$\hat{\alpha}_c = \frac{2}{T} \sum_{t=1}^T x(t) \cos(\omega t)$$

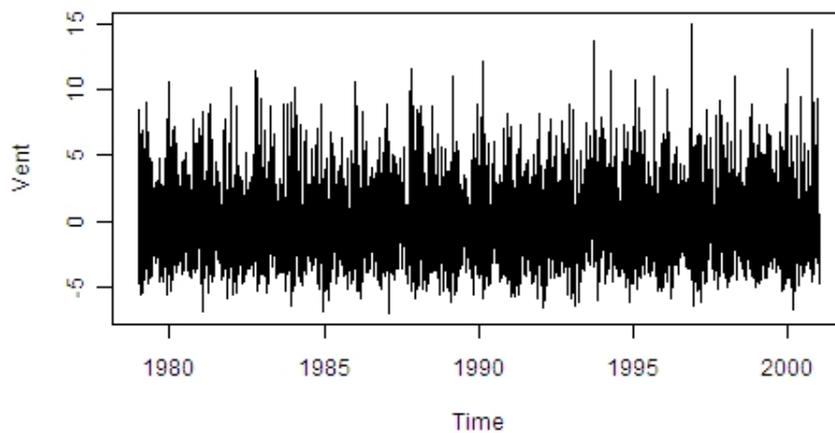
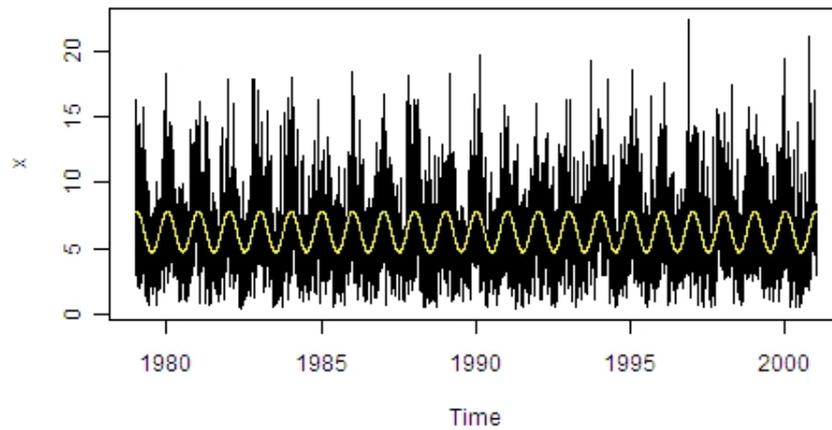
et

$$\hat{\alpha}_s = \frac{2}{T} \sum_{t=1}^T x(t) \sin(\omega t)$$

On reconnaît les coefficients de Fourier....

Ces estimateurs sont asymptotiquement sans biais et convergents.

La résolution du système précédent pour les données de vent donne $\hat{\mu} = 29.3$, $\alpha_c = 21,8$ et $\alpha_s = -7.7$.



Comparaison des différentes méthodes (p=2 mois)

Etude de la série résiduelle (modèle paramétrique) Il reste une forte saisonnalité. En fait, ici un modèle le modèle additif n'est pas adapté (cf TP?).

3.2.3 Différenciation

Définition - On appelle différence d'ordre τ de la série X_t les quantités

$$\nabla_t^{(\tau)} = X_{t+\tau} - X_t$$

définies pour $\pi \in \mathbb{N}$. Attention : ne pas confondre cet opérateur avec l'opérateur défini à la section 2 (modélisation de la tendance).

Si on suppose que la série $\{X_t\}$ suit un modèle de saisonnalité additif, c'est à dire que

$$X_t = S_t + Z_t$$

avec S_t une fonction périodique de période τ , avec τ le nombre de données par année, c'est à dire vérifiant $S_{t+\tau} = S_t$ et Z_t une série temporelle stationnaire (cf paragraphe 3). Alors il est facile de vérifier que $\nabla_\tau x_t = x_t - x_{t-\tau} = z_t - z_{t-\tau}$ est une série stationnaire (cf chapitre 3).

Retour sur l'exemple 2 (poissons).

3.2.4 Comparaison des différentes méthodes

Avantages des méthodes décrites au 2.a et 2.b sur la méthode du 2.c :

- Fournit un modèle facilement interprétable et qui permet, par exemple, de comparer plusieurs séries temporelles entre elles.
- Peut-être utilisé en prévision.

Les méthodes 2.a et 2.c s'ajustent à de nombreuses séries temporelles sans modification.

Chapter 4

Premiers pas vers la modélisation de la composante aléatoire

4.1 Généralités sur les processus stationnaires

4.1.1 Définitions

Définition 1. On appelle processus à temps discret une suite $\{X_1, \dots, X_T\}$ de variables aléatoires.

Remarque - Dans la suite, sauf mention contraire, on supposera que X_t est à valeurs réelles.

Définition 2. - Soit $X = \{X_t\}_{t \in \mathbb{Z}}$ un processus tel que $E[X_t^2] < \infty$ pour $t \in \mathbb{Z}$. On dira que ce processus est (faiblement) stationnaire (ou stationnaire d'ordre 2) si les 2 conditions suivantes sont vérifiées pour tout t :

- $E(X_t) = m$ pour tout $t \in \{1, \dots, T\}$
- $E(X_t X_{t+h}) = \gamma(h)$ est indépendant de t

On appellera alors fonction d'autocovariance de X_t la fonction C définie par

$$C(k) = E(X_t X_{t+k}) - E(X_t)E(X_{t+k})$$

et fonction d'autocorrélation la fonction ρ définie par

$$\rho(k) = \frac{E(X_t X_{t+k}) - E(X_t)E(X_{t+k})}{\sqrt{E(X_t^2)E(X_{t+k}^2)}}$$

.

Remarques :

1. On a $C(t) = C(-t)$ (fonction paire) donc il suffit de connaître la fonction d'autocovariance sur $[0, T]$. De même pour ρ .
2. Pour tout k , $|C(k)| \leq C(0)$.
3. La fonction d'autocovariance est définie positive.

Par la suite le terme stationnaire fera référence à la stationnarité faible.

4.1.2 Quelques commentaires sur l'autocorrélation

En général, on utilise l'autocorrélation pour caractériser les dépendances linéaires dans des séries résiduelles (ie des séries temporelles corrigées de la tendance et la saison). En effet, la tendance et la saison sont des composantes déterministes et ça a peu de sens d'estimer des propriétés statistiques de quantités déterministes. De plus, si la série étudiée a ses caractéristiques qui évoluent dans le temps, il peut être difficile d'estimer ses propriétés statistiques car on dispose en général d'une seule réalisation du processus ce qui n'est pas suffisant pour faire de l'estimation.

Cependant, il est très utile de comprendre quelle sera l'allure de l'auto-correlation empirique d'une série brute comportant une tendance et/ou une saison.

Proposition 1. *On considère une tendance linéaire pure*

$$x_i = ai + b, \quad i = 1, \dots, n,$$

pour h fixé, $\hat{\rho}(h)$ tend vers 1 quand n tend vers l'infini.

Preuve à faire en exercice : donner l'expression de la moyenne puis montrer que $\hat{\rho}(k)$ ne dépend ni de a ni de b . Enfin utiliser les résultats connus sur les sommes de puissances d'entiers successifs.

$$1 + 2 + \dots + N = \frac{N(N+1)}{2} \text{ et } 1 + 2^2 + \dots + N^2 = \frac{N(N+1)(2N+1)}{6}.$$

On admettra que le résultat est le même pour une tendance polynomiale.

Proposition 2. *On considère une série périodique pure de période T :*

$$x_i = a \cos\left(\frac{2\pi i}{T}\right), \quad i = 1, \dots, n,$$

pour h fixé, $\hat{\rho}(h)$ tend vers $\cos\left(\frac{2\pi h}{T}\right)$ quand n tend vers l'infini.

On admettra ce résultat.

En résumé, pour une longue série présentant une tendance polynomiale, l'autocorrélation empirique reste proche de un. Si une longue série comporte une saisonnalité, cette saisonnalité se voit sur la fonction d'autocorrélation empirique. Ainsi si une autocorrélation empirique présente ces caractéristiques, c'est qu'on n'a pas bien retiré la tendance et/ou correctement désaisonné la série.

4.2 Quelques exemples classiques de processus.

4.2.1 Bruit blanc

Définition 3. *On dira que le processus ϵ_t est un "bruit blanc" s'il forme une suite de variables indépendantes et identiquement distribuées. Il sera dit centré si $E[\epsilon_t] = 0$ et réduit si $Var[\epsilon_t] = 1$.*

Proposition 3. - *Soit X_t un bruit blanc vérifiant $E[X_t] = m$ et $Var[X_t] = \sigma^2$. Alors X_t est un processus stationnaire.*

On laisse la preuve en exercice.

La première condition signifie tout simplement que l'espérance du processus est indépendante du temps. La seconde condition implique bien entendu l'indépendance de la fonction d'autocovariance par rapport au temps (stationnarité). Mais elle implique en outre que les termes d'autocovariance (pour $h \geq 0$) sont tous nuls. Seule la variance est non nulle. Autrement dit, cela signifie que les bruits blancs sont des processus stationnaires particuliers sans "mémoire". Le niveau de la série considéré aujourd'hui n'a aucune incidence sur son niveau de demain, tout comme le niveau d'hier n'a aucune incidence sur le niveau d'aujourd'hui.

4.2.2 Processus gaussien

En statistique les vecteurs et les processus gaussien jouent un rôle particulier car ils sont faciles à manipuler. En particulier, ils sont complètement caractérisés par leur structure d'ordre deux. On peut donc facilement estimer et modéliser leurs propriétés statistiques.

Définition 4. *Le processus $(X_n)_{n \in \mathbb{Z}}$ est un processus gaussien si toutes ses lois marginales sont gaussiennes. Ca'est à dire si quelque soient k et j_1, \dots, j_k , le vecteur $(X_{j_1}, \dots, X_{j_k})$ est un vecteur gaussien.*

Il est clair qu'un processus gaussien stationnaire est strictement stationnaire.

Exercice - Soit $(X_n)_{n \in \mathbb{Z}}$ un suite de variables aléatoires gaussiennes indépendantes, centrées et toutes de variance 1. Parmi les processus suivants, lesquels sont gaussiens? Lesquels sont stationnaires? Donner alors l'expression de leur fonction de covariance.

1. $Y_n = X_n X_{n+1}$
2. $Y_n = X_n X_{n+1} \cdots X_{n+k}$
3. $Y_n = X_n^2 X_{n+1}^2$
4. $Y_n = X_n^2 + X_{n+1}^2$
5. $Y_n = n + X_n$
6. $Y_n = X_1 \cos(n\pi/2)$

4.2.3 Processus moyenne mobile

Définition 5. *On dira que le processus $\{Z_t\}$ suit un modèle "moyenne mobile d'ordre q " ($MA(q)$) s'il existe un bruit blanc centré $\{\epsilon_t\}$ et des constantes $\beta_k, k = 0, \dots, q$ tels que*

$$Z_t = \sum_{k=0}^q \beta_k \epsilon_{t-k}$$

Proposition 4. *Soit $\{X_t\}$ un processus suivant un modèle moyenne mobile, avec $\{\epsilon_t\}$ un bruit blanc centré vérifiant $E[\epsilon_t] = \sigma^2 < \infty$, alors le processus $\{X_t\}$ est stationnaire. Sa fonction d'autocovariance est donnée par*

$$C(h) = \sum_{k=0}^{q-|h|} \beta_k^2 \sigma^2 \text{ si } |h| \in 0, \dots, q \text{ et } 0 \text{ sinon}$$

Preuve : Un tel processus est stationnaire : écrire la moyenne et la covariance.

