

EISTI



Ecole  
Internationale  
des Sciences  
du Traitement  
de l'Information

SERIES TEMPORELLES

-

PROCESSUS ARMA



# SOMMAIRE

Sommaire .....	3
Références.....	4
<b>Chapitre 1 : Introduction, exemples et généralités .....</b>	<b>5</b>
1. Exemples.....	5
2. Objectifs de l'analyse d'une série temporelle.....	6
2.1. Premier objectif : décrire.....	6
2.2. Deuxième objectif : modéliser .....	9
2.3. Troisième objectif : prévoir .....	12
3. Processus stationnaires .....	12
<b>Chapitre 2 : Lissage par moyenne mobile .....</b>	<b>14</b>
1. Désaisonnalisation .....	14
1.1. Moyenne mobile centrée.....	14
1.2. Moyenne mobile différence .....	15
2. Prévision d'une série sans composante saisonnière et sans tendance .....	16
3. Prévision d'une série sans composante saisonnière et avec tendance.....	17
<b>Chapitre 3 : Prévision par lissage exponentiel.....</b>	<b>19</b>
1. Prévision d'une série sans composante saisonnière et sans tendance .....	19
2. Prévision d'une série sans composante saisonnière et avec tendance.....	20
3. Ajout d'une composante saisonnière .....	21
4. Lissage exponentiel de Holt-Winter .....	23
4.1. Série sans composante saisonnière .....	23
4.2. Série avec composante saisonnière .....	24
<b>Chapitre 4 : Prévision par régression linéaire .....</b>	<b>26</b>
1. Le modèle linéaire .....	26
2. Identification du modèle .....	27
2.1. Estimation par moindres carrés ordinaires.....	27
2.2. Validation du modèle .....	29
2.3. Prévision à l'horizon h .....	32
<b>Chapitre 5 : Erreurs de prévision et auto corrélation.....</b>	<b>34</b>
1. Test sur la moyenne des résidus .....	34
2. Tests d'auto corrélation des résidus .....	35
2.1. Tests individuels de bruit blanc .....	35
2.2. Test global de bruit blanc .....	35
3. Quelques commandes SAS .....	36
<b>Chapitre 6 : Modèle ARMA.....</b>	<b>38</b>
1. Processus autorégressif (AR) .....	38
1.1. Processus autorégressif d'ordre 1 – AR(1) .....	38
1.2. Processus autorégressif d'ordre p – AR(p) .....	40
1.3. Quelques commandes SAS.....	43
2. Processus moyenne mobile (MA).....	44
2.1. Processus moyenne mobile d'ordre 1 – MA(1).....	44
2.2. Processus moyenne mobile d'ordre q – MA(q).....	45
2.3. Quelques commandes SAS.....	47
3. Processus ARMA .....	48
4. Identification d'un processus ARMA .....	51
4.1. Choix des paramètres p et q d'un modèle ARMA .....	51
4.2. Estimation des paramètres d'un modèle ARMA .....	53
4.3. Choix du modèle ARMA .....	54
4.4. Exemple.....	55
5. Prévision dans un processus ARMA.....	59
5.1. Prévision avec un processus AR.....	59

5.2.	Prévision dans les processus MA.....	60
5.3.	Prévision dans les processus ARMA.....	61
5.4.	Intervalle de confiance.....	62

## REFERENCES

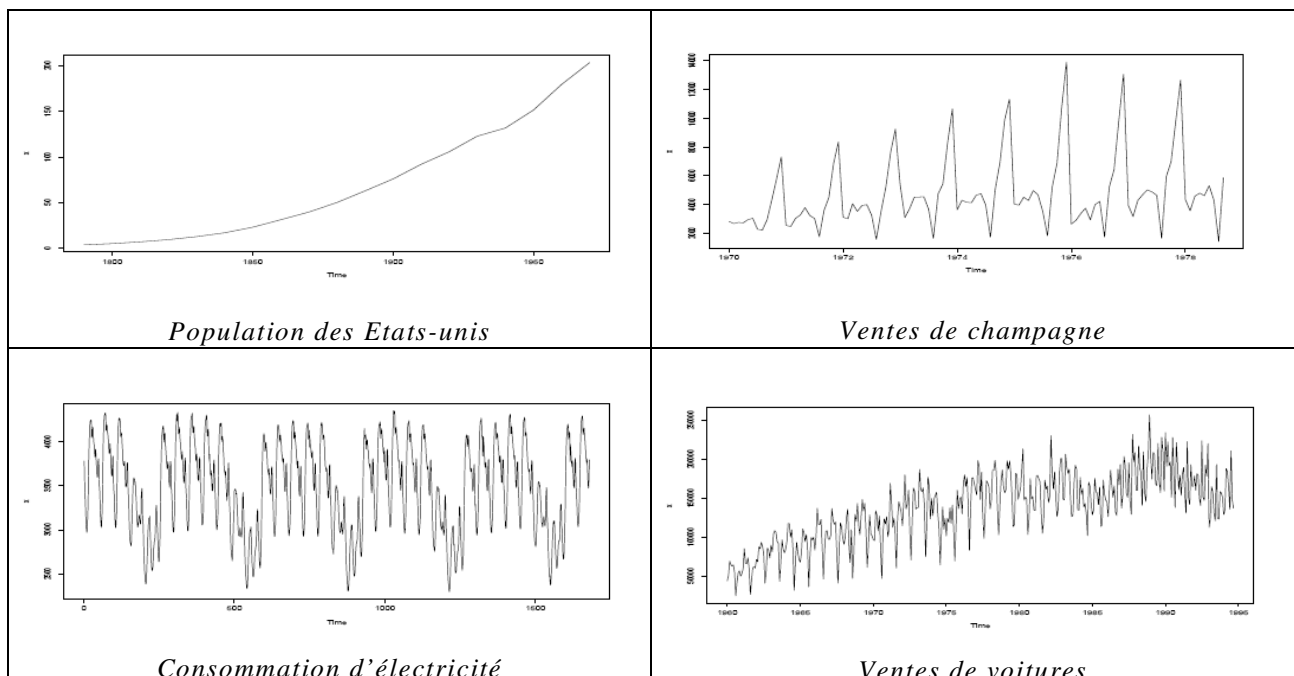
- Aragon Y. – Introduction aux séries temporelles – Master statistique et économétrie
- Charpentier A. – Cours séries temporelles – DESS Mathématiques de la décision & DESS Actuariat
- Mélard G. Méthode de prévision à court terme. Edition Ellipses
- Viano M.C. – Cours de séries temporelles – Maîtrise d'économétrie
- Mélard G. – Initiation à l'analyse des séries temporelles et à la prévision – Revue Modulad 2006, n°35 (revue électronique gratuite)

# CHAPITRE 1 : INTRODUCTION, EXEMPLES ET GENERALITES

Une série temporelle (ou série chronologique) est une suite d'observations indexées par le temps  $(x_1, \dots, x_T)$ . L'indice temps peut représenter suivant les études, l'année, le jour, l'heure, la minute, etc... L'étude des séries temporelles a pour but de construire un modèle statistique permettant de prévoir les observations futures à partir de la connaissance des observations présentes et passées. Bien entendu, cet objectif sous entend une dépendance dans le temps des observations. Il ne sert à rien de prévoir le résultat du 11<sup>ème</sup> lancer d'un dé connaissant 10 lancers précédents. Cette notion de dépendance est primordiale pour modéliser une série chronologique.

## 1. EXEMPLES

Les séries temporelles s'appliquent à des domaines très variés. De nombreux exemples sont présents dans la littérature et parmi les plus connus, nous avons



- La population américaine. Il s'agit un relevé tous les 10 ans de 1790 à 1980. On observe une croissance de la population avec le temps sans grande variabilité.
- Les ventes de champagne. Il s'agit d'un relevé mensuel de bouteilles vendues entre mars 1970 et septembre 1978. Le graphique montre deux saisonnalités

avec un pic correspondant aux fêtes de fin d'année. On note de plus que la moyenne et la variabilité augmentent au cours du temps

- La consommation d'électricité. Il s'agit d'un relevé sur toute la France à raison d'une donnée toutes les demi-heures pendant 35 jours en juin-juillet 1991. On remarque deux périodes, une correspondant à la journée (jour-nuit) et l'autre à la semaine (semaine-week-end).
- Les ventes de voitures. Il s'agit d'un relevé mensuel entre 1960 et 1995. On remarque une tendance à la hausse ainsi que deux saisonnalités. La courbe présente aussi une rupture aux environs de 1983, ce qui rend cette série extrêmement difficile à modéliser.

Toutes ces séries sont très différentes mais toutes présentent une dépendance « temporelle » des observations. Pour chacune d'elles, on comprend bien l'intérêt stratégique à vouloir prévoir les valeurs de la série pour des temps futurs.

## 2. OBJECTIFS DE L'ANALYSE D'UNE SERIE TEMPORELLE

### 2.1. Premier objectif : décrire

Comme pour toutes séries de données, la première étape consiste à décrire et résumer la série. On utilise pour ce faire les représentations graphiques et indices numériques usuels. Afin d'illustrer notre propos, nous allons considérer des données de la SNCF comptabilisant le trafic voyageurs entre 1963 et 1980 (Gouriéroux et Monfort)

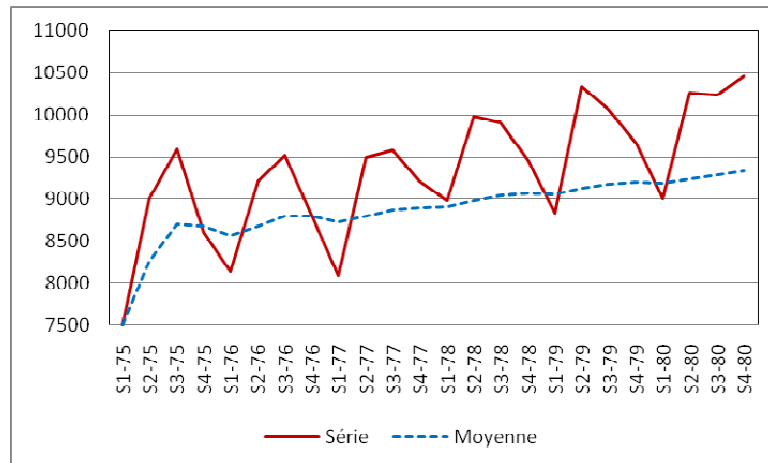
	JAN	FEB	MAR	AVR	MAI	JUN	JUL	AOU	SEP	OCT	NOV	DEC	Moyenne
1963	1750	1560	1820	2090	1910	2410	3140	2850	2090	1850	1680	2420	2131
1964	1710	1600	1800	2120	2100	2460	3200	2960	2190	1870	1770	2270	2171
1965	1670	1640	1770	2190	2020	2610	3190	2860	2140	1870	1760	2360	2173
1966	1810	1640	1860	1990	2110	2500	3030	2900	2160	1940	1750	2330	2168
1967	1850	1590	1880	2210	2110	2480	2880	2670	2100	1920	1670	2520	2157
1968	1834	1792	1860	2138	2115	2485	2581	2689	2038	1936	1734	2391	2133
1969	1798	1850	1981	2085	2120	2491	2834	2725	1932	2085	1856	2553	2193
1970	1854	1823	2005	2418	2219	2722	2912	2771	2153	2136	1910	2537	2288
1971	2008	1835	2120	2304	2264	2175	2928	2738	2178	2137	2009	2546	2270
1972	2084	2034	2152	2522	2318	2684	2971	2759	2267	2152	1978	2723	2387
1973	2081	2112	2279	2661	2281	2929	3089	2803	2296	2210	2135	2862	2478
1974	2223	2248	2421	2710	2505	3021	3327	3044	2607	2525	2160	2876	2639
1975	2481	2428	2596	2923	2795	3287	3598	3118	2875	2754	2588	3266	2892
1976	2667	2668	2804	2806	2976	3430	3705	3053	2764	2802	2707	3307	2974
1977	2706	2586	2796	2978	3053	3468	3649	3095	2839	2966	2868	3375	3032
1978	2820	2857	3306	3333	3141	3512	3744	3179	2984	2950	2896	3611	3194
1979	3313	2644	2872	3267	3391	3682	3937	3284	2849	3085	3043	3541	3242
1980	2848	2913	3248	3250	3375	3640	3771	3259	3206	3269	3181	4008	3331
Moyenne	2195	2101	2309	2555	2489	2888	3249	2931	2426	2359	2205	2861	2547

### La moyenne

A chaque temps  $t$ , il est possible de calculer la moyenne des observations précédentes,

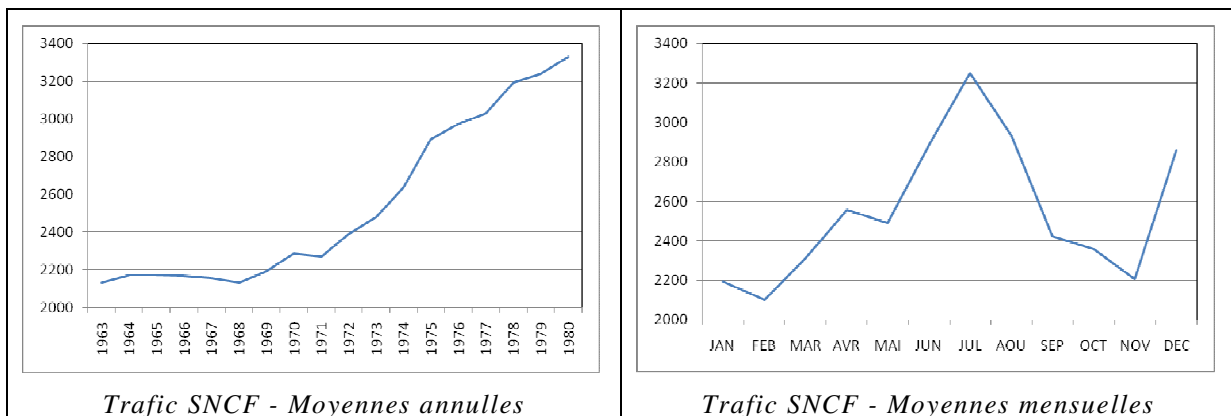
$$\bar{x}_t = \frac{1}{t} \sum_{i=1}^t x_i .$$

Ce premier indicateur donne une idée de la tendance centrale. On constate par exemple que le trafic voyageur suit une croissance moyenne linéaire.



*Trafic SNCF - Moyenne*

Les données étant mensuelles, il est possible de les moyenner par année pour faire apparaître la tendance ou par mois pour faire apparaître une périodicité.

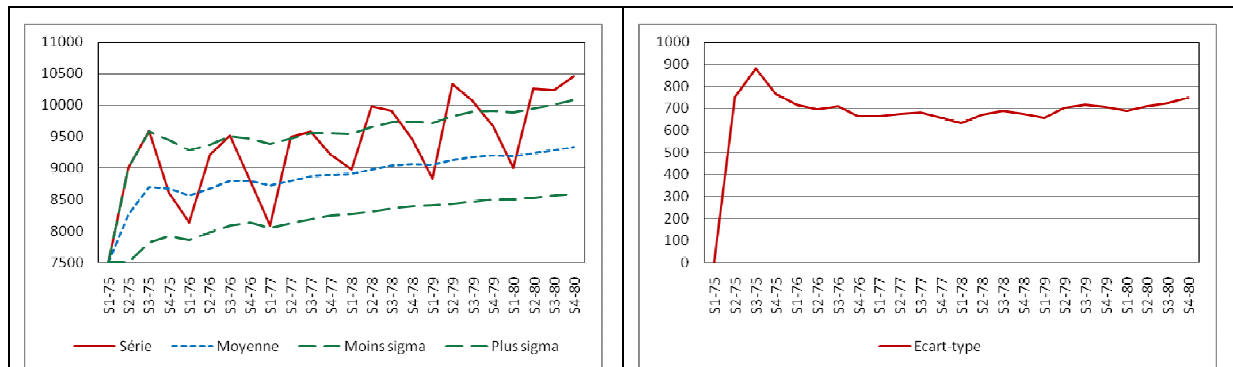


### La dispersion

De la même façon, à chaque temps  $t$ , il est possible de calculer la variance empirique sur les observations précédentes

$$\hat{\sigma}_t^2 = \frac{1}{t} \sum_{i=1}^t (x_i - \bar{x}_t)^2 .$$

On pourra ainsi juger de la dispersion des observations autour de la tendance centrale mais aussi de l'évolution de la variance. Nous verrons dans les chapitres suivants qu'il est nécessaire que la variance soit constante (condition d'homoscédasticité)



*Trafic SNCF – Moyenne et dispersion*

### La dépendance

Une question primordiale dans la modélisation des séries chronologiques est la dépendance temporelle des données. On utilise pour cela l'auto-covariance empirique.

$$\hat{\sigma}_t(1) = \frac{1}{t} \sum_{i=1}^{t-1} (x_i - \bar{x}_t)(x_{i+1} - \bar{x}_t)$$

indique la dépendance entre deux observations successives,

$$\hat{\sigma}_t(2) = \frac{1}{t} \sum_{i=1}^{t-2} (x_i - \bar{x}_t)(x_{i+2} - \bar{x}_t)$$

indique la dépendance entre deux observations ayant deux unités de temps d'écart,

$$\hat{\sigma}_t(h) = \frac{1}{t} \sum_{i=1}^{t-h} (x_i - \bar{x}_t)(x_{i+h} - \bar{x}_t)$$

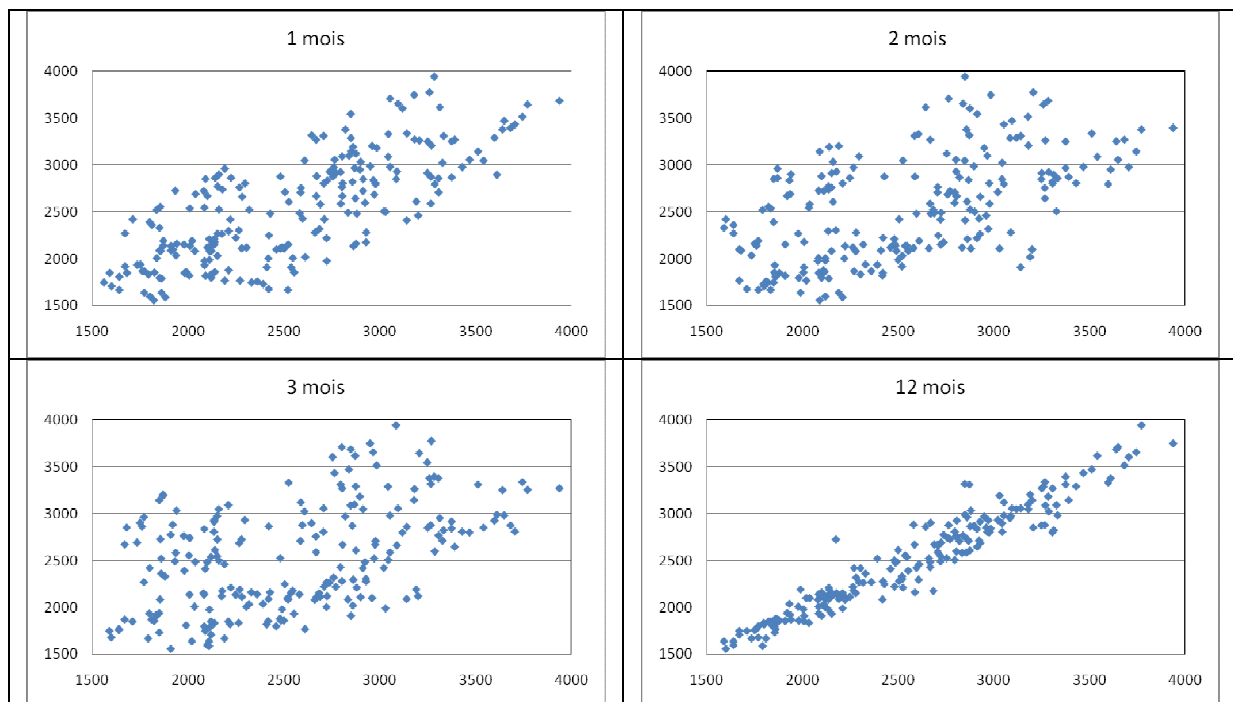
etc...