Théorème : Décomposition de Wold - Tout processus stationnaire d'ordre deux $(X_t; t \in \mathbb{Z})$ peut être représenté sous la forme :

$$X_t = \sum_{k=1}^{\infty} \beta_k \epsilon_k + \kappa_t$$

où les paramètres β_k satisfont $\beta_0 = 1$, $\beta_k \in \mathbb{R}$ pour $k > 0$, $\sum_{k=1}^{\infty} \beta_k^2 < \infty$ et où ϵ_t est un bruit blanc. On dit que la somme des chocs passés correspond à la composante linéaire stochastique de X_t . Le terme κ_t désigne la composante linéaire déterministe telle que $Cov(\kappa_t, \epsilon_{t-k}) = 0$ pour tout $k \in \mathbb{Z}$.

Ainsi, d'après le théorème de Wold, si l'on omet la composante déterministe κ_t , tout processus stationnaire peut s'écrire comme une somme pondérée infinie de chocs passés, ces chocs étant représentés par un bruit blanc de variance finie. L'implication forte de ce théorème est que, si l'on connaît les pondérations $\beta_k, k \in \mathbb{N}$, et si l'on connaît la variance du bruit blanc, on est mesure de proposer une représentation de n'importe quel processus stationnaire. Cette représentation est aussi qualifiée de représentation moyenne mobile infinie. Reste à comprendre ce que peut être cette composante linéaire déterministe κ_t . La condition $Cov(\kappa_t, \epsilon_{t-k}) = 0$ implique que ce terme est, par définition (déterministe), indépendant des chocs. Alors le cas le plus simple est celui d'un processus stationnaire $(X_t; t \in \mathbb{Z})$ d'espérance non nulle, tel que $E(X_t) = m$. Puisque le bruit blanc est par définition un processus centré, une somme pondérée de ces chocs est elle-même centrée. Par conséquent, la représentation de Wold du processus $(X_t; t \in \mathbb{Z})$ suppose que l'on ajoute à cette somme pondérée des chocs passés, une composante déterministe qui n'est autre que l'espérance du processus.

4.2.4 Processus autorégressif

Définition - On dira que le processus X_t suit un modèle autorégressif d'ordre p ($AR(p)$) s'il existe un bruit blanc centré réduit $\{\epsilon_t\}$, tel que ϵ_t soit indépendant de X_0, \dots, X_{t-1} et des constantes $\alpha_1, \dots, \alpha_p$ et σ tels que pour $t \in \{1, \dots, T\}$ on ait

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + \alpha_2(X_{t-2} - \mu) + \dots + \alpha_p(X_{t-p} - \mu) + \sigma \epsilon_t$$

Chapter 5

Auto-corrélation partielle

L'auto corrélation partielle est une notion importante pour les séries temporelles, mais elle est un peu délicate et nécessite des "rappels" de la géométrie dans un espace de variables aléatoires.

5.1 Rappels

Les variables aléatoires réelles de carré intégrable définies sur un même espace probabilité (Ω, \mathcal{A}, P) forment un espace vectoriel qu'on peut munir du produit scalaire $\langle X, Y \rangle = E[XY]$. On note cet espace $L^2 = L^2(\Omega, \mathcal{A}, P)$.

- *Orthogonalité.* Deux variables X et Y de L^2 sont orthogonales lorsque $E[XY] = 0$.
- *Distance.* On définit la distance entre deux variables comme

$$d(X, Y) = \sqrt{E[(X - Y)^2]}$$

- *Espace engendré.* On considère N variables aléatoires X_1, \dots, X_N . Le sous espace engendré par ces variables, qui sera noté $vec(X_1, \dots, X_N)$ est le sous ensemble de toutes les combinaisons linéaires de ces variables.

Projection Le projeté d'une variable aléatoire Y sur le sous espace $vec(X_1, \dots, X_N)$ sera noté

$$P_{vec(X_1, \dots, X_N)}(Y).$$

C'est la variable aléatoire de cet espace qui est la plus proche de Y au sens de la distance L^2 vue plus haut. C'est donc la combinaison linéaire $P_{vec(X_1, \dots, X_N)}(Y) = \gamma_1 X_1 + \dots + \gamma_N X_N$ telle que

$$E[(Y - \gamma_1 X_1 - \dots - \gamma_N X_N)^2] \leq E[(Y - \lambda_1 X_1 - \dots - \lambda_N X_N)^2] \quad \forall \lambda_1, \dots, \lambda_N.$$

On peut l'interpréter comme la meilleure explication linéaire de la variable Y en terme de X_1, \dots, X_N .

Pour calculer le projeté, on utilise la propriété suivante

Proposition 5. *Le projeté $P_{vec(X_1, \dots, X_N)}(Y) = \gamma_1 X_1 + \dots + \gamma_N X_N$ est caractérisé par le fait que $Y - P_{vec(X_1, \dots, X_N)}(Y)$ est orthogonal aux variables X_1, \dots, X_N .*

Ceci implique que

$$\forall k = 1, \dots, N, \quad E[(Y - \gamma_1 X_1 - \dots - \gamma_N X_N) X_k] = 0$$

Soit encore

$$\gamma_1 E(X_1 X_k) + \dots + \gamma_N E(X_N X_k) = E(Y X_k)$$

Ce qui s'écrit aussi sous la forme d'un système d'équations

$$\begin{pmatrix} E(X_1^2) & \dots & E(X_1 X_N) \\ \dots & \dots & \dots \\ E(X_1 X_N) & \dots & E(X_N^2) \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \dots \\ \gamma_N \end{pmatrix} = \begin{pmatrix} E(Y X_1) \\ \dots \\ E(Y X_N) \end{pmatrix}$$

Si la matrice est inversible, on obtient facilement les γ .

D'après la définition il est facile de comprendre que si Y est elle même une combinaison linéaire des variables X_1, \dots, X_n alors Y est son propre projeté. Par ailleurs si Y est orthogonale à toutes les variables X_j alors le projeté est nul. Et les variables X_j ne permettent pas d'expliquer Y .

5.2 Coefficient de corrélation partielle entre deux variables

On se donne $N \geq 3$ variables aléatoires X_1, \dots, X_N et on va définir un coefficient de corrélation entre X_1 et X_N qui fasse abstraction de l'influence de X_2, \dots, X_{N-1} .

Cette notion est une réponse à la question suivante : il arrive que 2 phénomènes soient fortement corrélés, mais que cette corrélation soit due à l'influence d'un facteur extérieur et non pas à un fort lien entre les deux phénomènes. Exemples :

- Corrélation entre les ventes de lunettes noires et les ventes de glaces ; effet de l'ensoleillement.
- Corrélation entre le nombre d'admissions à l'hôpital et les ventes de glaces : canicules.
- Corrélation entre la longueur des cheveux et la taille des personnes. On a mélangé les hommes et les femmes dans les données et les hommes sont en moyenne plus grands que les femmes.

Définition 6. *Le coefficient de corrélation partielle en X_1 et X_N abstraction faite de l'influence de X_2, \dots, X_{N-1} est le coefficient de corrélation entre les deux variables auxquelles on a retranché leurs meilleures explications en terme de X_2, \dots, X_{N-1} , soit*

$$r_{X_2, \dots, X_{N-1}}(X_1, X_N) = \rho(X_1 - P_{X_2, \dots, X_{N-1}}(X_1), X_N - P_{X_2, \dots, X_{N-1}}(X_N))$$

Exercice - Considérons trois variables indépendantes X, Z_1, Z_2 centrées et de variance 1. On construit les variables

$$X_1 = X + Z_1 \quad \text{et} \quad X_2 = X + Z_2.$$

Calculer la corrélation entre X_1 et X_2 puis l'auto corrélation partielle $r_{[X]}(X_1, X_2)$ entre X_1 et X_2 . Commentez.

5.3 Autocorrélation partielle d'un processus stationnaire

On considère maintenant un processus stationnaire centré $(\dots, X_{-1}, X_0, X_1, \dots)$.

5.3.1 Définition

On définit l'auto-corrélation partielle $r(h)$ pour $h \neq 0$ de la façon suivante.

Définition 7.

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

Exemple de l'AR(1) - Soit X un processus centré tel que $X_t = 0.9X_{t-1} + \epsilon_t$ avec ϵ un bruit blanc gaussien centré et réduit et ϵ_t indépendant de X_s pour tout $s < t$.

On calcule tout d'abord la fonction d'auto-corrélation.

$$\text{corr}(X_1, X_2) = \text{corr}(0.9X_1 + \epsilon_2, X_1) = 0.9$$

Intéressons nous maintenant aux projections sur le passé et/ou le futur.

$$P_{[X_2, \dots, X_{k-1}]}(X_k) = 0.9X_{(k-1)} \text{ et } P_{[X_2, \dots, X_{k-1}]}(X_1) = 0.9X_1.$$

Il est facile de vérifier que $P_{[X_2, \dots, X_{k-1}]}(X_k) = P_{[X_2, \dots, X_{k-1}]}(0.9X_{(k-1)} + \epsilon_k)$. Or ϵ_k est, par définition, indépendant du passé. On a donc bien $P_{[X_2, \dots, X_{k-1}]}(X_k) = 0.9X_{(k-1)}$.

Considérons maintenant, $P_{[X_2, \dots, X_{k-1}]}(X_1)$. Dans le cas où $k = 3$, on écrit $P_{[X_2]}(X_1) = \gamma X_2$ avec γ tel que

$$E((X_1 - \gamma X_2)^2)$$

soit minimum. On peut calculer la dérivée de $E((X_1 - \gamma X_2)^2)$ par rapport à γ : $\rho(1) - \gamma\rho(0)$. Et en posant que cette dérivée est nulle, on obtient $\gamma = .9$.

$$\text{corr}(X_k - 0.9X_{k-1}, X_1 - 0.9X_2) = \text{corr}(\epsilon_k, X_1 - 0.9X_2) = 0$$

car ϵ_k est indépendant de x_1 et de X_2

De façon générale, pour un processus stationnaire, l'auto-corrélation partielle s'écrit

$$r(h) = \text{corr}\left(X_n - \sum_{j=1}^{h-1} \gamma_j X_{n+j}, X_{n+h} - \sum_{j=1}^{h-1} \gamma_{h-j} X_{n+j}\right)$$

avec

$$\begin{pmatrix} C(0) & \cdots & C(h-2) \\ \cdots & \cdots & \cdots \\ C(h-2) & \cdots & C(0) \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \cdots \\ \gamma_{h-1} \end{pmatrix} = \begin{pmatrix} C(1) \\ \cdots \\ C(h-1) \end{pmatrix}$$

5.3.2 Algorithme de Durbin-Watson

Cet algorithme itératif permet de calculer rapidement les fonctions d'auto-corrélation partielle.

Proposition 6. - Si le processus $(X_n)_{n \in \mathbb{Z}}$ est stationnaire, pour tout $h \geq 2$, $r(h)$ est le coefficient de X_n dans l'expression du projeté de X_{n+h} sur l'espace engendré par (X_n, \dots, X_{n+h-1}) .

Partant de ce résultat, l'algorithme se construit de la façon suivante.

$$P_{[X_n, \dots, X_{n+h-1}]} = \sum_{j=0}^{h-1} \gamma_{h-j}^{(h)} X_j$$

et d'après la proposition, $r(h) = \gamma_h^{(h)}$. Par ailleurs, par définition du projeté,

$$\begin{pmatrix} \gamma_1^{(h)} \\ \dots \\ \gamma_h^{(h)} \end{pmatrix} = \begin{pmatrix} C(0) & \dots & C(h-1) \\ \dots & \dots & \dots \\ C(h-1) & \dots & C(0) \end{pmatrix}^{-1} \begin{pmatrix} C(1) \\ \dots \\ C(h) \end{pmatrix}$$

Chapter 6

Modèles Auto-régressifs

6.1 Modèles autorégressifs d'ordre 1

Quitte à considérer le processus $X_t - \mu$, on peut supposer que $E[X_t] = 0$. On suppose donc que X_t est centré.

Notons

$$X_t = \alpha_1 X_{t-1} + \epsilon_t$$

avec ϵ_t un bruit blanc centré de variance σ^2 . On vérifie aisément que

$$X_t = \epsilon_t + \alpha_1 \epsilon_{t-1} + \dots + \alpha_1^k \epsilon_{t-k} + X_{t-k-1}$$

Ainsi, si $|\alpha_1| < 1$ et X_t stationnaire alors $\|X_t\|^2 = E(X_t^2)$ est constant et

$$\left\| X_t - \sum_{j=0}^k \alpha_1^j \epsilon_{t-j} \right\|^2 = \alpha_1^{2k+2} \|X_{t-k-1}\|^2 \rightarrow 0 \text{ quand } k \rightarrow \infty$$

Comme $\sum_{j=0}^{\infty} |\alpha_1^j| \epsilon_{t-j}$ est convergent en moyenne quadratique, X_t admet une représentation en moyenne mobile

$$X_t = \sum_{j=0}^{\infty} \alpha_1^j \epsilon_{t-j}$$

On en déduit que $E[X_t] = 0$ pour tout t et

$$\begin{aligned} \text{Cov}(X_{t+k} X_t) &= \lim_{n \rightarrow \infty} E \left[\left(\sum_{j=0}^n \alpha_1^j \epsilon_{t+k-j} \right) \left(\sum_{j=0}^n \alpha_1^j \epsilon_{t-j} \right) \right] \\ &= \sigma^2 \alpha^{2|k|} \sum_{j=0}^{\infty} \alpha_1^{2j} \\ &= \sigma^2 \frac{\alpha^{2|k|}}{1 - \alpha_1^2} \end{aligned}$$

Si $|\alpha_1| > 1$ la série $\sum_{j=0}^{\infty} |\alpha_1^j| \epsilon_{t-j}$ n'est pas convergent dans L^2 . Et il ne peut exister de solution stationnaire (processus explosif).

Rappel: somme d'une suite géométrique $s_n = \sum_{k=0}^{n-1} q^k = \frac{1-q^{n+1}}{1-q} \rightarrow \frac{1}{1-q}$ si $|q| < 1$.

Théorème 1. *Le modèle défini par l'équation $X_t = \alpha X_{t-1} + \sigma \epsilon_t$ possède une solution stationnaire si et seulement si $|\alpha| < 1$. Dans ce cas, la solution stationnaire vérifie $E[X_t] = \alpha x_0$ et $E[X_t^2] = \frac{\sigma^2}{1-\alpha^2}$ et sa fonction d'auto-covariance est donnée par $C(k) = \frac{\sigma^2}{1-\alpha^2} \alpha^{|k|}$ avec $|k| > 0$.*

On remarque que $\alpha^k = e^{k \log(\alpha)}$. La décroissance de la fonction d'auto-covariance d'un processus auto-régressif d'ordre 1 est donc exponentielle.

Preuve du théorème

Stationnarité : il suffit de remarquer que pour tout $t \in \mathbb{Z}$,

$$X_t = \epsilon_t + \alpha \epsilon_{t-1} + \alpha^2 \epsilon_{t-2} + \dots$$

ainsi si la série $\sum_{k=1}^{\infty} \alpha^k$ est convergente on a un processus stationnaire dont le moment d'ordre 2 est fini.

Considérons X_t soit stationnaire et tel que $E[X_t] = 0$ et $E[X_t^2] = E[X_{t'}^2]$. On a alors

$$E[X_t^2] = \alpha^2 E[X_t^2] + \sigma^2$$

donc

$$E[X_t^2] = \frac{\sigma^2}{1-\alpha^2}.$$

On a de plus

$$C(k) = E[X_t X_{t-k}] = \alpha E[X_{t-1} X_{t-k}] = \alpha^2 E[X_{t-2} X_{t-k}].$$

On montre ainsi, par itération, que $C(k) = \frac{\sigma^2}{1-\alpha^2} \alpha^k$.

6.2 Modèles auto-régressifs d'ordre p

Théorème 2. - *Le modèle défini par l'équation*

$$\begin{aligned} X_t &= \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \sigma^2 \epsilon_t \\ X_0 &= 0 \end{aligned}$$

pour $t \geq p$ possède une solution stationnaire si et seulement si les racines (complexes) de son polynôme caractéristique

$$1 - (\alpha_1 x + \dots + \alpha_p x^p)$$

sont de module strictement supérieur à 1.

Remarque : modèle d'ordre $p = 1$: on retrouve la condition du théorème précédent.

On montre facilement que le processus est stationnaire en écrivant sa moyenne et sa covariance.

Exercice - Donner les conditions de stationnarité d'un processus AR(2).

Idée de la solution ON veut que les racines du polynôme caractéristique soient plus grandes que 1 en module :

$$\left| \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2} \right| > 1$$

Or on a

$$\begin{aligned} G_1 &= \frac{2\phi_2}{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}} \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}} \\ &= \frac{2(\phi_1 + \sqrt{\phi_1^2 + 4\phi_2})}{\phi_1^2 - \phi_1^2 + 4\phi_2} = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2} \end{aligned}$$

SI les racines sont réelles, $|G_1| < 1$ implique

$$\begin{aligned} \phi_1 + \sqrt{\phi_1^2 + 4\phi_2} &< 2 \\ \phi_1^2 + 4\phi_2 &< 4 - 4\phi_1 + \phi_1^2 \\ \phi_1 + \phi_2 &< 1 \end{aligned}$$

On a aussi $\phi_2 - \phi_1 < 1$ par l'autre racine.

Si les racines sont complexes, elles sont conjuguées et la condition s'écrit donc $|G_1| = |G_2| < 1$. En résolvant $|G_1|^2 < 1$, on obtient la condition $\phi_2 > -1$.

Proposition - Soit X_t un processus stationnaire suivant un modèle $AR(p)$ de paramètres $\alpha_1, \dots, \alpha_p, \sigma^2$. On a alors $E[X_t^2] = \sum_{k=1}^p \alpha_k C(-k) + \sigma^2$ et la fonction d'autocovariance $C(k)$ vérifie les relations ci-dessous (équations de Yule Walker) :

$$\begin{bmatrix} C(1) \\ C(2) \\ C(3) \\ \vdots \end{bmatrix} = \begin{bmatrix} C(0) & C(-1) & C(-2) & \cdots & 1 \\ C(1) & C(0) & C(-1) & \cdots & 0 \\ C(2) & C(1) & C(0) & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \end{bmatrix}$$

Démonstration

L'équation définissant le processus AR est

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

En multipliant les deux membres par X_{t-j} et en prenant l'espérance, on obtient

$$E[X_t X_{t-j}] = E \left[\sum_{i=1}^p \varphi_i X_{t-i} X_{t-j} \right] + E[\varepsilon_t X_{t-j}]$$

Or, il se trouve que $E[X_t X_{t-j}] = C(j)$ par définition de la fonction d'autocovariance. Les termes du bruit blanc sont indépendants les uns des autres et, de plus, X_{t-j} est indépendant de ε_t où j est plus grand que zéro. Pour $j > 0$, $E[\varepsilon_t X_{t-j}] = 0$. Pour $j = 0$,

$$E[\varepsilon_t X_t] = E \left[\varepsilon_t \left(\sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t \right) \right] = \sum_{i=1}^p \varphi_i E[\varepsilon_t X_{t-i}] + E[\varepsilon_t^2] = 0 + \sigma_\varepsilon^2$$

Maintenant, on a pour $j \geq 0$,

$$C(j) = E \left[\sum_{i=1}^p \varphi_i X_{t-i} X_{t-j} \right] + \sigma_\varepsilon^2 \delta_j$$

Par ailleurs,

$$E \left[\sum_{i=1}^p \varphi_i X_{t-i} X_{t-j} \right] = \sum_{i=1}^p \varphi_i E[X_t X_{t-j+i}] = \sum_{i=1}^p \varphi_i C(j-i)$$

qui donne les équations de Yule-Walker:

$$C(j) = \sum_{i=1}^p \varphi_i C_{j-i} + \sigma_\epsilon^2 \delta_0(j)$$

pour $j \geq 0$. Pour $j < 0$,

$$C(j) = C(-j) = \sum_{i=1}^p \varphi_i C(|j| - i) + \sigma^2 \delta_0(j)j$$

Remarque : Ces formules peuvent être utilisées pour calculer récursivement la fonction d'autocovariance.

Exercice : On considère le modèle AR(2) défini par

$$X_t = 0.9X_{t-1} - 0.3X_{t-2} + 0.1\epsilon_t$$

- Vérifier que ce processus admet une solution stationnaire
- Calculer les valeurs de $E[X_t]$, $C(1), \dots, C(5)$ du modèle précédent.

6.3 Inférence statistique pour les modèles auto-régressifs

6.3.1 Estimation des paramètres

On dispose d'une observation $\{x_0, \dots, x_T\}$ de longueur $T + 1$ d'un processus stationnaire X_t supposé suivre un modèle AR(p), c'est à dire vérifiant

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \sigma \epsilon_t$$

avec $t \in \mathbb{Z}$ et $X_0 = x_0$ et $\alpha_1, \dots, \alpha_p, \sigma$ des paramètres inconnus. On cherche alors à estimer ces paramètres à l'aide des observations disponibles.

Pour cela différentes méthodes sont possibles

- Moindres carrés (non étudié)
- Maximum de vraisemblance (non étudié). L'estimation d'un modèle AR(P) par la méthode du maximum de vraisemblance est délicate car la fonction de vraisemblance est très complexe et n'a pas de dérivée analytique. Cette difficulté provient de l'interdépendance des valeurs, ainsi que du fait que les observations antérieures ne sont pas toutes disponibles pour les p premières valeurs.
- Méthode des moments : utilisation des équations de Yule-Walker.