Il est cependant difficile de juger de l'importance d'une dépendance à l'aide de l'auto-covariance (pas de valeur limite). C'est pourquoi, on utilise l'auto-corrélation qui n'est rien d'autre que l'auto-covariance moyennée par la variance,

$$\hat{\rho}_t(h) = \frac{\hat{\sigma}_t(h)}{\hat{\sigma}_t^2}$$

L'auto-corrélation prend des valeurs comprises entre -1 et 1 et plus elle est proche de  $\pm 1$ , plus la dépendance est forte.

Il est possible d'avoir une première idée de l'auto-corrélation en visualisant les nuages de points des séries  $\{x_t\}$  et  $\{x_{t+h}\}$ . Plus le nuage de points est allongé, plus l'auto-corrélation est proche de  $\pm 1$ . Sur l'exemple du trafic SNCF, on note la dépendance entre  $x_t$  et  $x_{t+1}$  mais surtout celle  $x_t$  et  $x_{t+12}$ . Ce qui semble indiquer une période dans les données.



Trafic SNCF - Dépendance entre  $\{x_t\}$  et  $\{x_{t+1}\}, \{x_{t+2}\}, \{x_{t+12}\}$ .

## 2.2 Deuxième objectif : modéliser

Le deuxième objectif est de modéliser la série de données afin d'expliquer ses variations à l'aide de peu de paramètres. On cherche donc un modèle,  $f$ , tel que

$$x_t = f(x_{t-1}, x_{t-2}, \dots) + \varepsilon_t$$

où  $\varepsilon_t$  représente une erreur.

### 2.2.1. Tendence et saisonnalité

Il est classique d'introduire dans le modèle une tendance  $m_t$  et un effet saisonnier  $s_t$ , soit dans un modèle additif,

$$x_t = m_t + s_t + \varepsilon_t,$$

soit dans un modèle multiplicatif,

$$x_t = m_t \times s_t \times \varepsilon_t.$$

où l'erreur  $\varepsilon_t$  ne contient plus ni de tendance, ni d'effet saisonnier.

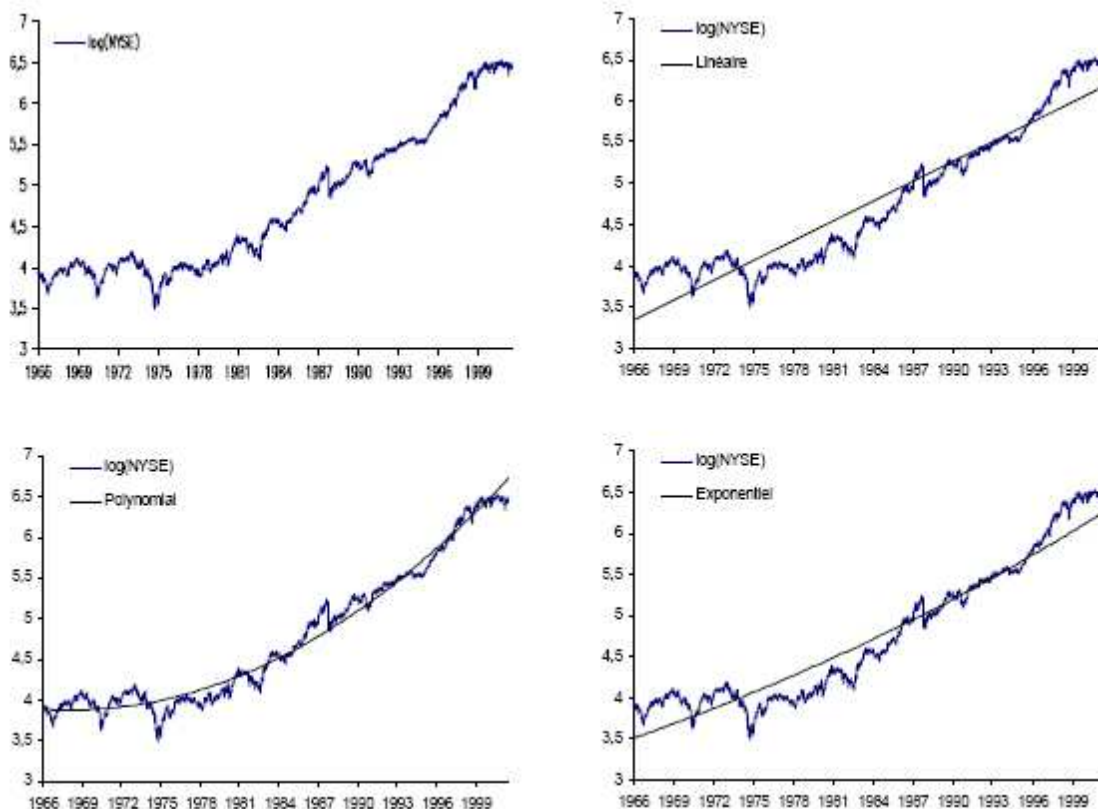
La tendance n'est pas nécessairement constante mais pourra être une fonction du temps  $t$  (en général, une fonction polynomiale). Cette composante reflète la croissance moyenne de la série. Elle peut prendre plusieurs formes dont les plus courantes sont

- linéaire :  $z_t = \beta_1 + \beta_2 t$ .
- exponentielle :  $z_t = \alpha r^t$ , ou  $z_t = \alpha(1+r)^t$ , ou encore  $z_t = \alpha \exp(rt)$

On se ramène au cas linéaire par transformation logarithmique. D'autres modèles existent tels que la forme quadratique,  $z_t = \beta_1 + \beta_2 t + \beta_3 t^2$ , le modèle de Gompertz,  $z_t = \exp(\alpha t + \gamma)$ , ...

### Exemple (Charpentier A.)

La série représente le logarithme de l'indice du New York Stock Exchange sur lequel trois tendances ont été ajustées : linéaire (en haut à droite), quadratique (en bas à gauche) et exponentiel (en bas à droite). La tendance quadratique semble la plus adaptée à la série de données mais il est difficile de lui trouver un sens physique contrairement à la tendance linéaire qui représente une croissance linéaire et la tendance exponentielle qui indique une augmentation constante (rapport entre  $X_t/X_{t-1}$  constant)



Si l'étude porte sur le court terme, il est nécessaire d'introduire l'effet saisonnier. Par exemple, en gestion de stock, il faut tenir en compte la saison pour passer commande aux fournisseurs. En revanche si l'étude concerne le moyen terme, l'effet saisonnier n'est pas important au regard des grandes tendances du phénomène. Il existe alors des procédures pour désaisonnaliser les séries.

### 2.2.2. Les erreurs

L'objectif est que le modèle explique entièrement la série de données. Cela implique que les erreurs  $\varepsilon_t$  ne doivent pas être corrélées car dans le cas contraire,  $\varepsilon_{t-1}$  contiendrait de l'information permettant d'expliquer  $\varepsilon_t$  et ainsi  $x_t$  ne serait pas

entièrement expliquée par  $f(x_{t-1}, x_{t-2}, \dots)$ . Nous reviendrons sur l'auto-corrélation des résidus dans le chapitre 5.

Une méthode pour valider la qualité d'un modèle est de regarder si celui-ci reproduit bien les données observées. On examine donc les résidus,

$$e_t = x_t - f(x_{t-1}, x_{t-2}, \dots), t=1, \dots, T$$

Evidemment, plus les résidus sont petits, plus le modèle est proche des valeurs observées. Il existe plusieurs fonctions d'erreurs pour mesurer cet écart.

L'écart absolu moyen, noté MAD pour « mean absolute deviation »,

$$\text{MAD}(e) = \frac{1}{T} \sum_{t=1}^T |e_t - \bar{e}|, \text{ où } \bar{e} = \frac{1}{T} \sum_{i=1}^T e_i.$$

Le carré moyen des erreurs, noté MSE pour « mean square error »,

$$\text{MSE}(e) = \frac{1}{T} \sum_{t=1}^T e_t^2.$$

L'erreur quadratique moyenne, notée RMSE pour « root mean square error »,

$$\text{RMSE}(e) = \sqrt{\frac{1}{T} \sum_{t=1}^T e_t^2}.$$

L'erreur absolue moyenne, notée MAE pour « mean absolute error »,

$$\text{MAE}(e) = \frac{1}{T} \sum_{t=1}^T |e_t|.$$

### 2.2.3. Opérateur de retard

Dans les modèles, l'observation de la série au temps  $t$  sera explicitée en fonction des observations passées. Il est alors pratique d'introduire un opérateur de retard ainsi qu'un opérateur de différence.

- Opérateur de retard B (backward) :  $Bx_t = x_{t-1}$
- Opérateur de différence :  $\Delta x_t = x_t - x_{t-1}$

On note que

$$B^2 x_t = B(Bx_t) = Bx_{t-1} = x_{t-2}$$

et plus généralement,

$$B^p x_t = x_{t-p}$$

On a de plus la relation,

$$\Delta x_t = x_t - Bx_t = (1 - B)x_t.$$

Par exemple le modèle de *lissage exponentiel simple* (chapitre 3)

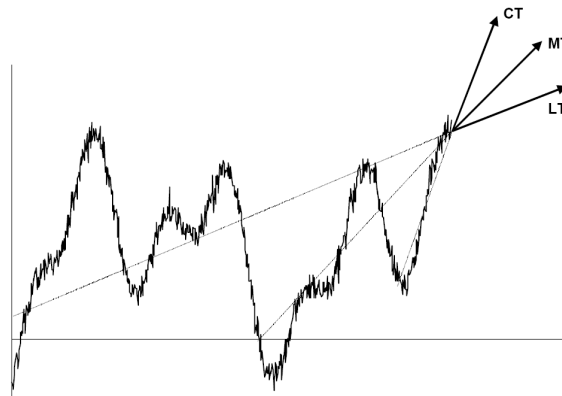
$$x_t = x_{t-1} + \varepsilon_t - (1 - \alpha)\varepsilon_{t-1},$$

où  $\varepsilon_t$  représente les résidus, peut s'écrire

$$\Delta x_t = [1 - (1 - \alpha)B]\varepsilon_t$$

### 2.3. Troisième objectif : prévoir

La prévision de valeurs à des dates futures connaissant le présent et le passé de la série peut être basée sur un modèle ou non. Supposons que l'on dispose des observations  $x_t$  aux temps  $1, \dots, T$ , alors on parle de prévision à l'horizon  $h$  pour toutes valeurs futures de la série,  $x_{T+h}$ .



Exemple issu de Charpentier A.

La prévision dépend du poids que l'on accorde aux observations passées. Les valeurs futures dépendent donc de l'objectif, à savoir si l'on souhaite faire de la prévision à court, moyen ou long terme.

## 3. PROCESSUS STATIONNAIRES

Afin de pouvoir établir un formalisme statistique, nous allons considérer les observations de la série temporelle  $\{x_1, \dots, x_T\}$  comme les réalisations de variables aléatoires  $\{X_1, \dots, X_T\}$ . Les variables aléatoires ne sont en général pas indépendantes. Nous parlerons alors du processus aléatoire  $\{X_t\}$ . La série chronologique observée constitue alors une observation du processus aléatoire. Il est bien évidemment impossible de tirer des conclusions à partir d'un échantillon constitué d'une seule observation. On est donc amené à émettre certaines hypothèses de telle sorte que le problème statistique puisse être résolu. On suppose ainsi que le processus est stationnaire.

### Définition

On dit qu'une suite de variables aléatoires est un *processus stationnaire* si

- les espérances sont constantes,  
$$E(X_t) = \mu, \forall t \in \{1, \dots, T\},$$
- les covariances sont stables par translation  $h$  du temps,  
$$\text{cov}(X_t, X_{t+h}) = \sigma(h), \forall t \in \{1, \dots, T\}.$$

On note qu'un processus stationnaire a une variance constante,

$$\text{var}(X_t) = \sigma(0) = \sigma^2.$$

Si l'espérance est nulle, on parle de processus *centré*.

### Définition de l'auto-corrélation

La fonction  $h \mapsto \sigma(h)$  est appelée *fonction d'auto-covariance* du processus stationnaire.

La fonction

$$\rho: h \mapsto \frac{\sigma(h)}{\sigma^2}$$

est appelée *fonction d'auto-corrélation*.

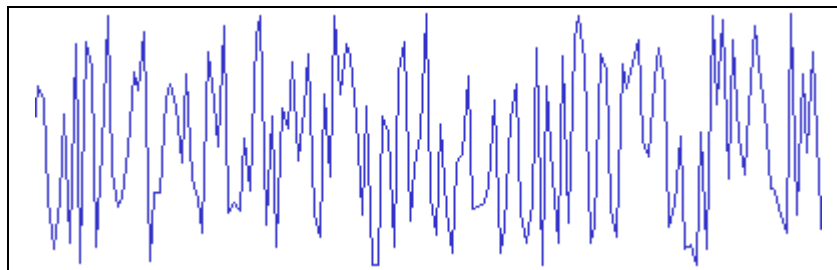
On remarque que ces deux fonctions sont paires,  $\sigma(-h) = \sigma(h)$  et  $\rho(-h) = \rho(h)$ .

### Bruits blancs

Un *bruit blanc* est un processus stationnaire donc la fonction d'auto-covariance est nulle, pour tout  $h$ ,

$$\text{cov}(X_t, X_{t+h}) = 0, \forall t \in \{1, \dots, T\}.$$

Les bruits blancs sont les processus stationnaires les plus élémentaires. Ils ne présentent pas d'intérêt dans la prévision mais servent de base à la construction de tous les modèles présentés dans ce cours.



Réalisation d'un bruit blanc

Il est courant d'utiliser comme bruit blanc, une suite de variables aléatoires indépendantes de loi normale centrée réduites. On parle alors de bruit blanc gaussien centré.

### Définition de l'auto-corrélation partielle

Les fonctions d'auto-covariance et d'auto-corrélation ne font intervenir que les variables aléatoires  $X_t$  et  $X_{t+h}$  sans tenir compte les variables intermédiaires  $X_{t+1}, \dots, X_{t+h-1}$ . La fonction *d'auto-corrélation partielle* pallie ce problème en calculant la corrélation entre les variables  $X_t$  et  $X_{t+h}$  auxquelles ont retranché leurs meilleures explications par les variables intermédiaires. Dans le cas d'un processus stationnaire, la fonction d'auto-corrélation partielle ne dépend que de la distance  $h$  et se définit par

$$r(h) = r(-h), \forall h \neq 0$$

$$r(1) = \rho(1)$$

$$r(h) = \text{corr}(\hat{X}_1, \hat{X}_{h+1}), \forall h \geq 2$$

où

$$\hat{X}_{h+1} = X_{h+1} - \text{EL}(X_{h+1} | X_2, \dots, X_h)$$

avec  $\text{EL}(X_{h+1} | X_2, \dots, X_h)$  la projection orthogonale de  $X_{h+1}$  sur l'espace formé par les combinaisons linéaires de  $X_2, \dots, X_h$ .

## CHAPITRE 2 : LISSAGE PAR MOYENNE MOBILE

L'objectif du lissage par moyenne mobile, « moving average » (MA), est d'éliminer certains aspects d'une série. Ce n'est en général pas un outil de prévision même si nous verrons qu'il est possible de prévoir des valeurs futures de la série.

### 1. DESAISONNALISATION

Supposons que la série s'écrit

$$x_t = z_t + s_t + \varepsilon_t, \text{ pour } t=1, \dots, T.$$

Le but est de trouver une transformation du processus qui annule la composante saisonnière  $S_t$ . On cherche un « filtre »  $\phi$  tel que la série

$$y_t = \phi(x_t) = z_t + \varepsilon_t$$

n'a plus de composante saisonnière

#### 1.1 Moyenne mobile centrée

La *moyenne mobile centrée d'ordre q* de la série  $\{x_t\}_{t=1, \dots, T}$  est définie par

- Si  $q$  est impair tel que  $q=2m+1$ ,

$$y_t = \frac{1}{2m+1} \sum_{i=-m}^m x_{t+i}$$

- Si  $q$  est pair tel que  $q=2m$ ,

$$y_t = \frac{1}{2m} \left[ \frac{1}{2} x_{t-m} + \sum_{i=-m+1}^{m-1} x_{t+i} + \frac{1}{2} x_{t+m} \right]$$

où  $m+1 \leq t \leq T-m$ .

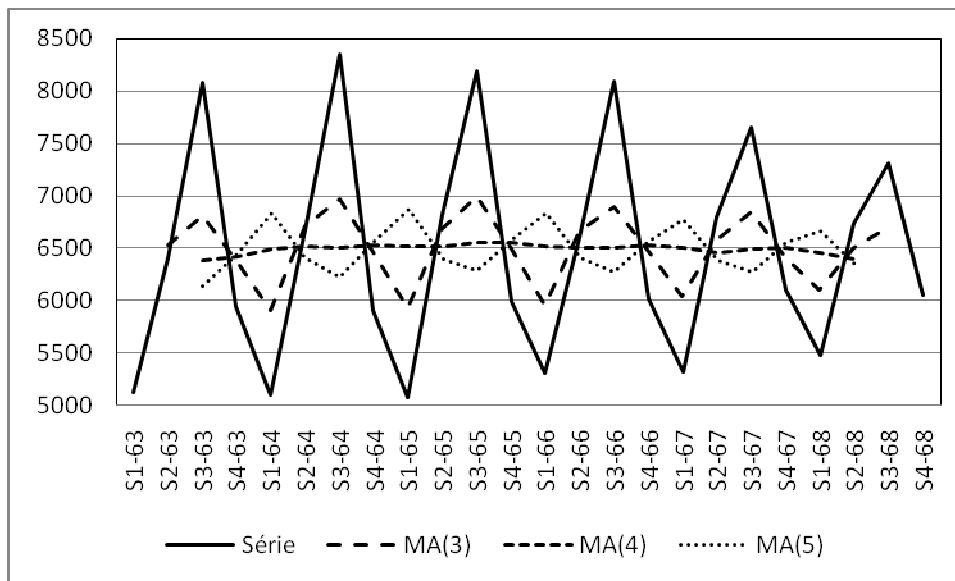
#### Exemple

La moyenne mobile centrée d'ordre 3 :  $y_t = \frac{1}{3} [x_{t-1} + x_t + x_{t+1}]$

La moyenne mobile centrée d'ordre 4 :  $y_t = \frac{1}{4} \left[ \frac{1}{2} x_{t-2} + x_{t-1} + x_t + x_{t+1} + \frac{1}{2} x_{t+2} \right]$

La moyenne mobile centrée d'ordre 5 :  $y_t = \frac{1}{5} [x_{t-2} + x_{t-1} + x_t + x_{t+1} + x_{t+2}]$

Si on applique cette désaisonnalisation à la série trimestrielle du trafic SNCF entre 1963 et 1968, on obtient



Moyennes mobiles centrées sur série trimestrielle du trafic voyageur SNCF entre 1963 et 1968

Ainsi la série obtenue par moyenne mobile centrée d'ordre 4

$$y_t = \frac{1}{4} \left[ \frac{1}{2} x_{t-2} + x_{t-1} + x_t + x_{t+1} + \frac{1}{2} x_{t+2} \right], 3 \leq t \leq T-2$$

est désaisonnalisée.