6.3.2 Estimateurs de Yule-Walker

La méthode consiste à reprendre les équations de Yule-Walker en inversant les relations : on exprime les coefficients en fonction des autocovariances. On applique alors le raisonnement de la méthode des moments : on trouve les paramètres estimés d'après les auto-covariances estimées.

On a vu précédemment que $\mu = E[X_1]$, $C(0) = \alpha_1 C(1) + \dots + \alpha_p C(p) + \sigma^2$ et

$$C(k) = \alpha_1 C(k-1) + \dots + \alpha_p C(k-p)$$

En prenant l'équation sous sa forme matricielle

$$\begin{bmatrix} C(0) \\ C(1) \\ C(2) \\ C(3) \\ \vdots \end{bmatrix} = \begin{bmatrix} C(-1) & C(-2) & C(-3) & \dots & 1 \\ C(0) & C(-1) & C(-2) & \dots & 0 \\ C(1) & C(0) & C(-1) & \dots & 0 \\ C(2) & C(1) & C(0) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \sigma_\varepsilon^2 \end{bmatrix}$$

que l'on écrira aussi

$$R_p \alpha = C$$

avec R_p la matrice et $C = (C(i))_{i=0, \dots, p}$ le second membre. Le vecteur des paramètres $\hat{\theta} = \begin{pmatrix} \hat{\alpha}_1 \\ \vdots \\ \hat{\sigma}_\varepsilon^2 \end{pmatrix}$

peut alors être obtenu.

Remarque - Les estimateurs de Yule-Walker sont tels que la moyenne et les $p + 1$ premières valeurs de la fonction d'auto-covariance du modèle estimé coïncident avec celles estimées sur les données.

Cas particuliers :

Modèles d'ordre $p = 1$. On obtient

$$\hat{R}_1 = \begin{bmatrix} \hat{C}(1) & 1 \\ \hat{C}(0) & 0 \end{bmatrix}$$

alors $\alpha_1 = \hat{C}(1)/\hat{C}(0)$ et $\hat{\sigma}^2 = \hat{C}(0) - \hat{C}(1)^2/\hat{C}(0)$

Modèles d'ordre $p = 2$. On obtient

$$\hat{R}_2 = \begin{bmatrix} \hat{C}(1) & \hat{C}(2) & 1 \\ \hat{C}(0) & \hat{C}(1) & 0 \\ \hat{C}(1) & \hat{C}(0) & 0 \end{bmatrix}$$

alors

$$\hat{R}_2^{-1} = \frac{1}{\hat{C}(0)^2 - \hat{C}(1)^2} \begin{bmatrix} 0 & \hat{C}(0) & \hat{C}(1) \\ 0 & -\hat{C}(1) & \hat{C}(0) \\ \hat{C}(0)^2 - \hat{C}(1)^2 & \hat{C}(1)(\hat{C}(2) - \hat{C}(0)) & \hat{C}(1)^2 - \hat{C}(0)\hat{C}(2) \end{bmatrix}$$

donc $\alpha_1 = (\hat{C}(0)\hat{C}(1) + \hat{C}(1)\hat{C}(2))/(\hat{C}(0)^2 - \hat{C}(1)^2)$, etc.

Les estimateurs de Yule-Walker sont évidemment consistants dès que les $\hat{C}(k)$ sont des estimateurs consistants des covariances $C(k)$. Le résultat suivant donne plus de précisions sur le

comportement asymptotique de ces estimateurs. Il reste vrai même dans le cas où on aurait résolu le système des équations de Yule-Walker avec une valeur de p qui ne serait pas la bonne.

Proposition 7. Soit (X_t) un processus $AR(p)$ stationnaire, soit $q \geq p$ et notons $\alpha \in \mathbb{R}^q$ le vecteur $\alpha = (\alpha_1, \dots, \alpha_p, 0, \dots, 0)$. Notons $\hat{\alpha}^{(n)}$ la solution de l'équation de Yule-Walker empirique dans \mathbb{R}^q soit $R_m^{(n)} \hat{\alpha}^{(n)} = C_{X,m}^{(n)}$. Alors

$$\sqrt{n}(\hat{\alpha}^{(n)} - \alpha) \underset{n \rightarrow \infty}{\mathcal{L}} N(0, \sigma^2 R_q^{-1})$$

(6.7)

Remarque - La matrice de covariance dépend de σ^2 de façon linéaire (est égale à σ^2 multiplié par la matrice de covariance pour un bruit de variance 1). Donc la matrice de covariance asymptotique $\sigma^2 R_m^{-1}$ ne dépend pas de σ^2 .

La proposition précise la variance asymptotique des estimateurs de Yule-Walker. Ces estimateurs ont la même variance asymptotique que ceux du maximum de vraisemblance. Mais, évidemment, ils ne sont utiles que lorsque on sait déjà que l'on a affaire avec un processus AR (et donc que la partie MA est nulle).

Dans le cadre d'un processus AR , la proposition fournit une approche à l'étude de la meilleure valeur de p . En effet, si $m > p$ et $\hat{\alpha}^{(n)} = (\hat{\alpha}_1^{(n)}, \dots, \hat{\alpha}_m^{(n)})$, (6.7) affirme que $\sqrt{n} \hat{\alpha}_m^{(n)} \rightarrow N(0, \lambda)$ pour $n \rightarrow \infty$, où λ est égal à σ^2 multiplié par l'élément diagonal d'indice m, m de la matrice R_m^{-1} . Or il est remarquable que cette quantité vaut toujours 1.

Remarque - ces formules peuvent être utilisées pour construire des intervalles de confiance asymptotiques en remplaçant σ et R_p par leurs estimations. Plus précisément, pour T grand, on a $\sqrt{T}(\hat{\alpha} - \alpha) \sim \mathcal{N}(0, \Sigma)$, donc en particulier $\sqrt{T}(\hat{\alpha}_i - \alpha_i) \sim \mathcal{N}(0, \Sigma_{ii})$ avec Σ_{ii} le i ème terme diagonal de la matrice $\Sigma = \sigma^2 R_p^{-1}$, puis

$$P(-q_{1-\alpha} \sqrt{\Sigma_{ii}} \leq \sqrt{T}(\hat{\alpha}_i - \alpha_i) \leq q_{1-\alpha} \sqrt{\Sigma_{ii}}) = 1 - \alpha$$

avec $q_{1-\alpha}$ le quantile d'ordre $1 - \alpha$ de la loi de Gauss centrée et réduite.

Exercice

On considère une série temporelle $\{x_t\}$ de longueur $T = 200$ telle que $\sum x_t = 13.1, \sum x_t^2 = 164.2, \sum x_t x_{t+1} = 6.1, \sum x_t x_{t+2} = -9.98, \sum x_t x_{t+3} = 13.4, \sum x_t x_{t+4} = 11.8$.

1. On suppose que cette série temporelle suit un modèle d'ordre 1

- a. Ecrire les équations de Yule-Walker
- b. En déduire une estimation des paramètres du modèle
- c. Donner un intervalle de confiance pour le paramètre α_1

2. On suppose que cette série temporelle suit un modèle d'ordre 2.

- a. Ecrire les équations de Yule-Walker
- b. En déduire une estimation des paramètres du modèle
- c. Donner un intervalle de confiance pour le paramètre α_1 puis pour α_2 .

6.3.3 Maximum de vraisemblance conditionnel

Une manière de simplifier la complexité de la fonction de vraisemblance est de conditionner cette fonction aux p premières observations. La fonction de log-vraisemblance devient:

$$\begin{aligned} \log \mathcal{L}(x_1, \dots, x_T) &= \frac{-(T-p)}{2} \log(2\pi) - \frac{-(T-p)}{2} \log(\sigma^2) \\ &\quad - \sum_{t=p+1}^T \frac{1}{2\sigma^2} (x_t - \mu - \alpha_1 x_{t-1} - \dots - \alpha_p x_{t-p})^2 \end{aligned}$$

La maximisation de cette fonction par rapport aux paramètres α correspond à la minimisation des erreurs du modèle. L'estimateur du maximum de vraisemblance conditionnel correspond ainsi à celui des moindres carrés. L'estimateur obtenu sera équivalent à l'estimateur inconditionnel dans de grands échantillons et tous deux ont la même distribution asymptotique (Hamilton 1994, p. 126). Il peut être biaisé.

6.4 Prédiction dans les modèles autorégressifs

Dans ce paragraphe, on suppose que $\{X_t\}$ est un processus stationnaire qui suit un modèle $AR(p)$, c'est à dire vérifie

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + \dots + \alpha_p(X_{t-p} - \mu) + \sigma \epsilon_t$$

avec ϵ_t un bruit blanc centré réduit.

Objectif: on cherche à prédire la valeur prise par le processus aux instants $t+1, t+2, \dots$ à partir de la connaissance des valeurs prises par ce processus jusqu'à l'instant t , c'est à dire de x_0, \dots, x_t .

Le modèle se réécrit sous la forme

$$X_t = \alpha_1(X_{t-1} - \mu) + \dots + \alpha_p(X_{t-p} - \mu) + \mu + \sigma \epsilon_t$$

En général, on utilise les quantités suivantes:

$$\hat{X}_{t+1|t} = \alpha_1(X_t - \mu) + \dots + \alpha_p(X_{t-p+1} - \mu) + \mu$$

pour prédire X_{t+1} à partir de X_1, \dots, X_t .

$$\hat{X}_{t+2|t} = \alpha_1(\hat{X}_{t+1|t} - \mu) + \dots + \alpha_p(X_{t-p+2} - \mu) + \mu$$

Et de façon générale

$$\hat{X}_{t+k|t} = \alpha_1(\hat{X}_{t+k-1|t} - \mu) + \dots + \alpha_p(X_{t+k-p|t} - \mu) + \mu$$

Remarque - Dans le cas des modèles d'ordre 1, on a $\hat{X}_{t+1|t} - \mu = \alpha_1(X_t - \mu)$, $\hat{X}_{t+2|t} - \mu = \alpha_1(\hat{X}_{t+1|t} - \mu) = \alpha_1^2(X_t - \mu)$, \dots . On vérifie aisément par récurrence que $\hat{X}_{t+k|t} - \mu = \alpha_1^k(X_t - \mu)$, donc en particulier que $\hat{X}_{t+k|t} \rightarrow \mu$ quand k tend vers l'infini.

Proposition 8. (*admise*) - Si $\{X_t\}$ est un processus stationnaire qui suit un modèle $AR(p)$, alors la meilleure prédiction, au sens des moindres carrés, de X_{t+1} connaissant x_t, \dots, x_{t-p+1} est donnée par \hat{x}_t défini par

$$\hat{x}_{t+1|t} = \alpha_1 x_t + \dots + \alpha_p x_{t-p+1} + \mu$$

Qualité de la prédiction

Prédiction à un pas de temps

On a par définition $\hat{X}_{t+1|t} - X_{t+1} = \sigma\epsilon_{t+1}$. ϵ_{t+1} représente donc l'erreur de prédiction à un pas de temps. En particulier, l'estimation est non-biaisée et la variance de l'erreur d'estimation est σ^2 . En général, afin de construire des intervalles de prédiction, on est amené à supposer que ϵ suit une loi de Gauss centrée et réduite. On en déduit alors que $\hat{X}_{t+1|t} - X_{t+1}$ suit aussi une loi de Gauss centrée et de variance σ^2 . L'intervalle $[\hat{X}_{t+1|t} - \sigma\Phi^{-1}(1 - .95/2), \hat{X}_{t+1|t} + \sigma\Phi^{-1}(1 - .95/2)]$ est appelé intervalle de prédiction à 95%. En pratique, σ et $\hat{X}_{t+1|t}$ sont inconnus et on les estime.