## 1.2 Moyenne mobile différence

Supposons que la composante saisonnière vérifie une relation de période  $p$  du type  $s_t = s_{t+p}$ , alors la *moyenne mobile différence* définie par

$$y_t = (1 - B^p)x_t$$

annule cette composante.

En effet,

$$y_t = (1 - B^p)x_t = x_t - x_{t-p} = (z_t - z_{t-p}) + (s_t - s_{t-p}) + (\varepsilon_t - \varepsilon_{t-p}) = z_t - z_{t-p} + \varepsilon_t$$

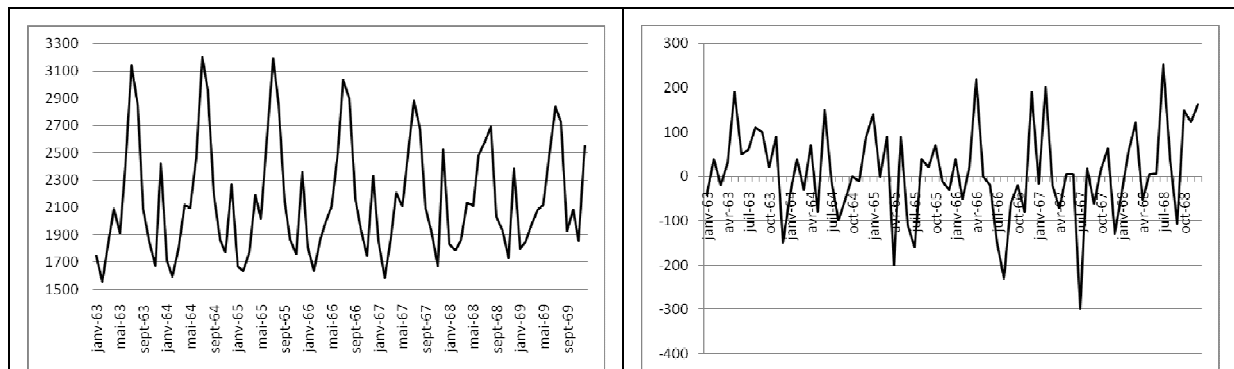
On remarque que la tendance de la série  $\{y_t\}$  n'est plus  $Z_t$  mais  $Z_t - Z_{t-p}$ .

### Exemple

Si on considère la série mensuelle du trafic voyageur alors la série

$$y_t = (1 - B^{12})x_t$$

ne présente plus de composante saisonnière.



Série mensuelle du trafic voyageur SNCF entre 1963 et 1969

## 2. PREVISION D'UNE SERIE SANS COMPOSANTE SAISONNIERE ET SANS TENDANCE

Supposons que la série  $\{y_t\}$  soit sans composante saisonnière (après éventuellement une phase de désaisonnalisation), et sans tendance alors on utilise une *moyenne mobile non centrée d'ordre q* pour calculer les prévisions,

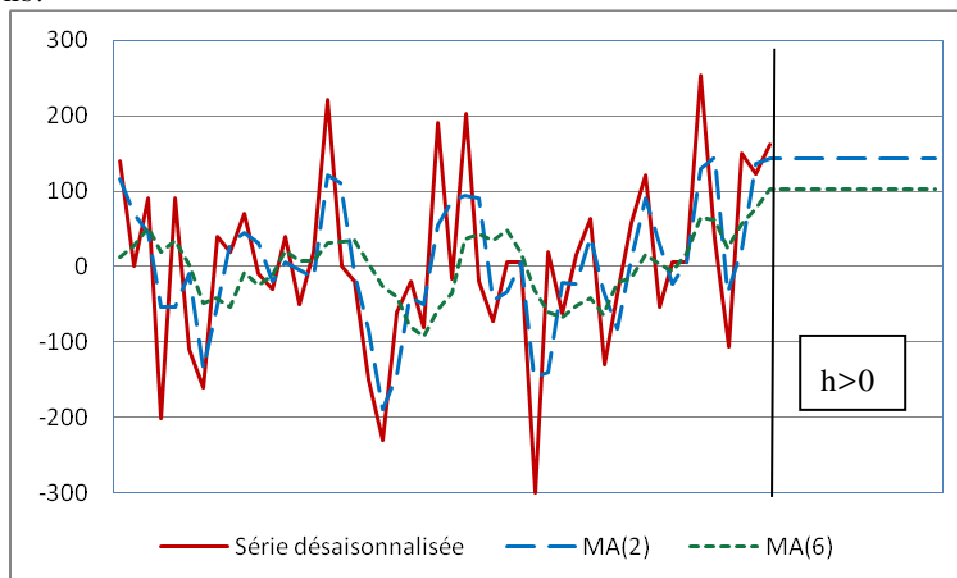
$$\hat{y}_t = \frac{1}{q} \sum_{i=0}^{q-1} y_{t-i}, \quad q-1 \leq t \leq T$$

Les prévisions à l'horizon  $h > 0$  sont alors constantes puisque la série ne présente pas de tendance,

$$\hat{y}_{T+h} = \hat{y}_T$$

### Exemple

Continuons l'exemple précédent des données mensuelles entre 1966 et 1970 du trafic voyageur. Une moyenne mobile différence a donné une série  $\{y_t\}$  sans composante saisonnière. Cette série semble ne pas présenter de tendance, on peut donc appliquer une moyenne mobile non centrée (ordres 2 et 6) pour calculer les prévisions.



Série mensuelle désaisonnalisée du trafic voyageur SNCF entre 1966 et 1970

On remarque que plus l'ordre de la moyenne mobile est grand, plus la prévision est lisse.

### 3. PREVISION D'UNE SERIE SANS COMPOSANTE SAISONNIERE ET AVEC TENDANCE

Supposons que la série  $\{y_t\}$  soit sans composante saisonnière, mais avec tendance linéaire de la forme

$$y_t = at + b + \varepsilon_t.$$

On calcule les moyennes mobiles non centrée d'ordre  $q$  suivantes

$$M_t = \frac{1}{q} \sum_{i=0}^{q-1} y_{t-i} \quad \text{et} \quad MM_t = \frac{1}{q} \sum_{i=0}^{q-1} M_{t-i}$$

On obtient alors les prévisions,

$$\hat{y}_t = 2M_t - MM_t, \quad \text{pour } t \leq T$$

Exercice : On pose  $y_t = at + b$

- 1) Calculez  $M_t$ . Que constatez-vous sur  $y_t - \hat{y}_t$  si  $\hat{y}_t = M_t$  ?
- 2) Calculez  $MM_t$ . Que constatez-vous sur  $y_t - \hat{y}_t$  si  $\hat{y}_t = 2M_t - MM_t$  ?

Les prévisions à l'horizon  $h > 0$  respectent la relation linéaire,

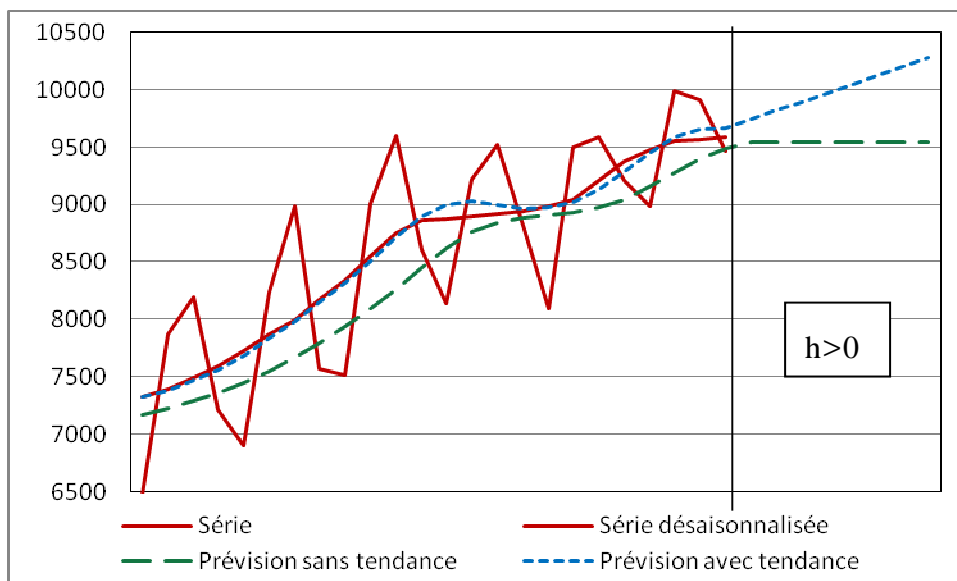
$$\hat{y}_{T+h} = \hat{y}_T + h\hat{a}_T$$

où

$$\hat{a}_T = \frac{2}{q-1} (M_T - MM_T) \quad (\text{pente})$$

#### Exemple

Prenons le trafic voyageur entre 1970 et 1980 pour la série trimestrielle. Celui-ci présente une forte saisonnalité qui peut être lissée par une moyenne mobile MA(4). On constate que la prévision faite en tenant compte de la tendance approche mieux la série désaisonnalisée que celle faite sans tenir compte de la tendance. On remarque que la prévision à l'horizon  $h > 0$  continue à tenir compte de la tendance.



Série trimestrielle du trafic voyageur SNCF entre 1973 et 1980

## CHAPITRE 3 : PREVISION PAR LISSAGE EXPONENTIEL

L'objectif du lissage exponentiel est de prévoir une série sans en chercher préalablement un modèle. Cette méthode s'applique à des séries qui ne présentent pas de composante saisonnière. Nous supposons donc par la suite que la série est de la forme

$$y_t = z_t + \varepsilon_t$$

où  $z_t$  représente la tendance.

### 1. PREVISION D'UNE SERIE SANS COMPOSANTE SAISONNIERE ET SANS TENDANCE

La méthode du *lissage exponentiel simple* permet de prévoir une série sans tendance,

$$y_t = b + \varepsilon_t.$$

Elle se définit comme une somme pondérée des observations passées,

$$\hat{y}_{t+h} = \alpha \sum_{k=0}^{t-1} (1-\alpha)^k y_{t-k},$$

où  $0 < \alpha < 1$ . On donne un poids d'autant moins important que les observations sont loii dans le passé avec une décroissance exponentielle (d'où le nom de la méthode),

- $\alpha$  proche de 0 permet de prendre en compte un passé lointain. Les prévisions sont alors rigides et peu sensibles aux variations conjoncturelles,
- $\alpha$  proche de 1 permet de ne prendre en compte que les valeurs récentes. Les prévisions sont plus souples et sensibles aux fluctuations.

Etant donné que dans l'expression ci-dessus,  $\hat{y}_{t+h}$  ne dépend pas de  $h$ , nous allons le noter  $\hat{y}_{t+1}$ , prévision faite pour l'instant  $t+1$  à l'aide des informations disponibles à l'instant  $t$ . Nous avons alors la formule de récurrence

$$\hat{y}_{t+1} = \alpha y_{t+1} + (1-\alpha)\hat{y}_t, \quad 1 \leq t \leq T-1$$

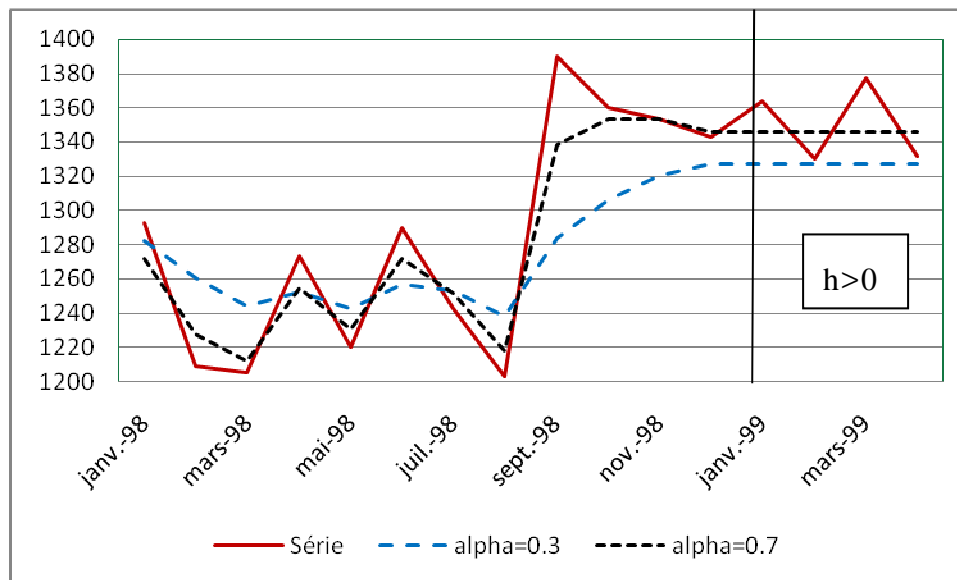
Les prévisions à l'horizon  $h > 0$  sont constantes puisque la série est sans tendance,

$$\hat{y}_{T+h} = \hat{y}_T$$

La formule de récurrence peut être initialisée

- soit par l'observation au temps 1,  $\hat{y}_1 = y_1$
- soit par la moyenne des valeurs observées
- soit par la moyenne des quelques-unes des premières observations

L'initialisation n'a pas beaucoup d'influence si t est grand.



Chiffre d'affaire d'une entreprise sur 18 mois entre janvier 1998 et avril 1999 (A. Charpentier)

On remarque sur le graphique que pour une valeur de  $\alpha=0.7$ , la prévision par lissage exponentiel retranscrit bien la rupture de septembre 1998. En revanche, la prévision pour  $\alpha=0.3$  est sous-estimée à partir de septembre 1998 car elle prend en compte un passé plus lointain. Les valeurs faibles du chiffre d'affaire observées avant la rupture pèsent donc dans la prévision.

**Remarque** : Le choix de la constante de lissage  $\alpha$  est donc stratégique. Il est assez usuel d'utiliser une valeur voisine de 0.7. Cependant le choix optimal de  $\alpha$  peut se faire par minimisation de l'erreur quadratique moyenne. C'est-à-dire de façon à ce que les prévisions « collent » au mieux avec les observations.

## 2. PREVISION D'UNE SERIE SANS COMPOSANTE SAISONNIERE ET AVEC TENDANCE

La méthode du *lissage exponentiel double* permet de prévoir une série avec tendance du type

$$y_t = at + b + \varepsilon_t.$$

De la même façon que pour la moyenne mobile avec tendance, on calcule

$$E_{t+1} = \alpha y_{t+1} + (1 - \alpha)E_t \text{ et } EE_{t+1} = \alpha EE_{t+1} + (1 - \alpha)EE_t$$

On obtient les prévisions avec les formules

$$\hat{y}_t = 2E_t - EE_t, \text{ pour } t \leq T$$

Les prévisions à l'horizon h respectent une relation linéaire,

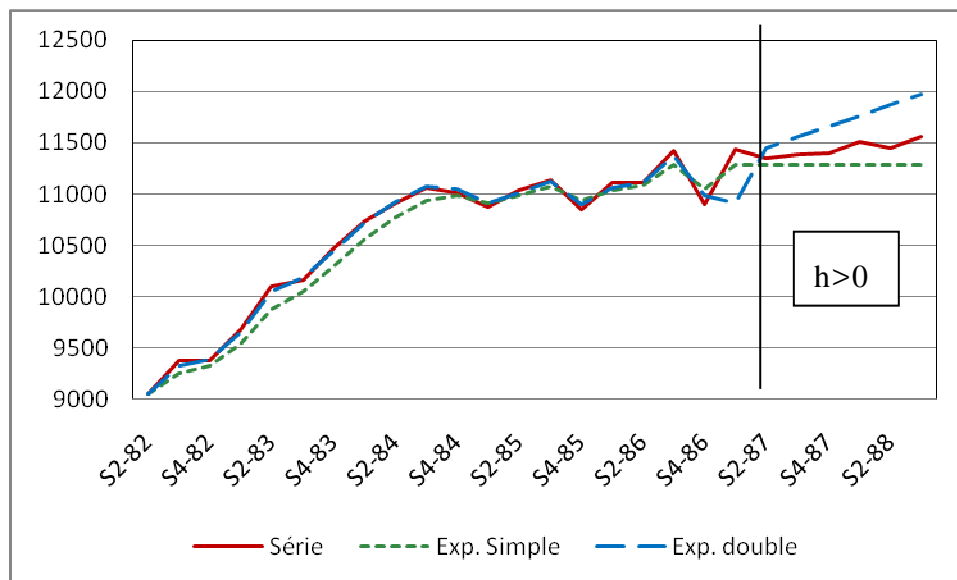
$$\hat{y}_{T+h} = \hat{a}_T h + \hat{b}_T$$

où

$$\hat{b}_T = 2E_T - EE_T \text{ (moyenne)}$$

$$\hat{a}_T = \frac{\alpha}{1-\alpha} (E_T - EE_T) \text{ (pente)}$$

Le graphique suivant représente un indice d'activité en 1982 et 1988 pour des données semestrielles. On vérifie que la prévision à l'horizon  $h > 0$  suit une équation de droite fonction de h alors qu'avec un lissage exponentielle simple la prévision reste constante.



Indice d'activité entre 1982 et 1988 données semestrielles (A. Charpentier)

### 3. AJOUT D'UNE COMPOSANTE SAISONNIERE

Les méthodes vues précédemment ne s'appliquent que pour des séries sans composantes saisonnières. Une première phase de désaisonnalisation a peut-être été nécessaire pour appliquer les méthodes de prévision. On peut alors souhaiter retrouver cette composante saisonnière dans la prévision.