Prédiction à deux pas de temps

On a par définition

$$\hat{X}_{t+2|t} - X_{t+2} = \sigma\epsilon_{t+2} + \alpha_1\sigma\epsilon_{t+1}$$

En particulier, l'estimation est non-biaisée et la variance de l'erreur d'estimation est $(1 + \alpha_1^2)\sigma^2$. On peut aussi construire des IP.

Remarque - L'extension au cas général est complexe, sauf dans le cas des modèles d'ordre $p = 1$. On a alors (donner formule avec $k=1$, $k=2$, puis $k=3$).

$$\hat{X}_{t+k|t} - X_{t+k} = \sigma\epsilon_{t+k} + \alpha_1\sigma\epsilon_{t+k-1} + \dots + \alpha_1^{k-1}\sigma\epsilon_{t+1}$$

En particulier, on a l'estimation est non-biaisée et la variance de l'erreur d'estimation est $\sigma^2 \sum_{j=0}^{k-1} \alpha_1^{2j}$.

On peut aussi construire des IP dans le cas gaussien :

$$[\hat{X}_{t+k|t} - \sigma\Phi^{-1}(1 - .95/2)\sqrt{1 + \alpha_1^2 + \dots + \alpha_{k-1}^2}, \hat{X}_{t+k|t} + \sigma\Phi^{-1}(1 - .95/2)\sqrt{1 + \alpha_1^2 + \dots + \alpha_{k-1}^2}]$$

est l'intervalle de prédiction de niveau 0.95 pour X_{t+h} lorsqu'on a observé X jusqu'au temps t . On vérifie aisément que la qualité de la prédiction se dégrade lorsque k augmente.

6.5 Sélection de modèle

Jusqu'à présent, nous avons supposé que l'ordre du modèle AR est connu. Cependant, en pratique, cette quantité est généralement inconnue, et on cherche alors à estimer sa valeur à partir des observations, ce qu'on appelle généralement le problème de la sélection de modèle.

Une première méthode consiste à réaliser des tests statistiques. En pratique on considère pour p variant de 1 à P les tests successifs permettant de comparer le modèle d'ordre p au modèle d'ordre $p + 1$.

$$H_0 : X \text{ suit un modèle } AR(p) \text{ contre } H_1 : \text{non } H_0$$

Par le théorème de normalité asymptotique des équations de Yule-Walker (voir section 3.1.1), on a directement, pour un processus $AR(p)$, un théorème central limite pour les pacf empiriques $\hat{r}(h)$, $h > p$:

$$\sqrt{T}\hat{r}(h) \rightarrow \mathcal{N}(0, 1) \text{ quand } T \rightarrow \infty$$

Ainsi sous H_0 , la probabilité que $\hat{r}(h) < \sqrt{1/T}\Phi^{-1}(1 - \alpha/2)$ est égale à α .

Une autre idée consiste à utiliser le modèle qui minimise l'erreur de prédiction à un pas de temps, c'est à dire la valeur de $\hat{X}_{t+1|t} - x_{t+1}$. Cependant, en pratique, plus on augmente la valeur de

p , plus l'erreur de prédiction diminue. On peut alors utiliser le critère Bayesian Information Criterion (BIC). On choisit alors le modèle qui minimise la quantité

$$BIC(p) = -2 \log(\mathcal{L}) + N_{par}(p) \log(T)$$

On obtient alors généralement un modèle pour lequel on a un bon compromis entre le nombre de paramètre et l'erreur de prédiction. En pratique, on obtient généralement des modèles parsimonieux qui s'ajuste bien aux données.

6.6 Validation de modèle

Lorsque l'on veut vérifier l'adéquation d'un modèle auto-régressif à des données observées, on utilise généralement le test de Portmanteau. Supposons tout d'abord que l'on veuille tester : H_0 : est un bruit blanc

H_1 : n'est pas un bruit blanc

(Attention : portemanteau signifie fourre-tout en anglais). Ce test évalue la corrélation existant entre les résidus.

Soit $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$ une famille de variables aléatoires de loi à densité. On souhaite encore tester H_0 : ε est un bruit blanc fort. On note

$$\hat{\rho}_\varepsilon(h) = \frac{\frac{1}{n-h} \sum_{i=1}^{n-h} \varepsilon_i \varepsilon_{i+h}}{\frac{1}{n} \sum_{i=1}^n \varepsilon_i^2}$$

Il s'agit d'un estimateur de la corrélation. En effet, de même que $\frac{1}{n} \sum_{i=1}^n \varepsilon_i^2$ est un estimateur de la variance $\text{var}(\varepsilon_t) = \mathbb{E}[\varepsilon_t^2]$, $\frac{1}{n-h} \sum_{i=1}^{n-h} \varepsilon_i \varepsilon_{i+h}$ est un estimateur de $\text{cov}(\varepsilon_t, \varepsilon_{t+h}) = \mathbb{E}[\varepsilon_t \varepsilon_{t+h}]$. Notons que sous H_0 cette covariance est nulle si $h \neq 0$.

Sous H_0 , pour n grand et $k < n/4$, $(\hat{\rho}_\varepsilon(h), h = 1, \dots, k)$ est approximativement un bruit blanc gaussien de variance $1/n$. Il s'ensuit que

$$Q_{BP} = n \sum_{h=1}^k \hat{\rho}_\varepsilon^2(h)$$

suit approximativement une loi du χ^2 à k degrés de liberté. Une trop grande valeur de Q_{BP} indiquerait une certaine corrélation entre les ε_i . On en déduit un test au niveau α de H_0 de zone de rejet $Q_{BP} > \chi_{1-\alpha}^2(k)$, avec $\chi_a^2(h)$ le quantile d'ordre a de la loi du χ^2 à k degrés de liberté. En fait, on peut améliorer sensiblement ce test en considérant plutôt la statistique

$$Q_{LB} = n(n+2) \sum_{h=1}^k \frac{\hat{\rho}_\varepsilon^2(h)}{n-h},$$

qui, sous H_0 , suit également, pour n grand, une loi du χ^2 à k degrés de liberté.

Ce test, sous sa première forme, est aussi appelé test de Box-Pierce ; sous sa seconde, il est connu sous le nom de test de Ljung-Box.

Chapter 7

Les processus moyenne mobile

7.1 Les processus moyenne mobile

Définition 8. *Etant donné un polynôme $B(z) = 1 + b_1z + \dots + b_qz^q$ de degré exactement q ($b_q \neq 0$) et dont toutes les racines sont de module strictement supérieur à 1, et étant donné un bruit blanc centré $(\epsilon_t)_{t \in \mathbb{Z}}$, on appelle processus moyenne mobile d'ordre q (on note $MA(q)$, Moving Average) de polynôme caractéristique B et d'innovation ϵ_n , le processus défini par*

$$X_t = \epsilon_t + b_1\epsilon_{t-1} + \dots + b_q\epsilon_{t-q}, \quad \forall n$$

Quelque soit la position des racines du polynôme B et processus $MA(q)$ est stationnaire, centré et on a

$$E(X_t \epsilon_{t+h}) = 0 \quad \forall h > 0$$

L'hypothèse sur les racines du polynôme permet de montrer que le processus $MA(q)$ admet une représentation en processus $AR(\infty)$

$$\epsilon_t = X_t + \beta_1 X_{t-1} + \dots, \quad \forall t$$

où les β_j sont les coefficients du développement en série de $B(z)^{-1}$ autour de zéro.

Pour un processus moyenne mobile $MA(q)$,

$$C(h) = 0, \quad \forall h > q$$

Ainsi l'auto-corrélation d'un processus MA a le même comportement que l'auto-corrélation partielle pour un AR . L'auto-corrélation partielle d'un processus MA est compliquée. On retiendra que l'auto corrélation partielle d'un processus MA décroît avec une vitesse exponentielle.

7.2 Inférence pour les processus MA

Comme pour les processus AR , on peut construire un algorithme itératif pour estimer les coefficients d'un processus MA basé sur la covariance empirique. Cependant les estimateurs du maximum de vraisemblance ont de meilleures propriétés statistiques. Ils sont utilisés par défaut dans R et SAS.

Exemple : Le modèle de moyenne mobile d'ordre 1

$$y_t = \mu + \epsilon_t + \lambda\epsilon_{t-1}, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

Le vecteur des paramètres à estimer est (μ, λ, σ^2) .

Approche conditionnelle

On suppose que $\epsilon_0 = 0$.

La distribution conditionnelle de $y_1 | \epsilon_0 = \mu + \epsilon_1 \sim \mathcal{N}(\mu, \sigma^2)$.

Plus généralement, si ϵ_{t-1} est connu, $y_t | \epsilon_{t-1} \sim \mathcal{N}(\mu + \lambda \epsilon_{t-1}, \sigma^2)$.

On réitère le calcul de $t = 1$ à T et par application successives de la formule de Bayes

$$\mathcal{L}(y_1, \dots, y_T | \epsilon_0, \mu, \lambda, \sigma^2) = \prod_{t=1}^T \varphi(y_t | \epsilon_{t-1}; \mu, \lambda, \sigma^2)$$

Avec φ la densité de la loi de Gauss centrée réduite. On obtient pour la log-vraisemblance, conditionnellement à $\epsilon_0 = 0$

$$\log(\mathcal{L}(y_1, \dots, y_T | \epsilon_0, \mu, \lambda, \sigma^2)) = -\frac{T}{2}(\log(2\pi) + \log(\sigma^2)) - \frac{1}{2\sigma^2} \sum_{t=1}^T \epsilon_t^2$$

L'expression peut paraître simple, elle est cependant non linéaire en le paramètre λ ; en effet

$$\epsilon_t = (y_t - \mu) + \lambda(y_{t-1} - \mu) + \lambda^2(y_{t-2} - \mu) + \dots + \lambda^t \epsilon_0.$$

Approche exacte

On utilise la structure multivariée gaussienne du processus. Un processus MA(1) gaussien possède pour matrice de variance-covariance

$$\sigma^2 \begin{bmatrix} 1 + \lambda^2 & -\lambda & 0 & \dots & 0 \\ -\lambda & 1 + \lambda^2 & -\lambda & \dots & 0 \\ \vdots & & & & \vdots \\ 0 & \dots & & & 1 + \lambda^2 \end{bmatrix} = \sigma^2 \Gamma$$

La log-vraisemblance s'écrit donc

$$-\frac{T}{2}(\log(2\pi) + \log(\sigma^2) + \log(\det(\Gamma))) - \frac{1}{2\sigma^2} (y - \mu)^T \Gamma^{-1} (y - \mu)$$

En fait c'est maintenant que le travail commence. On sait qu'une matrice symétrique se factorise $\Gamma = UDU^T$ avec U une matrice triangulaire inférieure et D une matrice diagonale de terme générique

$$d_{ii} = \sigma^2 \frac{1 + \lambda^2 + \dots + \lambda^{2i}}{1 + \lambda^2 + \dots + \lambda^{2(i-1)}}$$

En posant

$$\bar{y}_t = y_t - \mu - \lambda \frac{1 + \lambda^2 + \dots + \lambda^{2(t-2)}}{1 + \lambda^2 + \dots + \lambda^{2(t-1)}} y_{t-1}$$

On obtient

$$\log(\mathcal{L}(y_1, \dots, y_T | \epsilon_0, \mu, \lambda, \sigma^2)) = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \sum_{t=1}^T \log(d_{tt}) - \frac{1}{2} \sum_{t=1}^T \frac{\bar{y}_t^2}{d_{tt}^2}$$

La non linéarité des paramètres est évidente.