Reprenons le cas du trafic voyageur de la SNCF entre 1970 et 1978. Les données trimestrielles présentent une forte saisonnalité et la série est lissée grâce à la moyenne mobile centrée d'ordre 4

$$y_t = \frac{1}{4} \left[ \frac{1}{2} x_{t-2} + x_{t-1} + x_t + x_{t+1} + \frac{1}{2} x_{t+2} \right].$$

La composante saisonnière est alors obtenue en faisant la différence entre la série brute  $x_t$  et la série lissée  $y_t$  puis en calculant la moyenne pour chaque semestre. Par exemple la composante saisonnière du semestre 1 est

$$\Delta_{S1} = \frac{1}{8} [(x_{S1-71} - y_{S1-71}) + \dots + (x_{S1-78} - y_{S1-78})]$$

On normalise alors les composantes saisonnières de façon à ce que la somme sur les 4 trimestres soit nulle

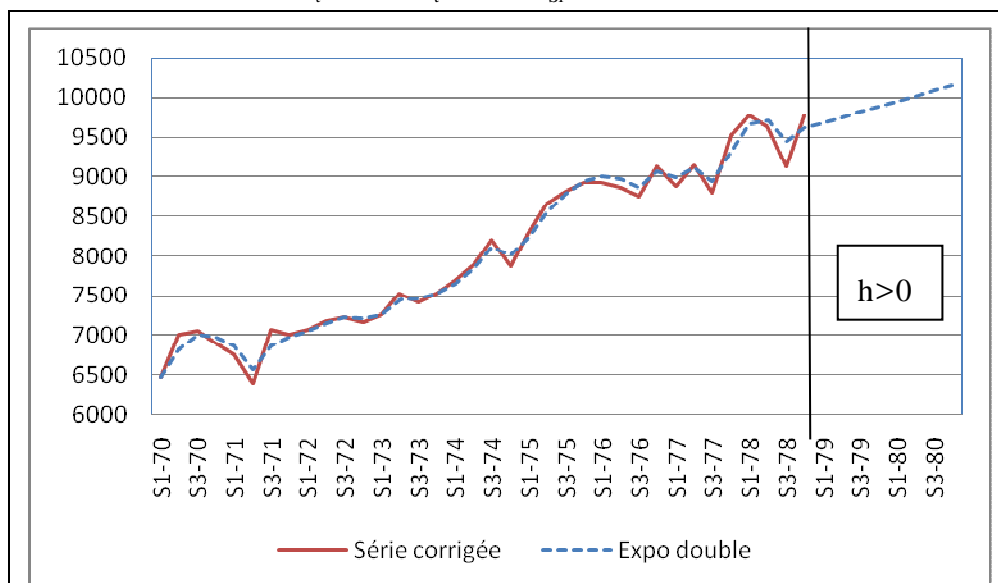
$$\delta_{S1} = \Delta_{S1} - \frac{1}{4} (\Delta_{S1} + \Delta_{S2} + \Delta_{S3} + \Delta_{S4}).$$

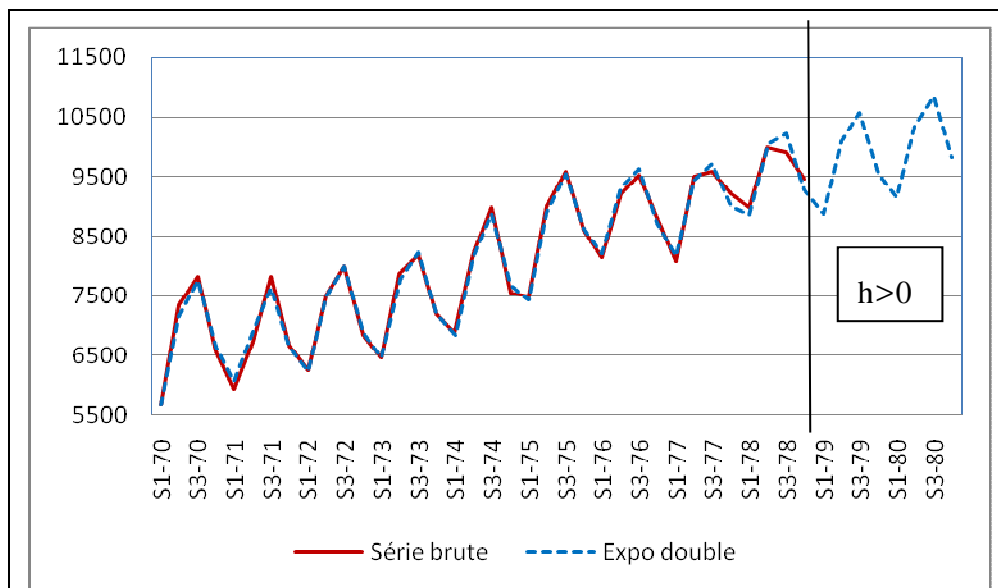
On construit alors la série corrigée de ses variations saisonnières comme la différence entre la série brute et la composante saisonnière du mois correspondant

$$\begin{aligned} z_t(S1) &= x_t(S1) - \delta_{S1} & z_t(S2) &= x_t(S2) - \delta_{S2} \\ z_t(S3) &= x_t(S3) - \delta_{S3} & z_t(S4) &= x_t(S4) - \delta_{S4} \end{aligned}$$

C'est sur la série corrigée que l'on applique une des méthodes de prévision vues précédemment et ainsi obtenir la prévision  $\hat{z}_t$ . La prévision sur la série initiale se fait en ajoutant la composante saisonnière du trimestre correspondant

$$\hat{x}_t(Si) = \hat{z}_t(Si) + \delta_{Si}, i=1, \dots, 4$$





Série trimestrielle du trafic voyageur SNCF entre 1970 et 1978

## 4. LISSAGE EXPONENTIEL DE HOLT-WINTER

### 4.1. Série sans composante saisonnière

Ce modèle traite les séries où la tendance et l'évolution de la tendance peuvent fluctuer différemment l'une de l'autre. Donc au lieu de faire un lissage exponentiel double, on va faire un lissage exponentiel simple sur la tendance et un autre lissage exponentiel simple sur l'évolution. Nous avons donc les deux relations de récurrence

$$\hat{y}_{t+1} = \alpha y_{t+1} + (1 - \alpha)(\hat{y}_t + \hat{T}_t)$$

$$\hat{T}_{t+1} = \gamma(\hat{y}_{t+1} - \hat{y}_t) + (1 - \gamma)\hat{T}_t$$

La première relation exprime que l'estimation de la série à l'instant t+1 est une moyenne pondérée de la valeur observée à l'instant t+1 avec la prévision faite à l'instant t.

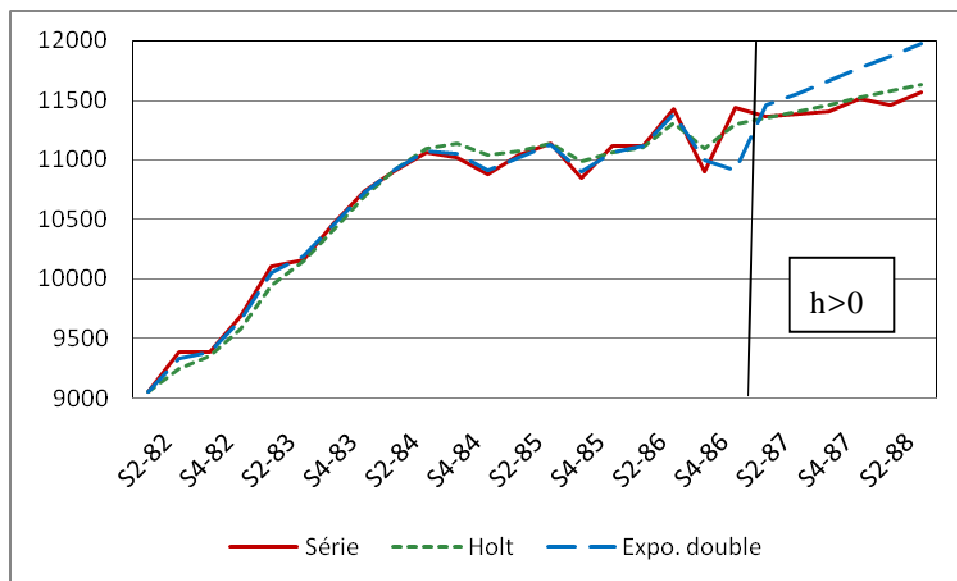
La deuxième relation exprime que l'estimation de la pente à l'instant t+1 est une moyenne pondérée de la pente de la droite passant par (t+1, y<sub>t+1</sub>) et (t, y<sub>t</sub>) d'une part avec la dernière estimation de cette pente T<sub>t</sub> d'autre part.

On prendra comme valeurs initiales

$$\begin{cases} \hat{y}_1 = y_1 \\ \hat{T}_1 = 0 \end{cases}$$

Les prévisions à l'horizon h sont données par la droite,

$$\hat{y}_{T+h} = \hat{y}_T + h\hat{T}_T$$



Indice d'activité entre 1982 et 1988 données semestrielles (A. Charpentier)

## 4.2. Série avec composante saisonnière

Le modèle de Holt-Winter permet de traiter directement une série avec une composante saisonnière,

$$y_t = z_t + s_t + \varepsilon_t,$$

où  $s_{t+p} = s_t$ , avec  $p$  le nombre de périodes couvrant le cycle saisonnier ( $p=4$  pour des données trimestrielles,  $p=12$  pour des données mensuelles). Au lissage de la moyenne et de la tendance, il faut ajouter le lissage de la saisonnalité. Nous avons donc les trois relations de récurrence pour  $t \geq p$ ,

$$\hat{y}_{t+1} = \alpha(y_{t+1} - \hat{s}_{t+1-p}) + (1-\alpha)(\hat{y}_t + \hat{\tau}_t)$$

$$\hat{\tau}_{t+1} = \gamma(\hat{y}_{t+1} - \hat{y}_t) + (1-\gamma)\hat{\tau}_t$$

$$\hat{s}_{t+1} = \beta(y_{t+1} - \hat{y}_{t+1}) + (1-\beta)\hat{s}_{t+1-p}$$

Les prévisions à l'horizon  $h$  sont données par la droite corrigée par la saisonnalité,

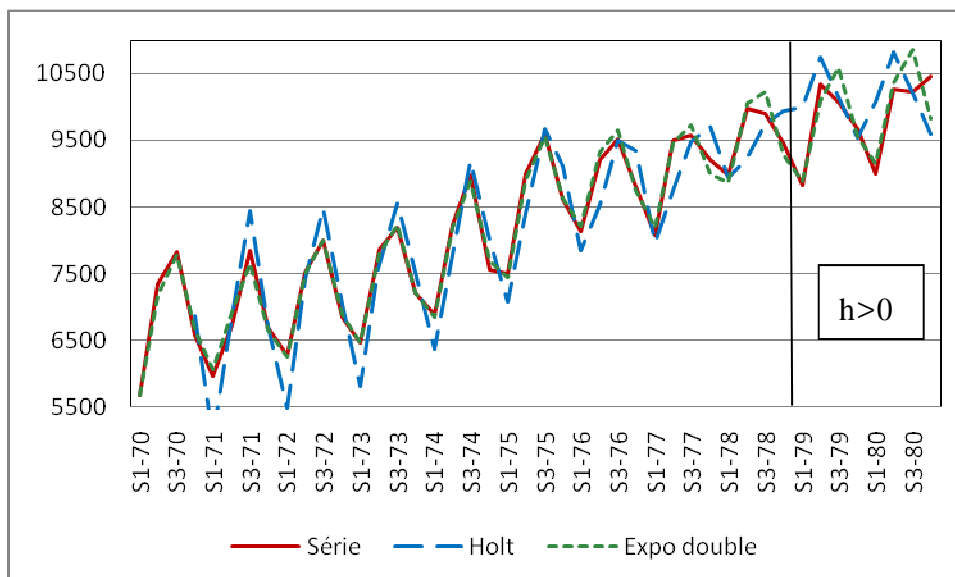
$$\hat{y}_{T+h} = \hat{y}_T + h\hat{\tau}_T + s_j, \text{ où } j=T-p+h$$

On prendra comme valeurs initiales

$$\begin{cases} \hat{y}_p = \frac{1}{p} \sum_{i=1}^p y_i \\ \hat{T}_p = \frac{1}{p^2} \left( \sum_{i=p+1}^{2p} y_i - \sum_{i=1}^p y_i \right) \\ \hat{s}_j = y_j - \hat{y}_p \text{ ou } \hat{s}_j = \bar{y} - \bar{y}_j \quad j=1, \dots, p \end{cases} .$$

où  $\bar{y}_j$  est la moyenne des observations sur la période j.

Il semble que dans l'exemple ci-dessous, la modélisation par lissage exponentiel double avec ajout de composante saisonnière donne de meilleurs résultats que le modèle de Holt. La difficulté de ce modèle est de trouver la valeur optimale pour maintenant trois constantes,  $\alpha$ ,  $\beta$ ,  $\gamma$ .



Série trimestrielle du trafic voyageur SNCF entre 1970 et 1980

# CHAPITRE 4 : PREVISION PAR REGRESSION LINEAIRE

## 1. LE MODELE LINEAIRE

On considère que la série  $\{X_t\}$  est la somme de deux composantes déterministes, une tendance  $Z_t$ , une saisonnalité  $S_t$ , et une composante aléatoire  $\varepsilon_t$ ,

$$X_t = Z_t + S_t + \varepsilon_t.$$

On suppose que  $Z_t$  et  $S_t$  sont des combinaisons linéaires de fonctions connues dans le temps, *i.e*

$$\begin{cases} Z_t = \beta_1 Z_t^1 + \dots + \beta_m Z_t^m \\ S_t = \gamma_1 S_t^1 + \dots + \gamma_n S_t^n \end{cases}.$$

La forme de la composante saisonnière  $S_t$  dépend du type de données et de la forme de la saisonnalité. On considérera ici des fonctions indicatrices,

$$S_t^i = \begin{cases} 0 & \text{si } t = \text{période } i \pmod{i} \\ 1 & \text{si } t \neq \text{période } i \pmod{i} \end{cases}$$

Par exemple pour des données trimestrielles on a

$$S_t = \gamma_1 S_t^1 + \gamma_2 S_t^2 + \gamma_3 S_t^3 + \gamma_4 S_t^4,$$

où  $S_t^i$  est la fonction indicatrice du trimestre  $i$ .

La composante tendancielle reflète la croissance moyenne de la série. Elle peut prendre plusieurs formes dont la plus courante est la tendance linéaire,

$$Z_t = \beta_1 + \beta_2 t,$$

On a alors  $Z_t^1 = 1$  et  $Z_t^2 = t$ .

La modélisation de la série repose sur l'estimation des paramètres  $\beta_1, \dots, \beta_m$  et  $\gamma_1, \dots, \gamma_n$  à partir des observations.

### Hypothèses sur le modèle

Comme pour toute régression linéaire, il convient de supposer qu'il n'existe pas de relation linéaire entre les variables explicatives (tendancielle et saisonnières).

On suppose les hypothèses classiques sur les erreurs à savoir les erreurs sont

- centrées :  $E(\varepsilon_t) = 0$

- de même variance :  $\text{var}(\varepsilon_t) = \sigma^2$
- non corrélées :  $\text{cov}(\varepsilon_t, \varepsilon_{t+h}) = 0, h > 0$ .

## 2. IDENTIFICATION DU MODELE

On considère donc le modèle de la forme

$$X_t = \sum_{i=1}^m \beta_i Z_t^i + \sum_{j=1}^n \gamma_j S_t^j + \varepsilon_t, t=1, \dots, T$$

où les paramètres  $\beta_i$  et  $\gamma_j$  sont à estimer à partir des observations de la série.

### 2.1. Estimation par moindres carrés ordinaires

De façon classique, la méthode des moindres carrés ordinaires détermine les paramètres inconnus du modèle de manière à minimiser le carré des erreurs,

$$(\hat{\beta}_i, \hat{\gamma}_j) = \arg \min \{\varepsilon_t^2\},$$

avec

$$\varepsilon_t = X_t - \sum_{i=1}^m \beta_i Z_t^i + \sum_{j=1}^n \gamma_j S_t^j.$$

Si on prend la notation matricielle du modèle,

$$X = \beta Z + \lambda S + \varepsilon$$

où  ${}^t\beta = [\beta_1, \dots, \beta_m]$  et  ${}^t\gamma = [\gamma_1, \dots, \gamma_n]$  et

$$Z = [Z_t^i]_{\substack{i=1, \dots, m \\ t=1, \dots, T}} = \begin{bmatrix} \vdots & & \vdots \\ Z^1 & \dots & Z^m \\ \vdots & & \vdots \end{bmatrix} \text{ et } S = [S_t^j]_{\substack{j=1, \dots, n \\ t=1, \dots, T}} = \begin{bmatrix} \vdots & & \vdots \\ S^1 & \dots & S^n \\ \vdots & & \vdots \end{bmatrix}$$

Simplifions les notations en posant

$$X = D\delta + \varepsilon,$$

où  $D = [Z \mid S]$  et  ${}^t\delta = [\beta_1 \dots \beta_m \mid \gamma_1 \dots \gamma_n]$ .

#### Exemple

Considérons le modèle dont la composante tendancielle est linéaire et la composante saisonnière est trimestrielle,

$$\begin{cases} Z_t = \beta_1 + \beta_2 t \\ S_t = \gamma_1 S_t^1 + \gamma_2 S_t^2 + \gamma_3 S_t^3 + \gamma_4 S_t^4 \end{cases}$$

où  $S_t^i$  représente le  $i^{\text{ème}}$  trimestre. Alors la matrice du modèle s'écrit

$$D = \left( \begin{array}{cc|cccc} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 & 0 \\ 1 & 3 & 0 & 0 & 1 & 0 \\ 1 & 4 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & & & & \\ 1 & T-3 & 1 & 0 & 0 & 0 \\ 1 & T-2 & 0 & 1 & 0 & 0 \\ 1 & T-1 & 0 & 0 & 1 & 0 \\ 1 & T & 0 & 0 & 0 & 1 \end{array} \right)$$

La méthode des moindres carrés donne alors l'estimation des paramètres

$$\hat{\delta} = [{}^tDD]^{-1} {}^tDX.$$

On obtient alors les prévisions des observations

$$\hat{X} = \hat{\delta}D.$$

Pour que l'estimateur des paramètres existe, il faut donc que la matrice  ${}^tDD$  soit inversible, *i.e* il faut que les colonnes de la matrice  $D$  soient linéairement indépendantes (absence de relation linéaire entre les variables explicatives). Ce qui n'est généralement pas le cas quand on la composante tendancielle prend en compte une constante. La colonne  $\mathbf{1}$ , associée au paramètre  $\beta_1$ , est alors la somme des colonnes associées aux composantes saisonnières. Pour pallier ce problème, on supprime la colonne  $\mathbf{1}$  de la matrice du modèle. On peut alors estimer les coefficients du modèle grâce aux moindres carrés. Si on note  $\tilde{\gamma}_i$  l'estimation du coefficient associé à  $S_t^i$  alors au niveau de la représentation on pose

$$\hat{\beta}_1 = \frac{1}{n} \sum_{i=1}^n \tilde{\gamma}_i \text{ et } \hat{\gamma}_i = \tilde{\gamma}_i - \hat{\beta}_1.$$

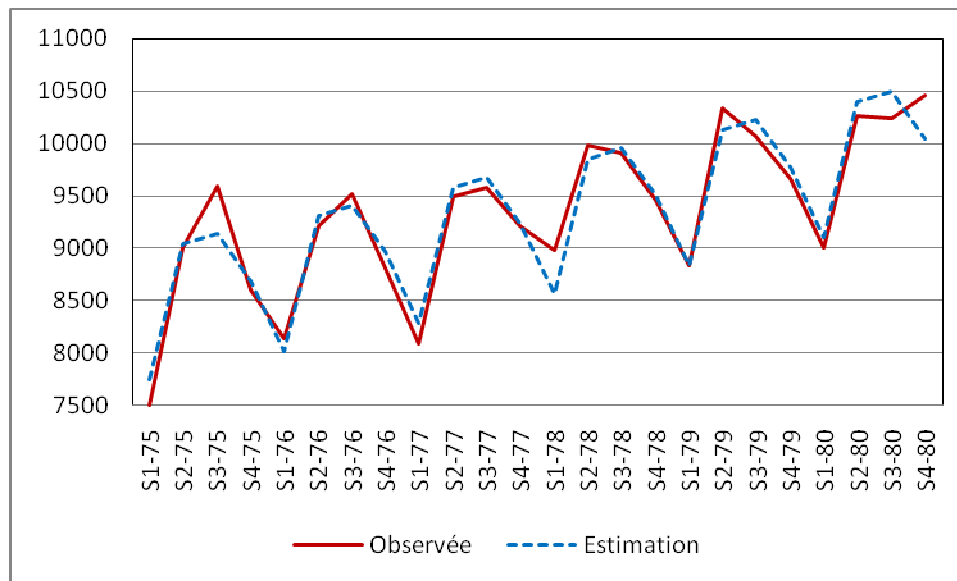
### Exemple

Les données trimestrielles du trafic SNCF entre 1975 et 1980 présente une composante tendancielle linéaire et une saisonnalité trimestrielle. Nous allons donc utiliser le modèle précédent. On obtient l'estimation des coefficients

Coefficients	$\beta_2$	$\gamma_1$	$\gamma_2$	$\gamma_3$	$\gamma_4$
Estimation	67.69	7681	8906	8938	8422

On en déduit l'estimation de  $\beta_1$  puis les coefficients  $\gamma$  corrigés

Coefficients	$\beta_1$	$\beta_2$	$\gamma_1$	$\gamma_2$	$\gamma_3$	$\gamma_4$
Estimation	8487	67.69	-806	418.9	541.6	-64.8



Série trimestrielle du trafic voyageur SNCF entre 1975 et 1980

## 2.2. Validation du modèle

Tout d'abord rappelons que sous l'hypothèse  $E(\epsilon_t) = 0$ , les estimateurs ci-dessus sont sans biais,

$$E(\hat{\delta}_i) = \delta_i.$$

Leur variance peut être estimée par

$$\text{var}(\hat{\delta}_i) = \hat{\sigma}_i^2 = s^2 [({}^t DD)^{-1}]_{ii} \quad \text{où} \quad s^2 = \frac{1}{T - n - m} \sum_{t=1}^T e_t^2,$$

avec  $e_t = x_t - \hat{x}_t$ .