7.3 Prédiction des processus MA

Pour ces processus, il est plus facile de projeter non pas sur l'espace engendré par X_1, \dots, X_t mais sur $X_{-\infty}, \dots, X_t$. C'est là que la représentation $AR(\infty)$ des processus MA est utile. En effet cette représentation implique que les espaces $X_{-\infty}, \dots, X_t$ et $\epsilon_{-\infty}, \dots, \epsilon_t$ sont égaux.

Proposition 9. *Pour tout t , le prédicteur à horizon 1 est*

$$\hat{X}_{n+1} = P_{[X_{-\infty}, \dots, X_t]}(X_{t+1}) = P_{[\epsilon_{-\infty}, \dots, \epsilon_t]}X_{t+1} = b_1\epsilon_t + \dots + b_q\epsilon_{t-q}$$

L'erreur de prévision à horizon 1 est

$$X_{t+1} - \hat{X}_{t+1} = X_{t+1} - (b_1\epsilon_t + \dots + b_q\epsilon_{t-q}) = \epsilon_{t+1}$$

On voit facilement à partir de la représentation

$$X_{t+h} = \epsilon_{t+h} + b_1\epsilon_{t+h-1} + \dots + b_q\epsilon_{t+h-q}$$

que le prédicteur est nul dès que l'horizon h dépasse q . L'erreur de prédiction est alors

$$X_{t+h} - \hat{X}_{t+h} = X_{n+h}$$

et la variance est $C(0)$.

Chapter 8

Les processus ARIMA

8.1 Les processus ARMA

L'inconvénient des modèles AR (resp. MA) est que si la fonction d'auto-corrélation partielle (resp. de corrélation) décroît très lentement, on doit avoir recours à des modèles d'ordre élevé pour bien caractériser les données. Les modèles ARMA, en combinant un processus AR et un processus MA, permettent de représenter un grand nombre de processus aléatoires stationnaires en impliquant un nombre raisonnable de paramètres. Ce chapitre détaille les principales méthodes de prédiction et d'identification de ces processus. Il est supposé à ce niveau que l'on a isolé, dans la série chronologique d'origine, sa partie aléatoire, en lui retirant tendances et facteurs saisonniers.

Exemple : ARMA(1,1)

$$x_t - \phi_1 x_{t-1} = \phi_0 + \epsilon_t - \theta_1 \epsilon_{t-1}$$

Il est facile d'écrire les la moyenne, la variance, l'auto-corrélation... pour un processus ARMA(1,1). On utilise l'hypothèse que le processus est stationnaire et on cherche les points fixes, comme pour le processus AR(1).

$$E(x_t) = \frac{\phi_0}{1 - \phi_1}$$
$$Var(x_t) = \frac{(1 - 2\phi_1\theta_1 + \theta_1^2)\sigma^2}{1 - \phi_1^2}$$

On remarque que pour que la variance soit positive, il faut que $\phi - 1^2 < 1$ et on retrouve la condition de stationnarité d'un processus AR(1).

8.1.1 Définition et propriétés

Définition 9. *Le processus X_t , où $t \in \mathbb{Z}$, est un ARMA(p, q) de moyenne nulle si : (X_t) est stationnaire,*

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \phi_0 + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q},$$

où ϵ_t est un bruit blanc centré et de variance σ^2 .

Une représentation équivalente et plus concise est la suivante :

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

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

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

$$\Theta(z) = 1 - \theta_1 B + \dots - \theta_p B^p$$

et B est l'opérateur "retard" défini par

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

Les processus AR et MA sont des ARMA particuliers.

Proposition 10. *Fonction de corrélation*

Pour tout processus aléatoire stationnaire dont la fonction de corrélation $\rho(\cdot)$ tend vers 0 à l'infini, et pour tout entier $k > 0$, il est possible de trouver un processus ARMA dont la fonction de corrélation est égale à celle du processus aléatoire jusqu'à l'ordre k .

Cette propriété est l'une des raisons de l'intérêt pratique des modèles ARMA; une autre raison est leur facilité de synthèse.

Proposition 11. *Mémoire courte*

Pour un processus ARMA, la fonction de corrélation est bornée géométriquement :

$$|\rho(k)| \leq Cr^k, \quad k = 1, 2, \dots, \quad C > 0, \quad 0 < r < 1$$

on dit qu'un processus ARMA est à mémoire courte (décroissance "rapide" de ρ).

Les processus ARMA admettent 3 représentations différentes. En effet, on peut aussi les écrire comme des processus AR ou MA infinis.

Définition 10. *Un processus ARMA est dit **inversible** s'il peut s'écrire comme un processus $AR(\infty)$. Une condition suffisante d'inversibilité, est que toutes les racines du polynôme $\Theta(\cdot)$ soient de module strictement plus grand que 1.*

Dans ce cas, x_t s'exprime comme une combinaison linéaire de l'innovation et du passé.

Un processus ARMA peut aussi s'écrire comme une combinaison linéaire de sa moyenne et des innovations (ou chocs) passés

$$x_t = \mu + \epsilon_T + \psi_1 \epsilon_{t-1} + \dots$$

On dit que le processus est **causal**. Les coefficients ϕ_i sont décroissants et tendent vers 0 quand i tend vers l'infini.

Définition 11. *Un processus ARMA défini par $\Phi(B)X_t = \Theta(B)Z_t$ est dit **causal** s'il admet une unique solution stationnaire de la forme d'un $MA(\infty)$. Il suffit que le polynôme Φ admette une inverse ie que ses racines soient de module strictement supérieur à 1.*

Proposition 12. *Soit $\{X_t\}$ un processus ARMA(p, q) tel que les polynômes Φ et Θ n'ont pas de racines communes. Alors $\{X_t\}$ est causal si et seulement si $\Phi(z) \neq 0$ pour tout $z \in F$ tel que $|z| \leq 1$. Les coefficients ψ_j sont déterminés par*

$$\psi(z) = \frac{\Theta(z)}{\Phi(z)}, \quad \text{pour tout } z \text{ tel que } |z| \leq 1$$

Si le processus ARMA est causal, on peut exprimer sa fonction d'auto-covariance :

$$C(h) = \sigma^2 \sum_{k=1}^{\infty} \psi_k \psi_{k+h}$$

8.1.2 Inférence

Il y a présomption de processus ARMA si les conditions suivantes sont satisfaites :

- le processus est stationnaire à l'analyse visuelle : pas de tendance, pas de saisonnalité, variance constante.
- la fonction de corrélation empirique est : à décroissance pas trop lente et sans pics périodiques.

Il faut noter que les fonctions d'auto-corrélation et d'auto-corrélation partielle ne sont pas informatives pour déterminer l'ordre d'un processus ARMA.

Il existe deux méthodes pour l'estimation des paramètres ϕ , θ et σ^2 :

- une méthode des moments basée sur la fonction de covariance empirique, dite robuste, qui fait une estimation directe des paramètres.
- une méthode d'estimation paramétrique basée sur le maximum de vraisemblance gaussien, applicable de façon optimale au cas des processus gaussiens.

La deuxième méthode donne généralement des résultats plus précis que la première, au moins dans le cas des processus gaussiens, mais sa mise en oeuvre passe par la minimisation d'une fonction de plusieurs variables qui n'admet pas de forme explicite. Il n'existe pas de résultats généraux sur la forme de cette fonction : convexité, existence de minima locaux, vitesse de variation . . . Il est donc préférable de disposer de valeurs approchées des paramètres pour l'initialisation ; on utilise par exemple la méthode des moments.

Proposition 13. *La vraisemblance L d'un processus ARMA gaussien, de moyenne nulle, de paramètres (ϕ, θ, σ^2) , est donnée par la formule suivante :*

$$L(\phi, \theta, \sigma^2) = (2\pi\sigma^2)^{-T/2} (r_0 \cdots r_{T-1})^{-1/2} \exp \left(-\frac{1}{2\sigma^2} \sum_j \frac{(X_j - \hat{X}_j)^2}{r_{j-1}} \right)$$

où $\hat{X}_{j+1} = \hat{X}_{j+1}(\phi, \theta, \sigma^2)$ est la prédiction de X_{j+1} sachant X_1, \dots, X_j , dans le modèle ARMA, et les r_j sont les variances des erreurs de prédiction, divisées par la variance de ϵ :

$$r_j = r_j(\phi, \theta, \sigma^2) = \frac{1}{\sigma^2} E((X_j - \hat{X}_j)^2).$$

On observe que pour calculer la vraisemblance on a besoin de la prévision à un pas de temps. Cette prévision est obtenue par un algorithme récursif.

Pour les processus ARMA, comme pour les autres modèles (AR, MA, ...), les statistiques comme la moyenne, la variance, etc calculées à partir des paramètres estimés sont considérés comme les statistiques de la **loi stationnaire** (ou asymptotique).

8.1.3 Prédiction

Le meilleur estimateur de X_{t+1} au sens de l'erreur en moyenne quadratique est l'espérance conditionnelle de X_{t+1} sachant X_t, \dots, X_1 c'est à dire que c'est \hat{X}_{t+1} , la projection de X_{n+1} sur le sous espace vectoriel engendré par X_t, \dots, X_1 . Autrement dit,

$$\hat{X}_{n+1}|n = \phi_0 + \sum_{i=1}^p \phi_i x_{n+1-i} - \sum_{i=1}^q \theta_i \epsilon_{n+1-i}$$

L'erreur associée est $e_n(1) = x_n - \hat{x}_{n+1|n} = \epsilon_{n+1}$ on a donc la variance de l'erreur.

Pour les prédictions suivantes, on utilise un algorithme itératif de Durbin-Watson. Et leur variance est calculée en utilisant la représentation MA du processus.

8.1.4 Validation

On valide le modèle comme pour les processus AR : analyse des résidus (test bruit blanc), validation croisée, etc.

8.2 Les processus ARIMA et SARIMA

8.2.1 Processus non stationnaires

La marche aléatoire est un processus ARIMA(0,1,0) : $p_t = p_{t-1} + \epsilon_t$.