### 2.2.1 Tests sur la pertinence du modèle

- Coefficient de détermination

Le coefficient de détermination,

$$R^2 = \frac{\sum_{t=1}^T (x_t - \bar{x})^2 - \sum_{t=1}^T (x_t - \hat{x}_t)^2}{\sum_{t=1}^T (x_t - \bar{x})^2}$$

permet de quantifier la variance expliquée par le modèle. En fait,  $R$  est le coefficient de corrélation multiple. Dans une régression avec constante, nous avons forcément,  $0 \leq R^2 \leq 1$ . Plus  $R^2$  est proche de 1, plus la relation linéaire entre variable expliquée et variables explicatives est justifiée.

Le coefficient de détermination a tendance à augmenter à mesure que l'on ajoute des variables dans le modèle. De ce fait, il est inopérant si l'on veut comparer des modèles comportant un nombre différent de variables. Il est conseillé dans ce cas d'utiliser le coefficient de détermination ajusté qui est corrigé des degrés de libertés,

$$R_a^2 = 1 - \frac{n + m - 1}{T - (n + m)} (1 - R^2).$$

- Signification du modèle

Le modèle n'est pas pertinent si tous les coefficients de la régression (sauf la constante) sont nuls. On rappelle que sous l'hypothèse  $H_0$  : «  $\delta_1 = \dots = \delta_{m+n} = 0$  » la variable aléatoire

$$F = \frac{R^2}{1 - R^2} \frac{T - (n + m)}{n + m - 1}$$

suit une loi de Fisher à  $(n+m-1; n+m)$  degrés de liberté.

Ainsi si le variable  $F$  prend une valeur supérieure à celle donnée par la table de la loi de Fisher associée à un niveau de confiance de  $\alpha$  (par exemple 5%), on peut rejeter l'hypothèse  $H_0$  et considérer que le modèle est pertinent.

- Signification d'un terme du modèle

Un des principes de la modélisation statistique est la parcimonie, c'est-à-dire qu'il importe de ne retenir que les termes essentiels du modèle. Pour ce faire, on teste les coefficients du modèle un à un. On rappelle que sous l'hypothèse  $H_0$  : «  $\delta_i = 0$  » la variable aléatoire

$$t_i = \frac{|\hat{\delta}_i|}{\hat{\sigma}_i}$$

suit une loi de Student à  $T-(n+m)$  degrés de liberté.

Ainsi si le variable  $t_i$  prend une valeur supérieure à celle donnée par la table de la loi de Student associée à un niveau de confiance de  $\alpha$  (par exemple 5%), on peut rejeter l'hypothèse  $H_0$  et considérer que le terme associé à  $\delta_i$  est pertinent dans le modèle.

### Exemple

Sur l'exemple précédent nous obtenons les résultats très satisfaisants suivants

$R^2$	$R_a^2$	F	$t_2$	$t_3$	$t_4$	$t_5$	$t_6$
0.93	0.91	292.7	9.96	6.76	3.39	3.52	0.49
F(5,18,1%)=4.25			t(18,1%)=2.55				

Seule la variable du trimestre 4 ne semble pas influente dans le modèle.

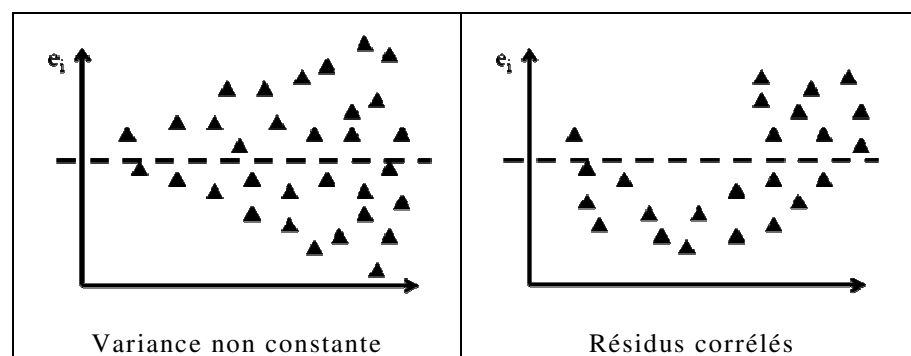
### 2.2.2. Tests sur les hypothèses du modèle

- Corrélation des variables explicatives

Les variables explicatives ne doivent pas avoir de relation linéaire entre elles. Afin de s'assurer de cette première hypothèse, il convient d'examiner la matrice de corrélation entre les variables explicatives.

- Etudes graphique des résidus

La loi des erreurs doit être normale centrée (hypothèse de normalité) et d'écart-type constant  $\sigma$  (hypothèse d'homoscédasticité). Il existe des tests statistiques pour valider l'hypothèse de normalité. Nous nous contenterons ici d'une étude graphique des résidus (résidus vs prévisions). Celui-ci ne doit laisser apparaître aucune tendance.



- Détection de valeurs aberrantes (outliers)

Une valeur aberrante peut remettre en question l'hypothèse de normalité. Pour détecter une valeur aberrante, on peut étudier les résidus studentisés,

$$\tilde{e}_t = \frac{x_t - \hat{x}_t}{s\sqrt{1 - h_t}},$$

où  $h_t = [D^t(DD)^{-1}D]_{tt}$ . En effet, lorsque T est grand, les résidus studentisés doivent rester entre -2 et 2. Un résidu sortant de cet intervalle peut indiquer une valeur aberrante. Cependant une valeur peut être aberrante sans que son résidu soit important. Une autre méthode pour détecter une valeur aberrante consiste à introduire une nouvelle variable explicative  $B_i$  égale à 0 sauf pour la ième observation. Plus le coefficient associé à  $B_i$  est significatif (test de Student), plus la valeur est aberrante.

- Auto corrélation des résidus

Les erreurs doivent être indépendantes. Dans le cas où les erreurs dépendent du temps, on va tester l'hypothèse de  $H_0$  de non corrélation des erreurs à l'aide du test de Durbin-Watson. La statistique de ce test est donné par

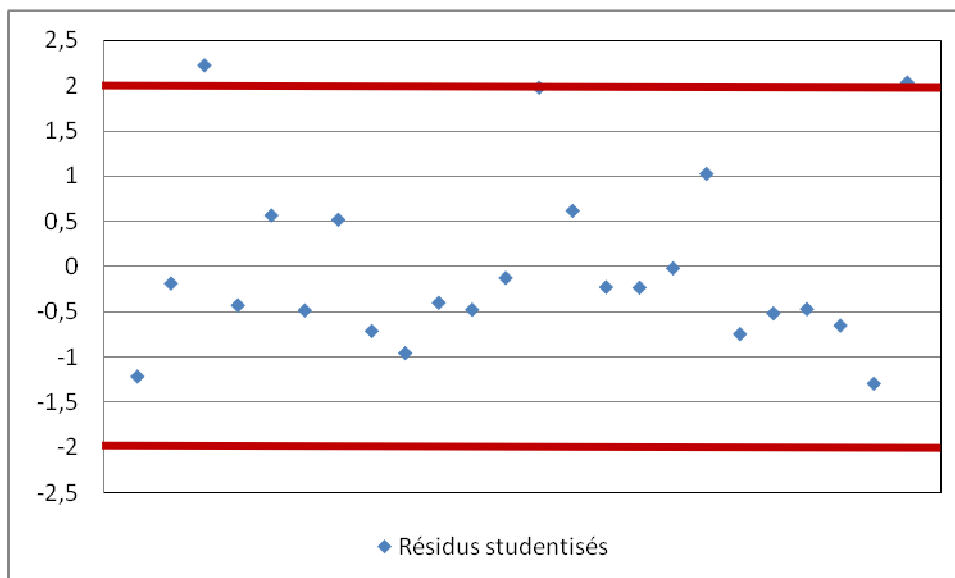
$$DW = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2}.$$

On a toujours  $0 \leq DW \leq 4$ . Des valeurs proches de 0 ou de 4 expriment une auto corrélation respectivement positive ou négative. Il est difficile de tabuler cette statistique. Dans la pratique on utilise la règle suivante, si

$$DW \in \left[ 2 - \frac{4}{\sqrt{T}}, 2 + \frac{4}{\sqrt{T}} \right],$$

alors on accepte l'absence d'auto corrélation, sinon on la refuse.

### Exemple



Les résidus sont centrés et ne présentent pas de structure particulière. Le test de Durbin-Watson,

$$DW = 1.82 \in [1.18 ; 2.82]$$

permet de valider l'absence d'auto corrélation des résidus. On détecte une valeur aberrante sur le graphique, ce qui est peu pour remettre en cause l'hypothèse de normalité. Les tests pour détecter des valeurs aberrantes en introduisant une variable  $B_i$  ne donnent rien au niveau 5% et détectent les observations n°3 et n°14 au niveau 1%.

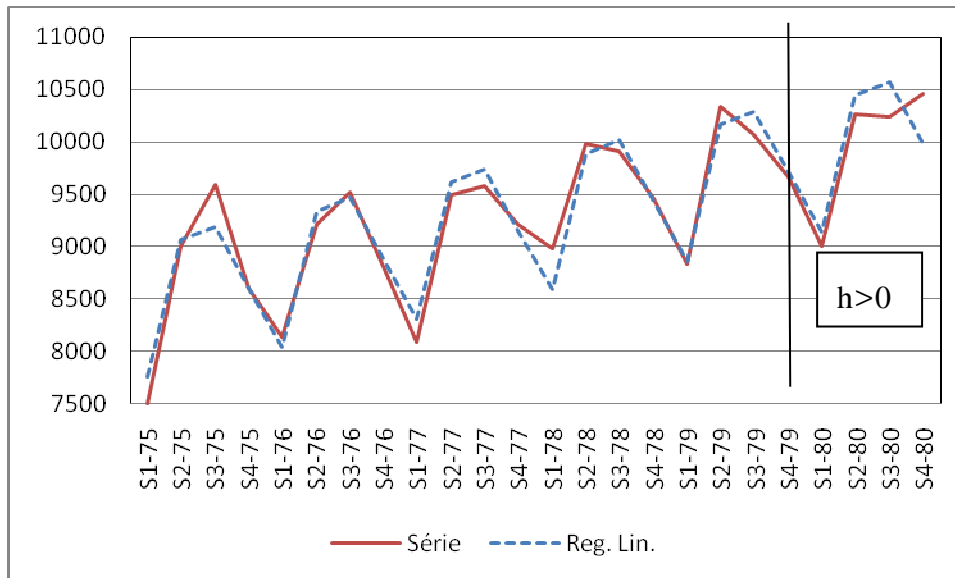
### 2.3. Prévision à l'horizon h

Soit  $h \geq 1$ . On suppose que le modèle reste valide en  $T+h$ . La variable  $X_{T+h}$  peut alors être approchée par

$$\hat{X}_{T+h} = \sum_{i=1}^m \hat{\beta}_i Z_{T+h}^i + \sum_{j=1}^n \hat{\gamma}_j S_{T+h}^j.$$

Cette prévision est la meilleure prévision sans biais linéaire en  $X_1, \dots, X_T$  (au sens de l'erreur quadratique moyenne).

Dans l'exemple précédent, on utilise les observations de 1975 à 1979 pour estimer les paramètres du modèle puis on prévoit les valeurs de l'année 1980 avec la formule ci-dessus.



*Série trimestrielle du trafic voyageur SNCF entre 1975 et 1980*

# CHAPITRE 5 : ERREURS DE PREVISION ET AUTO CORRELATION

Les différentes méthodes de prévision vues précédemment aboutissent toutes à des erreurs de prévision  $\varepsilon_t$ . Si la méthode utilisée est bonne, on peut émettre deux hypothèses sur les résidus.

- Premièrement, les erreurs de prévision doivent être centrées,  $E[\varepsilon_t]=0$ . Dans le cas contraire, les prévisions seraient systématiquement biaisées.
- Deuxièmement, les erreurs de prévision ne doivent plus contenir d'information sur les valeurs futures. En particulier, les séries  $\{\varepsilon_t\}$  et  $\{\varepsilon_{t+1}\}$  ne peuvent être corrélées. Dans le cas contraire,  $\varepsilon_t$  contiendrait de l'information permettant de prévoir  $\varepsilon_{t+1}$  et donc améliorer la prévision de  $y_{t+1}$ . On peut donc généraliser en affirmant qu'il ne doit pas y avoir de corrélation entre les séries  $\{\varepsilon_t\}$  et  $\{\varepsilon_{t+k}\}$  pour tout  $k>0$ .

Nous allons donc émettre que si la méthode est bonne alors le processus  $\{\varepsilon_t\}$  est un bruit blanc.

Pour tester les hypothèses sur les erreurs de prévision, nous allons utiliser les résidus,

$$e_t = x_t - \hat{x}_t, t=1, \dots, T$$

où  $x_t$  est la série observée et  $\hat{x}_t$  est la prévision obtenue par une des méthodes vues précédemment.

## 1. TEST SUR LA MOYENNE DES RESIDUS

Dans un premier temps, nous allons vérifier si les erreurs de prévision sont centrées. On suppose donc que les résidus sont les réalisations d'un échantillon tiré d'une population d'espérance  $m$  et de variance  $\sigma^2$ . Alors sous l'hypothèse  $H_0 : \langle m=0 \rangle$ , la statistique

$$Z = \frac{\bar{e}}{\text{RMSE}(e)} \sqrt{T}$$

suit approximativement une loi normale  $N(0,1)$  pour  $T$  assez grand.

Ainsi si la variable  $Z$  prend une valeur inférieure (en valeur absolue) à celle donnée par la table de la loi  $N(0,1)$  avec un niveau de confiance de  $\alpha$ , on peut accepter l'hypothèse  $H_0$  et considérer que les erreurs sont centrées.

## 2. TESTS D'AUTO CORRELATION DES RESIDUS

Pour tester l'auto corrélation des erreurs, on utilise l'auto corrélation empirique des résidus,

$$\hat{\rho}(h) = \frac{\hat{\sigma}^2(h)}{s^2}$$

où

$$\hat{\sigma}^2(h) = \frac{1}{T} \sum_{i=1}^{T-h} (e_i - \bar{e})(e_{i+h} - \bar{e}) \text{ et } s^2 = \frac{1}{T} \sum_{i=1}^T (e_i - \bar{e})^2.$$

### 2.1. Tests individuels de bruit blanc

Il s'agit ici de tester l'auto corrélation pour chaque retard  $h > 0$ . Sous l'hypothèse  $H_0$  que le processus  $\{\varepsilon_t\}$  est un bruit blanc centré alors la statistique

$$Z_h = \sqrt{T} \hat{\rho}(h)$$

suit approximativement une loi normale  $N(0,1)$  pourvu que  $T$  soit suffisamment grand.

Ainsi si la variable  $Z_h$  prend une valeur supérieure (en valeur absolue) à celle donnée par la table de la loi  $N(0,1)$  avec un niveau de confiance  $\alpha$ , on peut rejeter  $H_0$ , c'est-à-dire l'hypothèse de bruit blanc.

Pour tester un bruit blanc, il convient de tester plusieurs valeurs de retard  $h$ . Cette technique présente un inconvénient majeur concernant les probabilités d'acceptation ou de rejeter de l'hypothèse de bruit blanc. En effet, pour chaque test, il y a une probabilité de  $1-\alpha$  d'accepter à raison l'hypothèse d'une absence d'auto corrélation de retard  $h$ . Les auto corrélation étant indépendantes, on en déduit une probabilité  $1-(1-\alpha)^H$  de rejeter l'hypothèse de bruit blanc alors qu'elle est exacte si on teste les  $H$  premières auto corrélation. Par exemple, si on prend  $\alpha=5\%$ , il y a 5% de chance de rejeter à tort l'absence d'auto corrélation pour chaque test. Si on teste  $H=20$  retard, il y a alors 64% de chance de rejeter à tort l'hypothèse de bruit blanc. Pour pallier ce problème, à un seuil de 5%, on rejettera l'hypothèse de bruit blanc si plus de 5% des tests individuels ont rejeté l'hypothèse. Par exemple si on prend  $\alpha=5\%$  et  $H=20$ , on rejettera l'hypothèse de bruit blanc si au moins  $20 \times 0.05 = 1$  test individuel aboutit au rejet.

### 2.2. Test global de bruit blanc

Afin de ne pas effectuer des tests sur plusieurs retard, il est possible d'effectuer un test global (test du Portemanteau). Sous l'hypothèse  $H_0$  que le processus  $\{\varepsilon_t\}$  est un bruit blanc alors la statistique

$$Z = T \sum_{h=1}^H \hat{\rho}^2(h)$$

suit approximativement une loi du  $\chi^2$  à H degrés de liberté pourvu que T soit suffisamment grand.

Ainsi si la variable Z prend une valeur supérieure à celle donnée par la table de la loi du  $\chi^2$  avec un niveau de confiance  $\alpha$ , on peut rejeter  $H_0$ , c'est-à-dire l'hypothèse de bruit blanc.

Cependant, ce test manque de puissance et amène à accepter un bruit blanc alors qu'il peut y avoir une auto corrélation important pour l'un des retards.

### Méthodologie

Afin de pallier aux inconvénients des deux tests précédents, il est recommandé d'opérer de la façon suivante :

- Effectuer les tests individuels pour tous les retard de 1 à H (H pas trop grand devant T) en s'intéressant plus particulièrement aux retards suspects (h=1,2,4,et 12).
- Effecteur le test global et considérer en même temps le nombre de rejets des tests individuels.

## 3. QUELQUES COMMANDES SAS

Cours de Aragon Y. :

[http://w3.univ-tlse1.fr/GREMAQ/Statistique/Yvesweb/docs/IUP\\_st\\_cours.pdf](http://w3.univ-tlse1.fr/GREMAQ/Statistique/Yvesweb/docs/IUP_st_cours.pdf)

---

```
* Syntaxe pour la simulation du Bruit blanc gaussien de variance 1.5**2 . ;

data a;
do i =1 to 200;
z = 1.5* rannor(45297);
output;
end;
run;

* calcul de la FAC de la série précédente ;

proc arima data= a;
i var = z; run;
quit;
```

---



# CHAPITRE 6 : MODELE ARMA

Dans ce chapitre nous allons introduire des modèles permettant d'exprimer  $X_t$  en fonction des valeurs antérieures du processus. Ces modèles serviront ensuite à prévoir des valeurs futures de la série.

## 1. PROCESSUS AUTOREGRESSIF (AR)

Les processus autorégressifs forment une famille simple de modèles dans lesquels  $X_t$  s'exprime comme combinaison linéaire des valeurs antérieures. Dans un premier temps nous allons étudier le cas où on ne prend en compte que le temps précédent  $t$ .

### 1.1. Processus autorégressif d'ordre 1 – AR(1)

#### Définition

Soit  $\{\varepsilon_t\}$  un bruit blanc centré de variance  $\sigma^2$  et  $a$  une constante de l'intervalle  $]-1,1[$ . Un *processus autorégressif d'ordre 1* est un processus défini par

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

#### Propriété

Un processus autorégressif est stationnaire centré dont les fonctions d'auto-covariance et d'auto-corrélation sont

$$\sigma^2(h) = \sigma^2 a^h \frac{1}{1-a^2} \quad \text{et} \quad \rho(h) = a^h$$

(elles tendent vers zéro à vitesse exponentielle), et la fonction d'auto-corrélation partielle est nulle pour tout retard supérieur ou égal à 2,

$$r(h) = 0, \quad h \geq 2$$

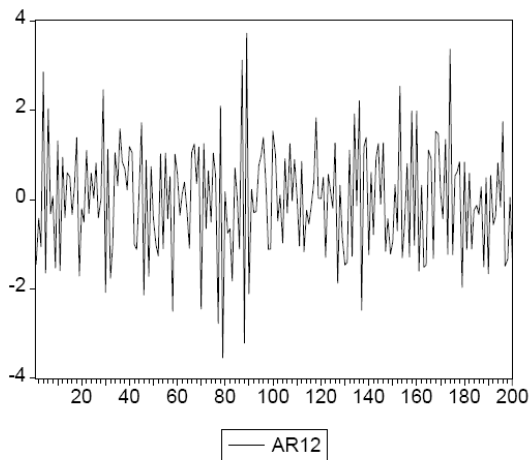
On remarque que par récurrence on peut écrire

$$X_t = \varepsilon_t + a\varepsilon_{t-1} + a^2\varepsilon_{t-2} + \dots = \sum_{i=0}^{\infty} a^i \varepsilon_{t-i}$$

On suppose que cette série converge et que les indices du bruit blanc peuvent être négatifs. La variable  $X_t$  peut donc s'écrire en fonction des  $\varepsilon_s$  où le temps  $s$  est égal ou antérieur à  $t$ .

On remarque alors que  $\varepsilon_{t+h}$  est orthogonal à  $X_t$  pour tout  $h > 0$ . Il représente donc ce que  $X_{t+1}$  apporte de nouveau par rapport à ce qui a été observé jusqu'à l'instant  $t$ . On parle alors de *bruit d'innovation*.

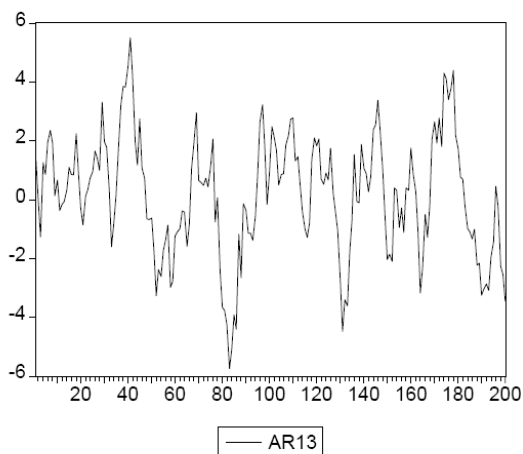
Exemple (Charpentier A. p. 98)  
Processus AR(1) avec a=-0,5



Sample: 1 1500  
Included observations: 1500

	Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	-0.510	-0.510	390.47	0.000		
2	0.252	-0.011	485.86	0.000		
3	-0.138	-0.018	514.48	0.000		
4	0.079	0.004	523.77	0.000		
5	-0.060	-0.022	529.21	0.000		
6	0.020	-0.028	529.84	0.000		
7	-0.011	-0.009	530.01	0.000		
8	0.017	0.014	530.46	0.000		
9	0.002	0.022	530.46	0.000		
10	0.004	0.014	530.49	0.000		
11	0.018	0.031	531.00	0.000		
12	-0.026	-0.007	531.99	0.000		
13	-0.028	-0.065	533.17	0.000		
14	-0.009	-0.058	533.28	0.000		
15	0.003	-0.019	533.30	0.000		
16	0.009	0.012	533.42	0.000		
17	0.013	0.029	533.66	0.000		
18	-0.020	-0.011	534.25	0.000		
19	0.017	-0.005	534.70	0.000		
20	-0.030	-0.026	536.05	0.000		
21	0.029	0.007	537.34	0.000		
22	-0.032	-0.007	538.89	0.000		

Processus AR(1) avec a=0.85



Sample: 1 1500  
Included observations: 1500

	Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.834	0.834	1044.4	0.000		
2	0.695	0.002	1771.9	0.000		
3	0.581	0.002	2279.8	0.000		
4	0.484	-0.005	2632.1	0.000		
5	0.400	-0.010	2872.9	0.000		
6	0.338	0.025	3045.1	0.000		
7	0.289	0.012	3171.4	0.000		
8	0.246	-0.005	3262.7	0.000		
9	0.202	-0.024	3324.2	0.000		
10	0.156	-0.030	3361.3	0.000		
11	0.109	-0.039	3379.2	0.000		
12	0.058	-0.048	3384.2	0.000		
13	0.026	0.017	3385.3	0.000		
14	0.020	0.055	3385.9	0.000		
15	0.025	0.033	3386.9	0.000		
16	0.029	0.002	3388.2	0.000		
17	0.024	-0.024	3389.1	0.000		
18	0.014	-0.020	3389.4	0.000		
19	0.009	0.015	3389.5	0.000		
20	0.007	0.010	3389.6	0.000		
21	0.012	0.024	3389.8	0.000		
22	0.014	-0.008	3390.1	0.000		

Démonstration

Par linéarité de l'espérance et étant donné que le bruit blanc est centré, on a  $E(X_t)=0$ .

Le produit scalaire est égal à  $E(X_t \varepsilon_{t+h})$  or les  $\varepsilon_t$  sont indépendants et centrés donc l'espérance est nulle, d'où l'orthogonalité.

Les  $\varepsilon_t$  étant indépendants, on en déduit que

$$\begin{aligned} \text{var}(X_t) &= \text{var}(\varepsilon_t) + a^2 \text{var}(\varepsilon_{t-1}) + a^4 \text{var}(\varepsilon_{t-2}) + \dots \\ &= \sigma^2 (1 + a^2 + a^4 + \dots) = \sigma^2 \sum_{i=0}^{\infty} a^{2i} = \sigma^2 \frac{1}{1-a^2} \end{aligned}$$

L'expression ci-dessus n'existe que si  $a \in ]-1,1[$ . Dans le cas contraire la variance est infinie.

Puisque le processus est centré, la covariance est l'espérance du produit,

$$\sigma^2(h) = \text{cov}(X_t, X_{t+h}) = E(X_t X_{t+h})$$

Or

$$X_t X_{t+h} = \sum_{i=0}^{\infty} a^i \varepsilon_{t-i} \times \sum_{j=0}^{\infty} a^j \varepsilon_{t+h-j} = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a^{i+j} \varepsilon_{t-i} \varepsilon_{t+h-j}$$

Les  $\varepsilon_k$  étant centrés et non corrélés, l'espérance du produit,  $E(\varepsilon_{t-i} \varepsilon_{t+h-j})$ , est nulle sauf quand  $t-i=t+h-j$  et dans ce cas elle est égale à la variance  $\sigma^2$ . On rappelle que la fonction d'auto-covariance est paire,  $\sigma^2(-h)=\sigma^2(h)$  donc  $t-i=t+h-j \Leftrightarrow i+j=2j-h \Leftrightarrow i+j=2j+h$ , d'où

$$\sigma^2(h) = \sigma^2 \sum_{j=0}^{\infty} a^{2j+h} = \sigma^2 a^h \frac{1}{1-a^2} \quad \text{et} \quad \rho(h) = a^h$$

Donc la covariance (ou auto-corrélation) est bien une fonction de  $h$  qui tend vers 0 à vitesse exponentielle quand  $h$  tend vers l'infini.

La fonction d'auto-corrélation partielle est définie par

$$\forall h \geq 2, r(h) = \text{corr}(\hat{X}_1, \hat{X}_{h+1}), \text{ où } \hat{X}_{h+1} = X_{h+1} - \text{EL}(X_{h+1} | X_2, \dots, X_h)$$

or

$$\text{EL}(X_{h+1} | X_2, \dots, X_h) = \text{EL}(aX_h + \varepsilon_{h+1} | X_2, \dots, X_h) = aX_h$$

car  $\varepsilon_{h+1}$  est orthogonal à  $X_2, \dots, X_h$ . D'où

$$\hat{X}_{h+1} = X_{h+1} - aX_h = \varepsilon_{h+1}.$$

Finalement,

$$r(h) = \text{corr}(\hat{X}_1, \varepsilon_{h+1}) = 0 \text{ par orthogonalité}$$

□

## 1.2. Processus autorégressif d'ordre $p$ – AR( $p$ )

Le modèle autorégressif d'ordre 1 est très simple et on se doute que l'évolution de  $X_t$  ne dépend pas toujours de sa valeur au temps précédent. Une généralisation consiste alors à introduire des temps supplémentaires.

### Définition

Soit  $\{\varepsilon_t\}$  un bruit blanc centré de variance  $\sigma^2$ . Un *processus autorégressif d'ordre  $p$*  est un processus défini par

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} + \varepsilon_t, \quad t \in \{p+1, \dots, T\}$$

tel que son *polynôme caractéristique*

$$p(z) = 1 - a_1 z - a_2 z^2 - \dots - a_p z^p$$

admet des racines de module strictement supérieure à 1.

### Exemple

Le processus défini par le polynôme caractéristique

$$p(z) = 1 - z + \frac{1}{2} z^2$$

est autorégressif puisque les racines de son polynôme caractéristique,  $z_1=1-i$  et  $z_2=1+i$  sont de module  $\sqrt{2}>1$ . Il est de la forme

$$X_t = X_{t-1} - \frac{1}{2} X_{t-2} + \varepsilon_t.$$

### Propriété

Un processus AR(p) défini comme précédemment est centré, stationnaire et les fonctions d'auto-covariance et d'auto-corrélation satisfont les équations de récurrence

$$\sigma(h) = a_1\sigma(h-1) + a_2\sigma(h-2) + \dots + a_p\sigma(h-p), \forall h > 0$$

$$\rho(h) = a_1\rho(h-1) + a_2\rho(h-2) + \dots + a_p\rho(h-p), \forall h > 0$$

et la fonction d'auto-corrélation partielle est nulle pour tout retard supérieur ou égal à p,

$$r(h) = 0, \forall h \geq p$$

### Exemple

Prenons l'exemple des modèles d'ordre 2,

$$X_t = a_1X_{t-1} + a_2X_{t-2} + \varepsilon_t.$$

Alors la fonction d'auto-corrélation vérifie l'équation de récurrence

$$\rho(h) = a_1\rho(h-1) + a_2\rho(h-2), \forall h > 0.$$

Cette équation définit une suite récurrente linéaire d'ordre 2. Elle est donc entièrement déterminée par les racines du polynôme caractéristique ainsi que par les deux premiers termes de la suite. Or on sait que  $\rho(0) = 1$ , il reste donc à calculer  $\rho(1)$ .

$$\sigma(1) = \text{cov}(X_t, X_{t-1}) = E(X_t X_{t-1}) = a_1 E(X_{t-1}^2) + a_2 E(X_{t-1} X_{t-2}) + E(X_{t-1} \varepsilon_t)$$

$$= a_1 E(X_{t-1}^2) + a_2 E(X_{t-1} X_{t-2}) \text{ car } \varepsilon_t \text{ est supposé indépendant du passé de } X_t$$

(cf. paragraphe suivant)

$$= a_1\sigma(0) + a_2\sigma(1)$$

$$\Leftrightarrow \sigma(1) = \frac{a_1\sigma^2}{1-a_2} \Leftrightarrow \rho(1) = \frac{\sigma(1)}{\sigma^2} = \frac{a_1}{1-a_2}.$$

Appliquons l'exemple précédent au modèle AR(2) suivant,

$$X_t = X_{t-1} - \frac{1}{2}X_{t-2} + \varepsilon_t.$$

Les racines du polynôme caractéristique sont complexes conjuguées telles que  $z_1 = 1 - i = \sqrt{2}e^{-i\pi/4}$ , donc la fonction d'auto-corrélation est de la forme

$$\rho(h) = (\sqrt{2})^h \left( K_1 \cos\left(h \frac{\pi}{4}\right) + K_2 \sin\left(h \frac{\pi}{4}\right) \right).$$

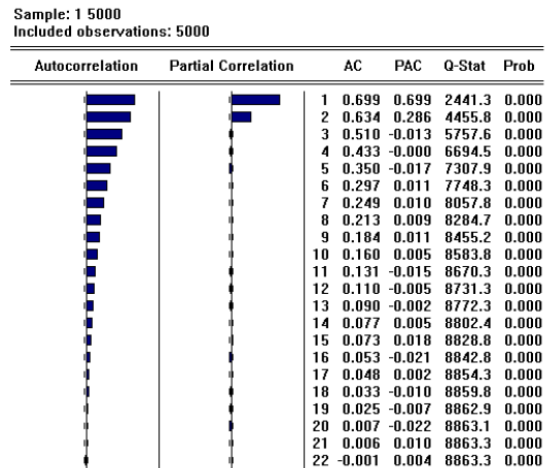
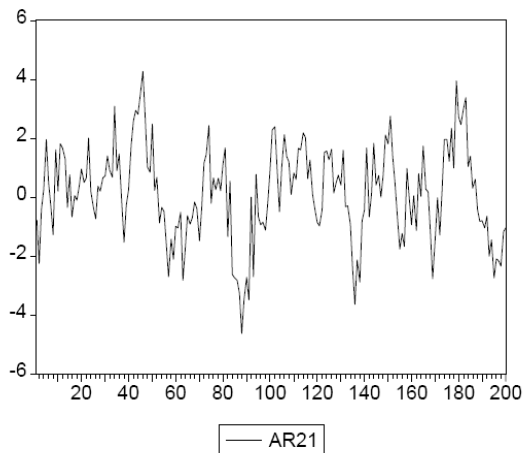
Or

$$\rho(0) = K_1 = 1 \text{ et } \rho(1) = K_1 + K_2 = \frac{1}{1 - (-\frac{1}{2})} \Rightarrow K_2 = -\frac{1}{3}$$

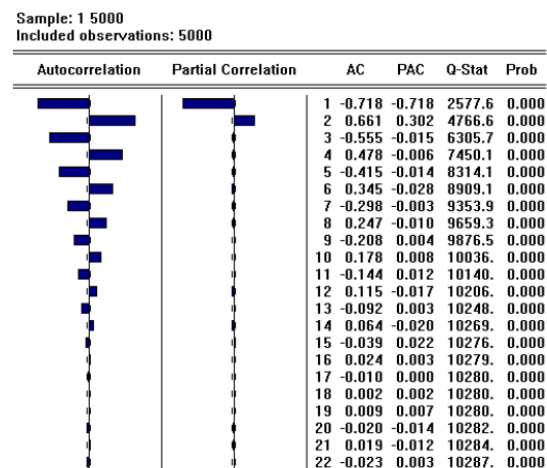
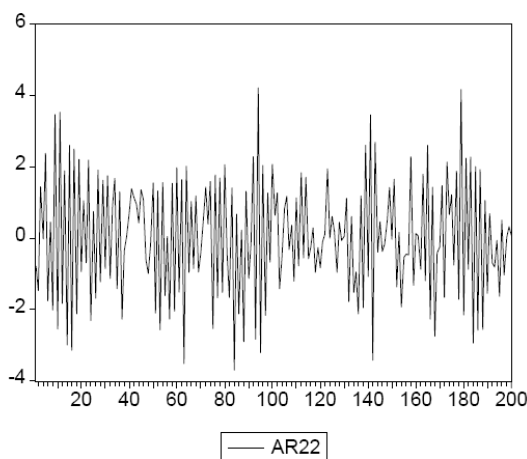
Finalement,

$$\rho(h) = (\sqrt{2})^h \left( \cos\left(h \frac{\pi}{4}\right) - \frac{1}{3} \sin\left(h \frac{\pi}{4}\right) \right).$$

Exemple (Charpentier A. p 100)  
Processus AR(2) avec  $a_1=0,5$  et  $a_2=0,3$



Processus AR(2) avec  $a_1=-0,5$  et  $a_2=0,3$



Ecriture sous la forme d'un MA( $\infty$ )

En utilisant un opérateur de retard, le modèle autorégressif d'ordre p peut s'écrire sous la forme

$$(1 - a_1B - a_2B^2 - \dots - a_pB^p)X_t = \varepsilon_t$$

Notons  $z_1, \dots, z_p$  les p racines du polynôme caractéristique, on a alors  $p(z)=(1-\alpha_1z)(1-\alpha_2z)\dots(1-\alpha_pz)$  où  $\alpha_n=z_n^{-1}$ . Chaque terme de la factorisation est inversible puisque  $|z_n|>1$  entraîne  $|\alpha_n|<1$  et  $(1-\alpha z)^{-1}=1+\alpha z+\alpha^2 z^2+\dots+\alpha^n z^n+\dots$ . Le polynôme est donc lui aussi inversible et  $p^{-1}(z)=1+\beta z+\beta^2 z^2+\dots+\beta^n z^n+\dots$ . On a alors

$$\begin{aligned} X_t &= (1 - a_1B - a_2B^2 - \dots - a_pB^p)^{-1}\varepsilon_t \\ &= (1 + \beta_1B + \beta_2B^2 + \dots + a_nB^n + \dots)\varepsilon_t \\ &= \varepsilon_t + \sum_{n \geq 1} \beta_n \varepsilon_{t-n} \end{aligned}$$

La variable  $X_t$  peut donc s'écrire en fonction des  $\varepsilon_s$  où le temps  $s$  est égal ou antérieur à  $t$ . Là encore on parle de bruit d'innovation car  $\varepsilon_{t+h}$  est orthogonal à  $X_t$  pour  $h > 0$ .