Si p_t est le prix du log d'un stock donné à la date t , alors p_0 pourrait être le prix du stock au moment de son premier appel public à l'épargne (c'est-à-dire le prix de l'introduction en bourse). Si ϵ a une distribution symétrique autour de zéro, alors conditionnellement à $p_t - 1$, p_t a 50% de chance d'aller vers le haut ou vers le bas, ce qui implique que p_t ira vers le haut ou vers le bas au hasard. Si nous considérons le modèle aléatoire comme un modèle AR(1) spécial, alors le coefficient de $p_t - 1$ est l'unité, qui ne satisfait pas à la condition de stationnarité faible d'un modèle AR(1). Une série de marche aléatoire n'est donc pas faiblement stationnaire, et nous l'appelons une série temporelle non stationnaire racine-unité.

8.2.2 Processus ARIMA

Les modèles ARIMA sont bien adaptés au cas précédent, et plus généralement au cas des séries temporelles présentant une décroissance très lente de la fonction d'auto-covariance (présence d'une tendance). En pratique, on modélise le processus des accroissements $X_t - X_{t-d}$ par un processus ARMA : ARIMA(p,d,q). Le plus souvent $d = 1$, ce qui permet d'éliminer une tendance linéaire. Le I de ARIMA signifie "intégré". En utilisant l'opérateur de retard, on écrit

$$\Phi(B)(1 - B)^d X_t = \Theta(B)\epsilon_t.$$

Plus généralement, ces processus permettent de modéliser les séries présentant des racines unités. En pratique, on fixe l'ordre du I au nombre de racines unités.

En finance, on considère généralement que les séries de prix ne sont pas stationnaires, mais la série des rendements,

$$r_t = \ln(p_t) - \ln(p_{t-1}),$$

est stationnaire. Dans ce cas, la série de prix de log est unitaire non stationnaire et peut donc être traitée comme un processus ARIMA.

8.2.3 Test racine unité

Pour tester si le prix p_t d'un produit financier suit un modèle de marche aléatoire ou un modèle de marche aléatoire avec tendance, on considère les modèles

$$\begin{aligned} p_t &= \phi_1 p_t + \epsilon_t \\ p_t &= \phi - 0 + \phi_1 p_t + \epsilon_t \end{aligned}$$

et les hypothèses de test :

$$H_0 : \phi_1 = 1, \quad H_1 : \phi_1 < 1$$

C'est le test de Dickey-Fuller. La statistique de test est

$$S = \frac{\hat{\phi}_1 - 1}{\sqrt{\text{Var}(\hat{\phi}_1)}} = \frac{\sum_{t=1}^T p_{t-1} \epsilon_t}{\hat{\sigma} \sqrt{\sum_{t=1}^T p_{t-1}^2}}$$

où $\hat{\phi}_1$ et $\hat{\sigma}$ sont les estimateurs des moindres carrés. Sous l'hypothèse nulle, cette statistique de test converge vers un processus connu (mouvement brownien si $\phi_0 = 0$) et les valeurs critiques du test sont obtenues par simulation.

Dans le cas d'un AR(p), on écrit

$$x_t = c_t + \beta x_{t-1} + \sum_{i=1}^{p-1} \Phi_i (x_{t-i} - x_{t-i-1}) + \epsilon_t$$

avec c une fonction déterministe du temps t et on teste $\beta = 1$. La statistique de test est

$$S = \frac{\hat{\beta} - 1}{\sqrt{\text{var}(\hat{\beta})}}$$

C'est le test augmenté de Dickey-Fuller (ADF test).

8.2.4 Modèle SARIMA

Le modèle SARIMA permet de prendre en plus en compte la saisonnalité. Elle est éliminée par différenciation :

$$\Phi(B)(1 - B)^d(1 - L^s)^D X_t = \Theta(B)\epsilon_t,$$

avec s la longueur de la saison (par exemple, $s=12$ si on a des données mensuelles avec une période d'un an).

Sous R : `arima(x, orderc(p,d,q))`. Voir les exemples dans l'aide en ligne : MA(1)+tendance+ saison.

Sous SAS : PROC ARIMA.

8.3 Modèles avec co-variables : ARMAX

$$\Phi(B)X_t = \Theta(B)\epsilon_t + D(B)Z_t$$

soit encore

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i} + \sum_{i=1}^b z_{t-d}$$

Chapter 9

Modèles hétéroscédastiques

9.1 Introduction

L'objectif de ce chapitre est d'étudier quelques méthodes statistiques et modèles économétriques disponibles dans la littérature pour modéliser la volatilité du rendement d'un actif. Les modèles sont appelés modèles hétéroscédastiques conditionnels.

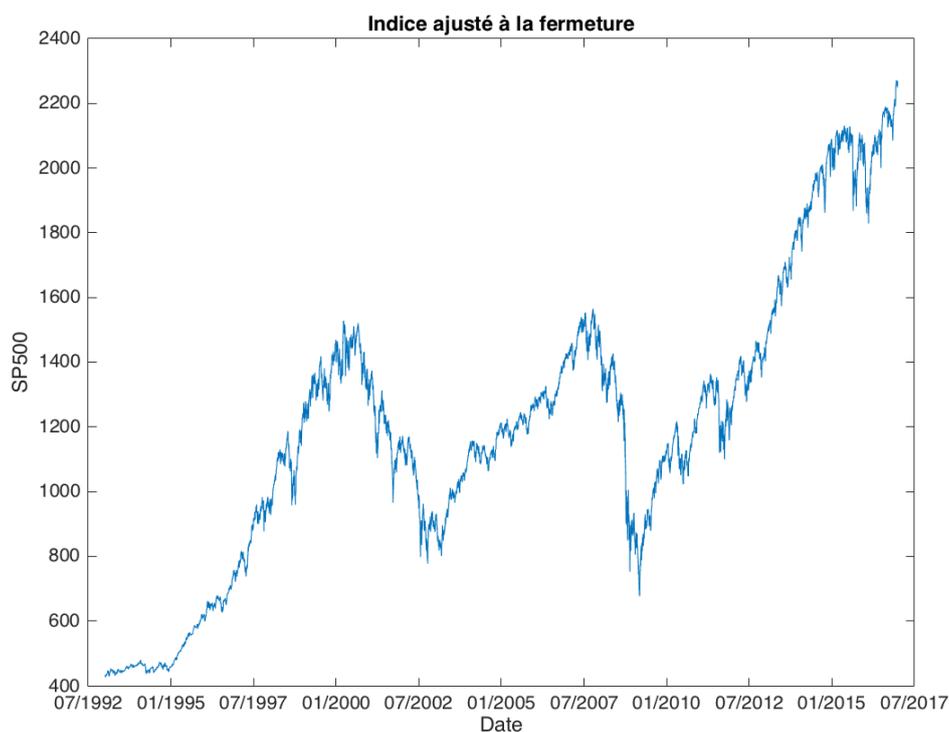


Figure 9.1: Exemple : indice SP500, valeur ajustée à la fermeture entre 1995 et 2016

9.2 Caractéristiques de la volatilité

Une caractéristique particulière de la volatilité des actions est qu'elle n'est pas directement observable. Par exemple, considérez les rendements journaliers de l'indice SP500 (Fig. ref-

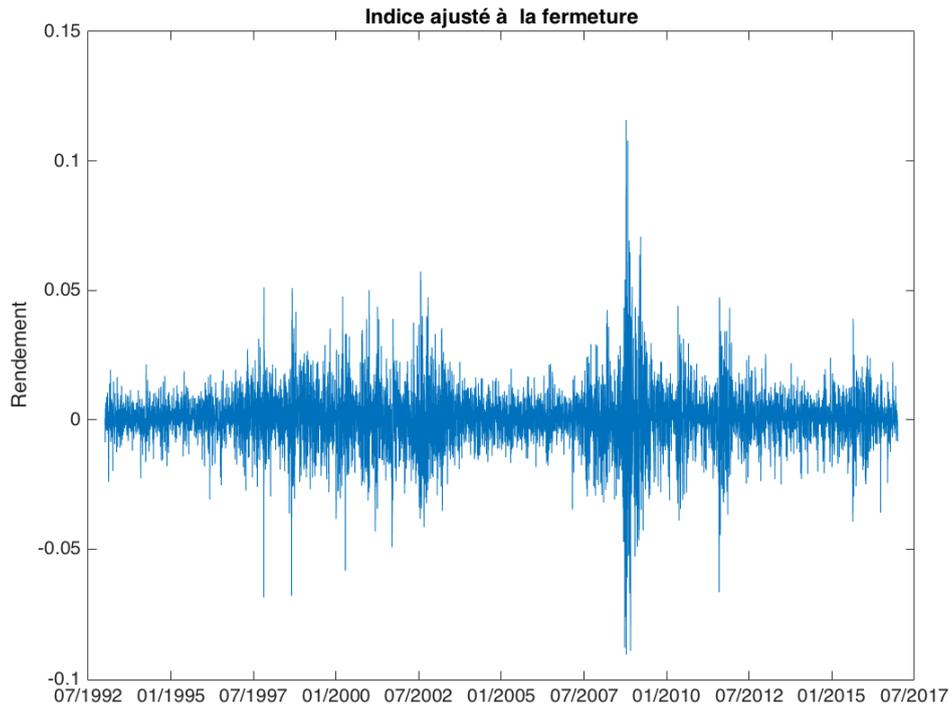


Figure 9.2: Rendement de l'indice SP500, valeur ajustée à la fermeture entre 1995 et 2016

fig:RSP). La volatilité quotidienne n'est pas directement observable à partir des données de rendement car il n'y a qu'une seule observation dans une journée de négociation. Si des données intra-journalières du stock, telles que des retours de 10 minutes, sont disponibles, alors on peut estimer la volatilité quotidienne.

Par exemple pour l'indice journalier SP500, on peut estimer la volatilité mensuelle en calculant la variance empirique dans une fenêtre glissante de largeur 1 mois.

Bien que la volatilité ne soit pas directement observable, elle présente certaines caractéristiques qui sont couramment observées dans le rendement des actifs.

1. Il existe des grappes de volatilité (c'est-à-dire la volatilité peut être élevée pour certaines périodes de temps et faible pour d'autres périodes).
2. La volatilité évolue dans le temps de façon continue, c'est-à-dire que les sauts de volatilité sont rares.
3. Troisièmement, la volatilité ne diverge pas à l'infini, c'est-à-dire que la volatilité ne varie dans une certaine fourchette fixe.
4. Quatrièmement, la volatilité semble réagir différemment à une forte hausse des prix ou à une forte baisse des prix, appelée effet de levier.