### 1.3 Quelques commandes SAS

Cours de Aragon Y. :

[http://w3.univ-tlse1.fr/GREMAQ/Statistique/Yvesweb/docs/IUP\\_st\\_cours.pdf](http://w3.univ-tlse1.fr/GREMAQ/Statistique/Yvesweb/docs/IUP_st_cours.pdf)

**Simulation d'un AR** On veut simuler 200 observations d'un processus stationnaire obéissant à :

$$y_t = 1.4 * y_{t-1} - 0.45 * y_{t-2} + a_t \quad (6.22)$$

où  $a_t$  est un bruit blanc gaussien de variance 1, c'est un AR(2) gaussien. L'étape data ci-dessous fait le travail.

---

```
title1 'AR(2) simule';
data a;
  um1 = 0; um2=0; /* valeurs initiales */
  do i = -50 to 200;
    a = rannor( 43201 );
    u = 1.4 * um1 - .45* um2 + a;
    if i > 0 then output;
    um2 = um1;
    um1 = u;
  end;
run;
```

---

Commentaires. Noter qu'il faut deux valeurs initiales  $y_1$  et  $y_2$ . Au début de la simulation le processus dépend de ces valeurs. C'est pourquoi on abandonne le début de la série, ici les 51 premières valeurs.

---

```
* simulation d'un AR(1)
data a;
c= 4;
phi= -.8;
y1=0;
  do i = -50 to 150;
    z = rannor( 32565 );
    y = c+ phi* y1 +z;
    y1=y;
    if i > 0 then output;
  end;
run;
proc arima data = a;
i var= y; run;
quit;
```

---

## 2. PROCESSUS MOYENNE MOBILE (MA)

### 2.1. Processus moyenne mobile d'ordre 1 – MA(1)

#### Définition

Soit  $\{\varepsilon_t\}$  un bruit blanc centré de variance  $\sigma^2$  et  $b$  une constante de l'intervalle  $] -1, 1[$ . Un *processus moyenne mobile d'ordre 1* est un processus défini par

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

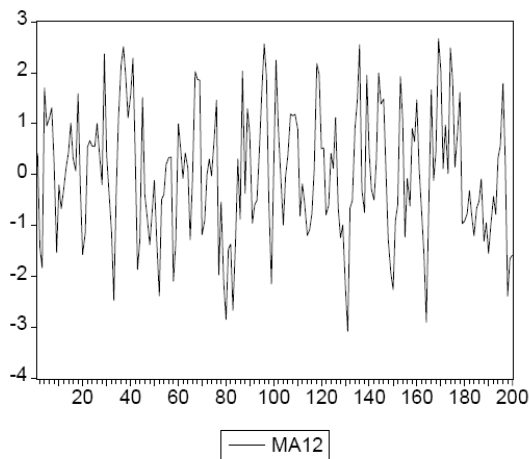
#### Propriété

Un processus moyenne mobile d'ordre 1 est stationnaire centré dont les fonctions d'auto-covariance et d'auto-corrélation sont

$$\sigma^2(h) = \begin{cases} \sigma^2 b & \text{si } h = 1 \\ 0 & \text{sin on} \end{cases} \quad \text{et} \quad \rho(h) = \begin{cases} \frac{b}{1+b^2} & \text{si } h = 1 \\ 0 & \text{sin on} \end{cases}$$

#### Exemple (Charpentier A. p 104)

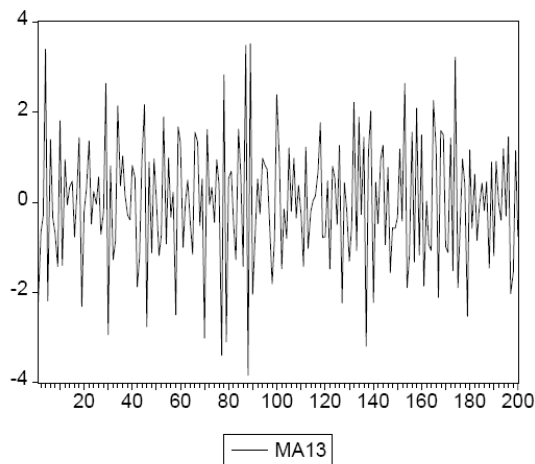
Processus MA(1) avec  $b=-0,5$



Sample: 1 1500  
Included observations: 1500

	Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.383	0.383	220.63	0.000		
2	-0.024	-0.200	221.46	0.000		
3	-0.013	0.090	221.72	0.000		
4	-0.017	-0.063	222.18	0.000		
5	-0.039	-0.008	224.44	0.000		
6	-0.023	-0.007	225.23	0.000		
7	0.011	0.020	225.41	0.000		
8	0.031	0.019	226.85	0.000		
9	0.033	0.018	228.50	0.000		
10	0.032	0.019	230.09	0.000		
11	0.005	-0.015	230.12	0.000		
12	-0.071	-0.075	237.85	0.000		
13	-0.103	-0.048	253.89	0.000		
14	-0.055	-0.006	258.56	0.000		
15	0.005	0.019	258.59	0.000		
16	0.035	0.023	260.49	0.000		
17	0.021	-0.008	261.16	0.000		
18	-0.009	-0.020	261.30	0.000		
19	-0.018	-0.009	261.81	0.000		
20	-0.018	-0.008	262.31	0.000		
21	-0.002	0.017	262.32	0.000		
22	-0.004	-0.007	262.34	0.000		

Processus MA(1) avec  $b=0,85$



Sample: 1 1500  
Included observations: 1500

	Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	-0.494	-0.494	366.71	0.000		
2	-0.001	-0.324	366.71	0.000		
3	-0.004	-0.242	366.73	0.000		
4	0.021	-0.160	367.38	0.000		
5	-0.027	-0.149	368.50	0.000		
6	0.000	-0.138	368.50	0.000		
7	0.003	-0.124	368.51	0.000		
8	0.010	-0.095	368.67	0.000		
9	-0.004	-0.078	368.70	0.000		
10	0.004	-0.060	368.73	0.000		
11	0.031	0.007	370.14	0.000		
12	-0.023	0.017	370.93	0.000		
13	-0.030	-0.028	372.32	0.000		
14	0.001	-0.059	372.32	0.000		
15	0.008	-0.067	372.42	0.000		
16	0.015	-0.041	372.78	0.000		
17	0.008	-0.008	372.87	0.000		
18	-0.017	-0.018	373.29	0.000		
19	0.011	-0.007	373.49	0.000		
20	-0.021	-0.035	374.14	0.000		
21	0.022	-0.014	374.85	0.000		
22	-0.017	-0.024	375.26	0.000		

Démonstration

Par linéarité de l'espérance et étant donné que le bruit blanc est centré, on a  $E(X_t)=0$ .

La fonction d'auto-covariance du bruit blanc étant nulle, on en déduit que

$$\text{var}(X_t) = \text{var}(\epsilon_t) + b^2 \text{var}(\epsilon_{t-1}) = \sigma^2(1 + b^2)$$

Puisque le processus est centré, la covariance est l'espérance du produit,

$$\begin{aligned} \sigma^2(h) &= \text{cov}(X_t, X_{t+h}) = E(X_t X_{t+h}) \\ &= E(\epsilon_t \epsilon_{t+h}) + bE(\epsilon_t \epsilon_{t+h-1}) + bE(\epsilon_{t-1} \epsilon_{t+h}) + b^2E(\epsilon_{t-1} \epsilon_{t+h-1}) \end{aligned}$$

Si  $h>1$  alors  $\sigma^2(h) = 0$  car les  $\epsilon_t$  sont indépendants. Si  $h=1$ ,  $\sigma(1) = bE(\epsilon_t^2) = b\sigma^2$ .

On en déduit immédiatement  $\rho(h)$ .  $\square$

De même que pour les processus AR, on aimerait pouvoir écrire le processus MA en fonction de son passé (observé) et pas seulement en fonction d'un bruit non observé. On remarque que le processus peut s'écrire à l'aide d'un opérateur de retard

$$X_t = (1 - bB)\epsilon_t.$$

Etant donné que  $b \in ]-1, 1[$ ,  $(1 - bB)^{-1}$  peut se développer en série entière,

$$(1 - bB)^{-1} = 1 + \sum_{i \geq 1} \alpha_i B^i.$$

D'où  $\epsilon_t = (1 - bB)^{-1} X_t \Rightarrow \epsilon_t = (1 + \sum_{i \geq 1} \alpha_i B^i) X_t \Rightarrow X_t = -\alpha_1 X_{t-1} - \alpha_2 X_{t-2} - \dots - \epsilon_t$ .

Le processus peut s'écrire comme une auto-régression infinie. On parle alors de processus *inversible*.

## 2.2. Processus moyenne mobile d'ordre q – MA(q)

Définition

Soit  $\{\epsilon_t\}$  un bruit blanc centré de variance  $\sigma^2$ . Un *processus moyenne mobile d'ordre q* est un processus défini par

$$X_t = \epsilon_t + b_1 \epsilon_{t-1} + \dots + b_q \epsilon_{t-q}.$$

Propriété

Un processus moyenne mobile d'ordre q est stationnaire centré dont les fonctions d'auto-covariance et d'auto-corrélation sont nulles à partir de l'ordre q+1.

Exemple

Prenons l'exemple des processus moyenne mobile d'ordre 2,

$$X_t = \epsilon_t + b_1 \epsilon_{t-1} + b_2 \epsilon_{t-2}.$$

Etant donné que les  $\epsilon_t$  sont indépendants, on a

$$\text{var}(X_t) = \text{var}(\epsilon_t) + b_1^2 \text{var}(\epsilon_{t-1}) + b_2^2 \text{var}(\epsilon_{t-2}) = \sigma^2(1 + b_1^2 + b_2^2).$$

On sait que la fonction d'auto-covariance est nulle pour  $h \geq 3$ .

$$\begin{aligned} \sigma(1) &= E(X_t X_{t-1}) = E[(\epsilon_t + b_1 \epsilon_{t-1} + b_2 \epsilon_{t-2})(\epsilon_{t-1} + b_1 \epsilon_{t-2} + b_2 \epsilon_{t-3})] \\ &= b_1 E[\epsilon_t^2] + b_1 b_2 E[\epsilon_{t-2}^2] = \sigma^2 b_2 (1 + b_1) \end{aligned}$$

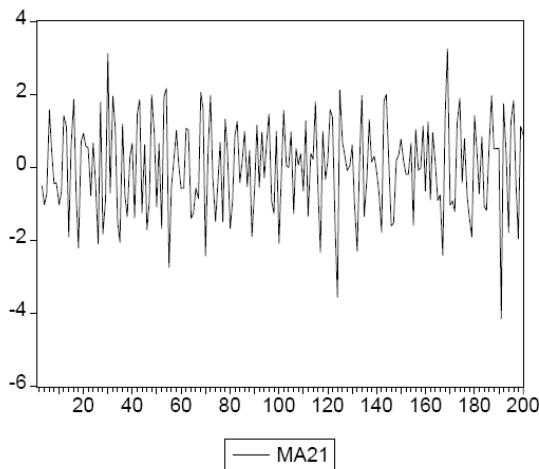
$$\Rightarrow \rho(1) = \frac{b_2(1 + b_1)}{1 + b_1^2 + b_2^2}$$

$$\begin{aligned} \sigma(2) &= E(X_t X_{t-2}) = E[(\epsilon_t + b_1 \epsilon_{t-1} + b_2 \epsilon_{t-2})(\epsilon_{t-2} + b_1 \epsilon_{t-3} + b_2 \epsilon_{t-4})] \\ &= b_2 E[\epsilon_{t-2}^2] = \sigma^2 b_2 \end{aligned}$$

$$\Rightarrow \rho(2) = \frac{b_2}{1 + b_1^2 + b_2^2}$$

**Exemple** (Charpentier A. p 105)

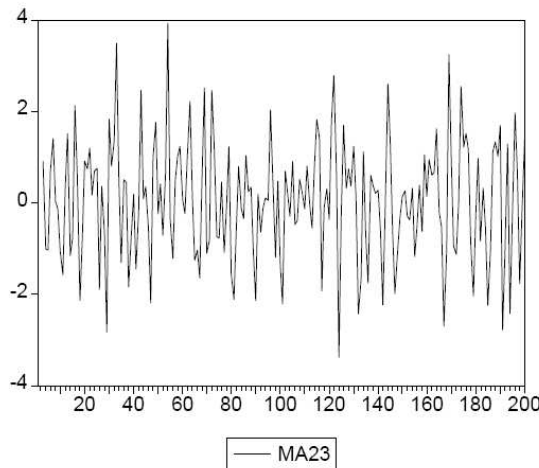
Processus MA(2) avec  $b_1=0,5$  et  $b_2=0,3$



Sample: 1 5000  
Included observations: 4998

	Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	-0.160	-0.160	128.18	0.000		
2	-0.316	-0.350	626.69	0.000		
3	-0.012	-0.160	627.43	0.000		
4	-0.022	-0.207	629.84	0.000		
5	0.006	-0.135	630.03	0.000		
6	0.007	-0.134	630.30	0.000		
7	-0.016	-0.129	631.63	0.000		
8	-0.002	-0.118	631.64	0.000		
9	0.021	-0.089	633.92	0.000		
10	0.020	-0.061	636.03	0.000		
11	-0.023	-0.080	638.76	0.000		
12	-0.007	-0.063	639.01	0.000		
13	0.016	-0.046	640.37	0.000		
14	-0.001	-0.043	640.37	0.000		
15	-0.017	-0.055	641.88	0.000		
16	-0.017	-0.070	643.37	0.000		
17	-0.006	-0.081	643.58	0.000		
18	0.034	-0.046	649.29	0.000		
19	0.016	-0.041	650.57	0.000		
20	-0.005	-0.030	650.67	0.000		
21	-0.022	-0.050	653.15	0.000		
22	-0.023	-0.071	655.92	0.000		

Processus MA(2) avec  $b_1=-0,5$  et  $b_2=-0,7$



Sample: 1 5000  
Included observations: 4998

	Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.186	0.186	173.38	0.000		
2	-0.306	-0.353	641.17	0.000		
3	-0.008	0.159	641.48	0.000		
4	-0.022	-0.205	643.82	0.000		
5	0.004	0.145	643.90	0.000		
6	0.006	-0.141	644.08	0.000		
7	-0.011	0.106	644.66	0.000		
8	0.003	-0.097	644.70	0.000		
9	0.032	0.122	649.86	0.000		
10	0.023	-0.077	652.44	0.000		
11	-0.017	0.070	653.96	0.000		
12	-0.011	-0.065	654.62	0.000		
13	0.010	0.069	655.14	0.000		
14	-0.009	-0.076	655.56	0.000		
15	-0.028	0.036	659.37	0.000		
16	-0.028	-0.084	663.39	0.000		
17	-0.003	0.055	663.43	0.000		
18	0.035	-0.031	669.48	0.000		
19	0.025	0.056	672.50	0.000		
20	-0.005	-0.041	672.60	0.000		
21	-0.029	0.013	676.74	0.000		
22	-0.020	-0.036	678.69	0.000		

Ecriture d'un MA sous la forme d'un AR( $\infty$ )

De même que pour l'ordre 1, le processus peut s'écrire avec un opérateur de retard,

$$X_t = (1 + b_1 B + \dots + b_q B^q) \epsilon_t.$$

On montre alors que le processus est inversible si les racines du polynôme  $1 + b_1z + \dots + b_qz^q$  sont toutes de module strictement supérieur à 1 et on peut le développer en série entière

$$(1 + b_1z + \dots + b_qz^q)^{-1} = 1 + \sum_{n \geq 1} \varepsilon_n z^n$$

D'où

$$(1 + b_1B + \dots + b_qB^q)^{-1}X_t = \varepsilon_t \Leftrightarrow X_t = -\sum_{n \geq 1} \alpha_n X_{t-n} + \varepsilon_t.$$

On remarque que techniquement la condition d'inversibilité d'un MA est identique à la condition de stationnarité d'un AR.

### 2.3. Quelques commandes SAS

Cours de Aragon Y.

[http://w3.univ-tlse1.fr/GREMAQ/Statistique/Yvesweb/docs/IUP\\_st\\_cours.pdf](http://w3.univ-tlse1.fr/GREMAQ/Statistique/Yvesweb/docs/IUP_st_cours.pdf)

**Simulation d'un MA** Un processus MA( $q$ ) est une combinaison linéaire de  $q$  v.a. non corrélées de moyenne 0, de variance constante. On peut simuler dans SAS des v.a. i.i.d.  $N(0,1)$ , par la fonction `rannor`. Elle génère des nombres pseudo aléatoires à partir d'un germe (ou graine) qu'on doit spécifier. Si on donne, à des emplois successifs de la fonction, le même germe constamment, on obtient toujours la même suite de simulations.

Exemple : Simulation d'une série de 100 observations suivant :

$$x_t = a_t - 0.8a_{t-1}$$

avec  $a_t$  v.a. i.i.d.  $N(0,1)$  ( $x_t$  MA(1) gaussien).

---

```
title1 'Serie MA(1) Simulee';
data a;
  a1 = 0;
  do i = -50 to 150;
    a = rannor( 32565 );
    u = a - .8 * a1;
    if i > 0 then output;
    a1 = a;
    u1 = u;
  end;
run;
```

---

On a initialisé la série du BB à 0 : `a1 = 0;`, puis on fait 150 tirages et on ne conserve que les 100 derniers (`if i > 0 then output;`).

Simulation du MA(2) (6.17)

---

```

data a;
mu =3;
zm1 =0;
zm2 =0;
do i=-10 to 200;
    zm0 = 1.5*rannor(54629);
    y = mu + 0.3*zm1-0.9*zm2 + zm0;
    zm2 = zm1;
    zm1=zm0;
    if i >0 then output;
end;
run;
*-----;
proc arima data=a;
i var= y;
run;
quit;

```

---

### 3. PROCESSUS ARMA

Ces processus se représentent sous une forme auto-régressive à second membre de type moyenne mobile. Ils généralisent ainsi les processus AR et MA vus précédemment.