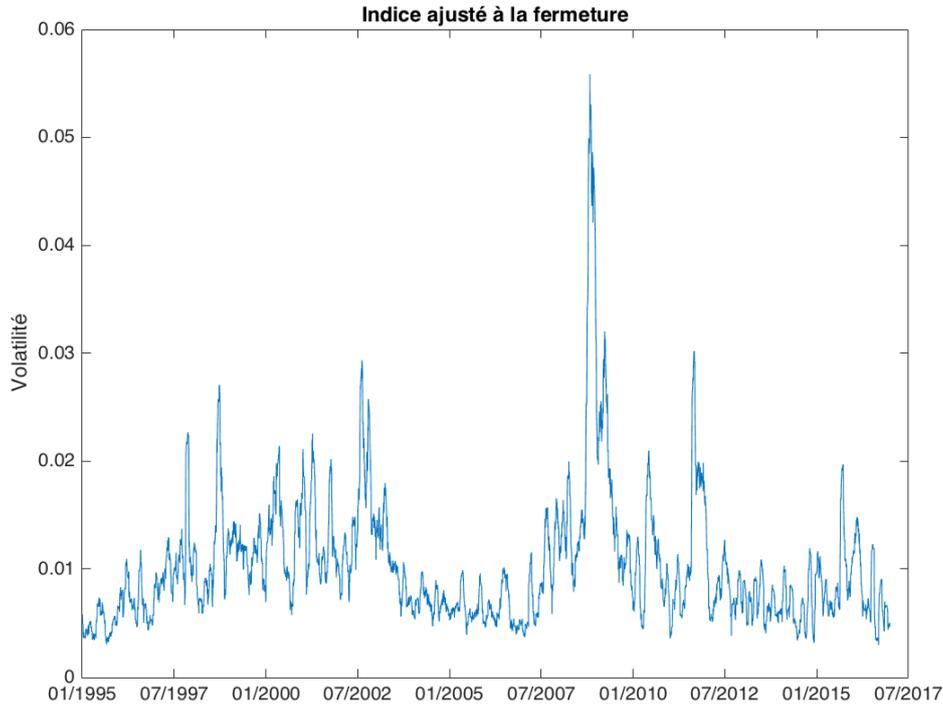


Figure 9.3: Volatilité du rendement de l'indice SP500, valeur ajustée à la fermeture entre 1995 et 2016

9.3 Structure du modèle ARCH

9.3.1 Généralités

Soit r_t le (logarithmique) du rendement d'un actif à l'indice temporel t . L'idée de base de l'étude de la volatilité est que la série $(r_t)_{t \in \mathbb{Z}}$ est non corrélée avec des corrélations sérielles mineures (Fig. 9.4), mais c'est néanmoins une série dépendante. On le constate car les r_t^2 (ou $\log(r_t)^2$) sont corrélés (Fig. 9.5). Les modèles ARCH permettent de décrire cette dépendance.

Dans les années 80, des développements ont été apportés dans l'étude de la non-linéarité de certaines séries, et sur leur modélisation. En 1982, Engle a introduit la classe des modèles ARCH (auto-régressifs conditionnellement hétéroscédastiques). Ces modèles ont été introduits pour palier à une observation empirique qui n'était pas prise en compte par les modèles : la volatilité conditionnelle d'une série (Y_t) n'a aucune raison d'être constante. Dans les modèles AR(1), la variance conditionnelle de X_t sachant X_{t-1} est constante : $Var(X_t|X_{t-1}) = \sigma^2$ où $Var(\epsilon_t) = \sigma^2$ (notion d'homoscédasticité). Engle a cherché un modèle dans lequel la variance conditionnelle de X_t sachant X_{t-1} dépendrait de X_{t-1} , et plus particulièrement, $Var(X_t|X_{t-1}) = (\alpha + \beta X_{t-1}^2)\sigma^2$. Pour cela, il a considéré les modèles la forme

$$X_t = \epsilon_t \sigma_t, \text{ où } \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2.$$

Ce modèle ARCH(1) a ensuite été généralisé sous la forme ARCH(p)

$$X_t = \epsilon_t \sigma_t, \text{ où } \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_p X_{t-p}^2$$

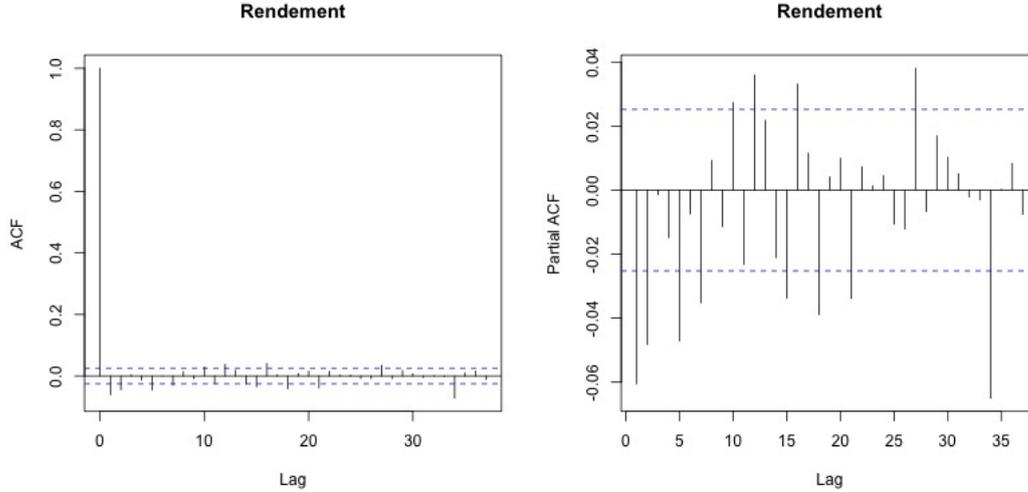


Figure 9.4: Rendement de l'indice SP500, valeur ajustée à la fermeture entre 1995 et 2016

puis sous la forme des modèles GARCH

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \epsilon_{t-j}$$

On dit que ϵ_t est le choc, ou l'innovation, du rendement d'un produit financier.

9.3.2 Modèle ARCH(1)

Pour bien comprendre les modèles GARCH, il est utile d'étudier le modèle ARCH(1)

$$X_t = \epsilon_t \sigma_t, \text{ où } \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2.$$

Il est facile de vérifier que X_t a une moyenne nulle

$$E(X_t) = E(E(X_t | X_{t-1})) = E(\sigma_t E(\epsilon_1)) = 0.$$

La variance est donnée par

$$Var(X_t) = E(X_t^2) = E(E(X_t^2 | X_{t-1})) = \alpha_0 + \alpha_1 E(X_{t-1}^2).$$

Sous les hypothèse de stationnarité, on obtient donc

$$Var(X_t) = \frac{\alpha_0}{1 - \alpha_1}.$$

Cette équation implique que $0 \leq \alpha_1 < 1$.

On peut aussi écrire le moment d'ordre 4 du processus ARCH(1) sous l'hypothèse que ϵ suit une loi de Gauss.

$$E(X_t^4 | X_{t-1} = x_{t-1}) = 3E^2(X_t^2 | X_{t-1} = x_{t-1}) = 3(\alpha_0 + \alpha_1 x_{t-1}^2)^2$$

Ainsi, on peut aussi écrire

$$m_4 = E(X_t^4) = E(X_t^4 | X_{t-1}) = 3(\alpha_0^2 + 2\alpha_0\alpha_1 Var(X_{t-1}) + \alpha_1 m_4)$$

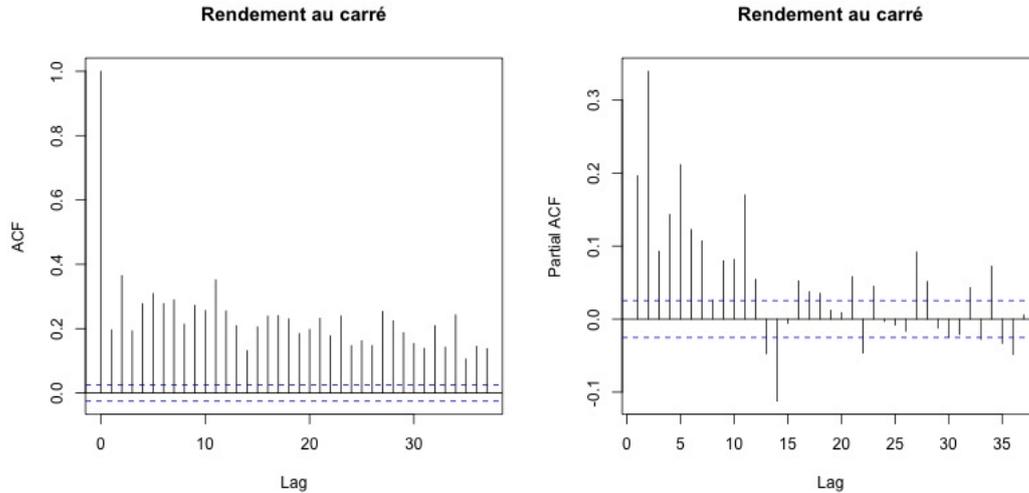


Figure 9.5: Rendement au carré de l'indice SP500, valeur ajustée à la fermeture entre 1995 et 2016

et

$$m_4 = \frac{3\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)}$$

Le moment d'ordre 4 est positif, on en déduit donc que $0 \leq \alpha_1 < 1/3$. D'autre part, le moment d'ordre 4 centré réduit, le kurtosis, est

$$\frac{m_4}{(\text{Var}(X_t))^2} = 3 \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} > 3.$$

On en déduit que le processus ARCH(1) a des queues plus lourdes que la loi de Gauss, ie que ce processus a une probabilité plus forte d'avoir des valeurs élevées.

9.3.3 Inconvénients

Les modèles ARCH présentent quelques inconvénients.

1. Le modèle ARCH suppose que les choc positifs et les chocs négatifs ont le même effet sur la volatilité car la variance dépend du carré des chocs passés. Cette hypothèse est rarement vérifiée dans les séries financières.

2. Les conditions nécessaires sur les coefficients α pour avoir une porcessus stationnaire sont restrictives.

3. Le modèle ARCH ne fournit aucune indication pour comprendre la source des variations d'une série chronologique financière. Elle fournit simplement une façon de décrire le comportement de la variance conditionnelle.

4. Les modèles ARCH sont susceptibles de surestimer la volatilité parce qu'ils réagissent lentement aux grands chocs isolés de la série.

9.4 Inférence

Si un effet ARCH est jugé significatif, on peut utiliser le PACF de x_t^2 pour déterminer l'ordre ARCH.

L'estimation des paramètres est obtenue par maximum de vraisemblance.

$$\log \mathcal{L}(x_{p+1}, \dots, x_T | \theta, x_1, \dots, x_p) = \log 2\pi - \sum_{t=p+1}^T \left(\frac{1}{2} \log \sigma_t^2 + \frac{1}{2} \frac{x_t^2}{\sigma_t^2} \right)$$

avec $\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2 + \dots + \alpha_p x_{t-p}^2$.

Pour un modèle ARCH bien spécifié, les résidus standardisés

$$\tilde{x}_t = \frac{x_t}{\sigma_t}$$

forment une suite de variables aléatoires iid.

Les prédictions sont obtenues récursivement comme dans le cas d'un AR.

9.5 Modèle GARCH

Bien que le modèle ARCH soit simple, il nécessite souvent de nombreux paramètres pour décrire adéquatement le processus de volatilité d'un rendement d'actif. Pas exemple pour modéliser le rendement de l'indice SP500, on ajuste un modèle ARCH(9).

Pour réduire le nombre de paramètres, on a introduit les modèles GARCH

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \epsilon_{t-j}.$$

Comme pour les modèles ARCH, on peut montrer que les modèles GARCH décrivent des processus à queue plus lourde que les processus gaussiens.

Pour prédire les processus GARCH, on utilise les mêmes techniques que pour les processus ARMA.

Pour certaines séries temporelles, les queues des processus GARCH ne sont pas suffisamment lourdes, on peut alors considérer des innovations de loi de Student.