#### Définition

Soit  $\{\varepsilon_t\}$  un bruit blanc centré de variance  $\sigma^2$ . Un *processus ARMA(p,q)* est un processus stationnaire défini par

$$X_t - a_1X_{t-1} - \dots - a_pX_{t-p} = \varepsilon_t + b_1\varepsilon_{t-1} + \dots + b_q\varepsilon_{t-q}$$

tel que les polynômes caractéristiques  $A(z) = 1 - a_1z - a_2z^2 - \dots - a_pz^p$  et  $B(z) = 1 + b_1z + \dots + b_qz^q$  ont des racines de module strictement supérieur à 1 et n'ont pas de racine commune.

On remarque que si les  $b_j$  sont nuls alors on retrouve les processus AR, et si les  $a_i$  sont nuls, on a un processus MA. Afin de distinguer les processus ARMA de ces deux cas extrêmes, on suppose que  $pq \neq 0$ .

#### Propriété

Un processus ARMA(p,q) est centré stationnaire dont les fonctions d'auto-covariance et d'auto-corrélation satisfont les équations de récurrence (équations de Yule-Walker)

$$\sigma(h) = a_1\sigma(h-1) + a_2\sigma(h-2) + \dots + a_p\sigma(h-p), \quad \forall h > q$$

$$\rho(h) = a_1\rho(h-1) + a_2\rho(h-2) + \dots + a_p\rho(h-p), \quad \forall h > q$$

et la fonction d'auto-corrélation partielle est nulle à partir de l'ordre  $p+1$ ,

$$r(h) = 0, \quad \forall h \geq p+1.$$

#### Exemple

Prenons le cas le plus simple d'un processus ARMA(1,1)

$$X_t - aX_{t-1} = \varepsilon_t + b\varepsilon_{t-1},$$

tel que  $0 < |a| < 1$  et  $0 < |b| < 1$ . Si on l'écrit avec un opérateur de retard, on a

$$(1 - aB)X_t = (1 + bB)\varepsilon_t.$$

$(1-aB)^{-1}$  peut se développer en série puisque  $|a| < 1$ , d'où

$$\begin{aligned} X_t &= (1 - aB)^{-1}(1 + bB)\varepsilon_t \\ &= (1 + aB + a^2B^2 + a^3B^3 + \dots)(1 + bB)\varepsilon_t \\ &= \varepsilon_t + (a + b)\sum_{n \geq 1} a^{n-1}\varepsilon_{t-n} . \end{aligned}$$

Etant donné que  $\{\varepsilon_t\}$  est un bruit blanc centré, on en déduit que  $X_t$  est centré, puis

$$\begin{aligned} \sigma(0) &= \text{var}(X_t) = \text{var}(\varepsilon_t) + (a + b)\sum_{n \geq 1} a^{n-1} \text{var}(\varepsilon_{t-n}) \\ &= \sigma^2 + (a + b)\sigma^2 \sum_{n \geq 1} a^{n-1} = \sigma^2 + (a + b)\sigma^2 \frac{1}{1 - a} = \sigma^2 \frac{1 + b}{1 - a} \end{aligned}$$

$$\begin{aligned} \sigma(1) &= E(X_t X_{t-1}) = E(aX_{t-1}^2 + \varepsilon_t X_{t-1} + b\varepsilon_{t-1} X_{t-1}) \\ &= a \text{var}(X_{t-1}) + b \text{var}(\varepsilon_{t-1}) \text{ car } X_{t-1} = \varepsilon_{t-1} + (a + b)\sum_{n \geq 1} a^{n-1}\varepsilon_{t-1-n} \\ &= a\sigma^2 \frac{1 + b}{1 - a} + b\sigma^2 = \sigma^2 \frac{a + b}{1 - a} \end{aligned}$$

$$\Rightarrow \rho(1) = \frac{1 + b}{a + b}$$

Ensuite la relation de récurrence permet de dire que

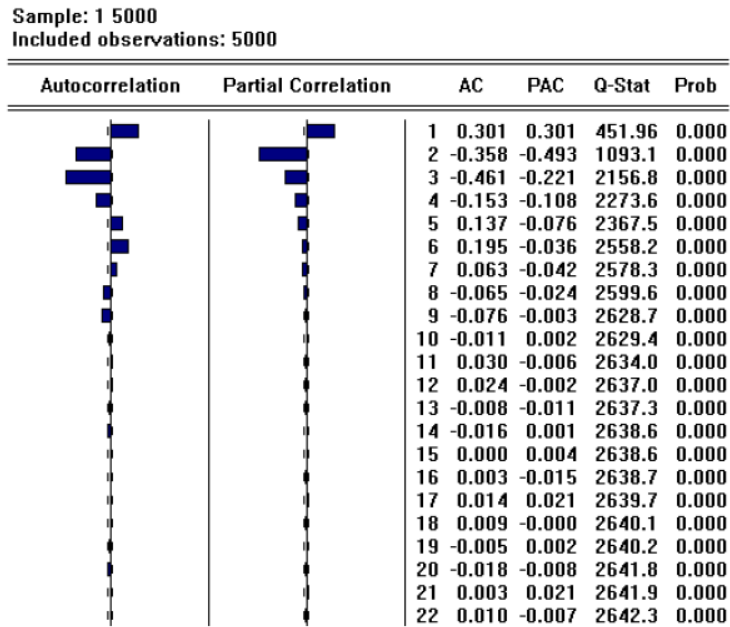
$$\forall h \geq 2, \sigma(h) = a^{h-1}\sigma(1) \text{ et } \rho(h) = a^{h-1}\rho(1).$$

Exemple (Charpentier A. p 108)

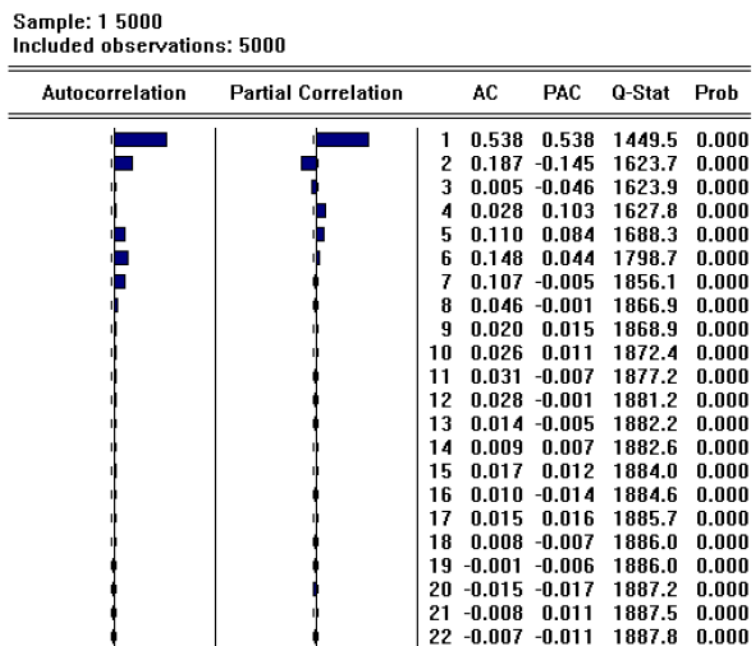
Sample: 1 5000  
Included observations: 5000

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.567	0.567	1607.5	0.000
		2	0.605	0.418	3439.3	0.000
		3	0.470	0.059	4542.8	0.000
		4	0.369	-0.088	5224.2	0.000
		5	0.274	-0.083	5600.6	0.000
		6	0.218	0.000	5839.6	0.000
		7	0.168	0.029	5981.6	0.000
		8	0.134	0.018	6071.3	0.000
		9	0.109	0.007	6131.0	0.000
		10	0.091	0.003	6172.7	0.000
		11	0.067	-0.015	6195.5	0.000
		12	0.056	-0.005	6211.2	0.000
		13	0.041	-0.002	6219.6	0.000
		14	0.032	0.002	6224.8	0.000
		15	0.036	0.024	6231.2	0.000
		16	0.015	-0.018	6232.4	0.000
		17	0.019	-0.005	6234.3	0.000
		18	0.006	-0.011	6234.4	0.000
		19	0.004	-0.001	6234.5	0.000
		20	-0.015	-0.023	6235.6	0.000
		21	-0.010	0.003	6236.1	0.000
		22	-0.014	0.010	6237.2	0.000

ARMA(2,1)



ARMA(1,2)



ARMA(2,2)

Écriture sous la forme d'un AR( $\infty$ ) ou d'un MA( $\infty$ )

En utilisant un opérateur de retard, un processus ARMA(p,q) peut s'écrire

$$(1 - a_1B - \dots - a_pB^p)X_t = (1 + b_1B + \dots + b_qB^q)\varepsilon_t.$$

Etant donné que les racines du polynôme caractéristique A sont de module strictement supérieure à 1, celui-ci est inversible et peut se décomposer en série entière, d'où

$$\begin{aligned} X_t &= (1 - a_1 B - \dots - a_p B^p)^{-1} (1 + b_1 B + \dots + b_q B^q) \varepsilon_t \\ &= (1 + \beta_1 B + \dots + \beta_n B^n + \dots) \varepsilon_t \\ &= \varepsilon_t + \sum_{n \geq 1} \beta_n \varepsilon_{t-n} \end{aligned}$$

On obtient ainsi l'écriture du processus sous la forme d'un MA( $\infty$ ). Comme nous l'avons vu dans le cas ARMA(1,1), cette écriture permet d'établir la stationnarité du processus et les formules de récurrence des fonctions d'auto-covariance et d'auto-corrélation.

Etant donné que les racines du polynôme caractéristique B sont de module strictement supérieur à 1, celui-ci est inversible et peut se décomposer en série entière, d'où

$$\begin{aligned} \varepsilon_t &= (1 - a_1 B - \dots - a_p B^p) (1 + b_1 B + \dots + b_q B^q)^{-1} X_t \\ &= (1 + \alpha_1 B + \dots + \alpha_n B^n + \dots) X_t \\ \Leftrightarrow X_t &= \sum_{n \geq 1} \alpha_n X_{t-n} + \varepsilon_t \end{aligned}$$

On obtient ainsi l'écriture du processus sous la forme d'un AR( $\infty$ ). Celle-ci servira dans le chapitre sur la prévision d'un modèle ARMA.

#### Simulation d'un ARMA sous SAS

[http://support.sas.com/documentation/cdl/en/imlug/59656/HTML/default/genstate\\_xpls\\_sect14.htm](http://support.sas.com/documentation/cdl/en/imlug/59656/HTML/default/genstate_xpls_sect14.htm)

## 4. IDENTIFICATION D'UN PROCESSUS ARMA

Identifier un modèle ARMA consiste à

- choisir les paramètres p et q du processus ARMA(p,q)
- Estimer les coefficients  $a_1, \dots, a_p$  du processus AR,  $b_1, \dots, b_q$  du processus MA et la variance du bruit d'innovation  $\sigma^2$ .
- Valider le modèle

Nous supposons que le processus est stationnaire. Si cela n'est pas le cas, il faut alors le différencier jusqu'à le rendre stationnaire.

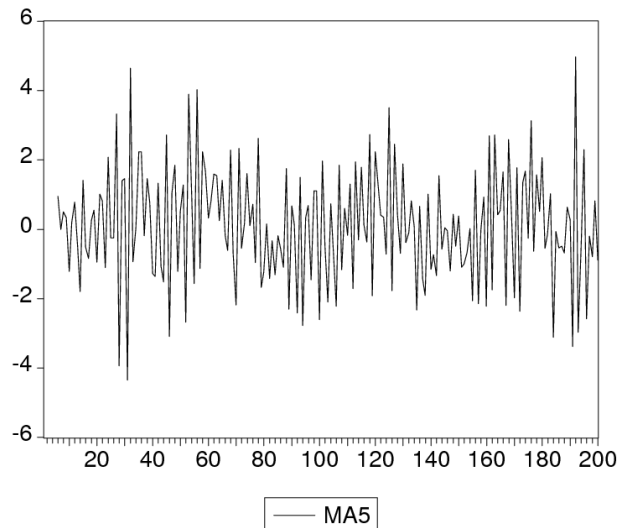
### 4.1. Choix des paramètres p et q d'un modèle ARMA

#### Identification de q

Nous avons vu précédemment qu'un processus MA(q) est stationnaire centré dont

les fonctions d'auto-covariance et d'auto-corrélation sont nulles à partir de l'ordre q+1.

Cette dernière propriété est utile pour deviner l'ordre de la moyenne mobile convenable pour modéliser une série. En effet, en présence d'un corrélogramme empirique non significativement différent de 0 à partir d'un certain ordre  $k$ , on pensera à modéliser la série correspondante par un  $MA(k-1)$ . Prenons l'exemple ci-dessous (Charpentier A.) d'un processus  $MA(5)$  simulé. On constate que la fonction d'auto-corrélation est significativement nulle à partir de 6.



Exemple simulé d'un  $MA(5)$  – Charpentier A.

Sample: 1 5000  
Included observations: 4995

	Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	-0.394	-0.394	776.33	0.000		
2	0.277	0.144	1159.8	0.000		
3	0.348	0.602	1764.0	0.000		
4	-0.207	0.154	1979.2	0.000		
5	0.268	-0.127	2338.3	0.000		
6	0.005	-0.203	2338.4	0.000		
7	0.005	0.006	2338.5	0.000		
8	0.012	-0.090	2339.2	0.000		
9	-0.005	-0.106	2339.3	0.000		
10	0.015	0.046	2340.4	0.000		
11	-0.012	0.211	2341.1	0.000		
12	-0.002	0.079	2341.1	0.000		
13	0.007	-0.081	2341.4	0.000		
14	-0.019	-0.105	2343.1	0.000		
15	-0.001	-0.033	2343.1	0.000		
16	-0.001	-0.027	2343.1	0.000		
17	-0.024	-0.045	2346.0	0.000		
18	0.021	0.065	2348.2	0.000		
19	-0.017	0.133	2349.5	0.000		
20	0.001	0.055	2349.5	0.000		
21	0.011	-0.063	2350.2	0.000		
22	-0.021	-0.098	2352.4	0.000		

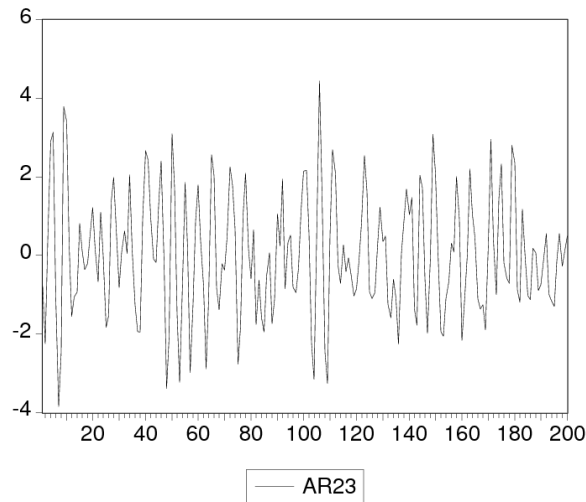
### Identification de $p$

Nous avons vu précédemment que

la fonction d'auto-corrélation partielle d'un  
processus  $AR(p)$  est nulle pour  $h > p$

Ainsi, en présence d'un corrélogramme partiel empirique non significativement différent de 0 à partir d'un certain ordre  $k$ , on pensera à modéliser la série correspondante par un  $AR(k-1)$ . Prenons l'exemple ci-dessous (Charpentier A.)

d'un processus AR(2) simulé. On constate que la fonction d'auto-corrélation partielle est significativement nulle à partir de 3.



Exemple simulé d'un AR(2) – Charpentier A

Sample: 1 5000  
Included observations: 5000

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.294	0.294	431.72	0.000
		2	-0.556	-0.703	1976.5	0.000
		3	-0.483	0.004	3142.1	0.000
		4	0.148	-0.003	3252.4	0.000
		5	0.405	-0.015	4072.7	0.000
		6	0.091	-0.006	4114.4	0.000
		7	-0.252	-0.034	4431.8	0.000
		8	-0.198	-0.013	4627.6	0.000
		9	0.084	0.001	4662.8	0.000
		10	0.192	0.004	4846.7	0.000
		11	0.040	0.003	4854.6	0.000
		12	-0.118	0.004	4925.0	0.000
		13	-0.095	-0.011	4970.6	0.000
		14	0.034	-0.000	4976.4	0.000
		15	0.067	-0.005	5014.1	0.000
		16	0.020	-0.003	5016.1	0.000
		17	-0.040	0.027	5024.2	0.000
		18	-0.033	-0.010	5029.8	0.000
		19	0.004	-0.002	5029.9	0.000
		20	0.021	-0.001	5032.0	0.000
		21	0.020	0.018	5033.9	0.000
		22	-0.006	-0.022	5034.1	0.000

## 4.2. Estimation des paramètres d'un modèle ARMA

Maintenant que les paramètres  $p$  et  $q$  sont choisis, il reste à estimer les coefficients  $a_1, \dots, a_p$  du processus AR,  $b_1, \dots, b_q$  du processus MA et la variance du bruit d'innovation  $\sigma^2$ . Nous ne détaillerons pas les méthodes de calcul des coefficients. Nous utiliserons directement les résultats fournis par le logiciel. On précise seulement que les méthodes d'estimation supposent que les erreurs sont des variables aléatoires indépendantes de loi normale.

Dans l'exemple ci-dessous, nous considérons un modèle ARMA(1,1). On obtient le modèle

$$X_t = 0.767X_{t-1} + \varepsilon_t - 0.463\varepsilon_{t-1}$$

LS // Dependent Variable is X  
Sample: 2 5000  
Included observations: 4999 after adjusting endpoints  
Convergence achieved after 4 iterations

Variable	Coefficient	Std. Error	T-Statistic	Prob.
AR(1)	0.767134	0.019055	40.25967	0.0000
MA(1)	-0.463423	0.026398	-17.55536	0.0000
R-squared	0.183040	Mean dependent var	0.014388	
Adjusted R-squared	0.182877	S.D. dependent var	1.115828	
S.E. of regression	1.008651	Akaike info criterion	0.017628	
Sum squared resid	5083.836	Schwartz criterion	0.020236	
Log likelihood	-7135.336	F-statistic	1119.579	
Durbin-Watson stat	2.002189	Prob(F-statistic)	0.000000	
Inverted AR Roots	.77			
Inverted MA Roots	.46			

### 4.3. Choix du modèle ARMA

#### Critères du pouvoir prédictif

Dans un modèle ARMA, l'erreur de prévision à l'horizon 1 dépend de la variance du résidu. On peut alors choisir le modèle conduisant à la plus petite erreur de prévision. Plusieurs indicateurs sont possibles

- Minimiser la variance du résidu ou la somme des carrés des résidus
- Maximiser le coefficient de détermination  $R^2$ , ou sa version modifiée
- Maximiser la statistique de Fisher

#### Critères d'information

Les critères d'information permettent de mesurer l'écart entre le modèle proposé et la vraie loi (Akaike, Schwartz,...)

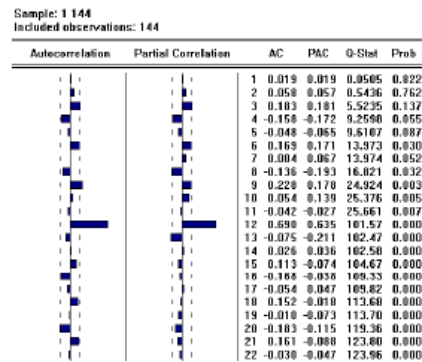
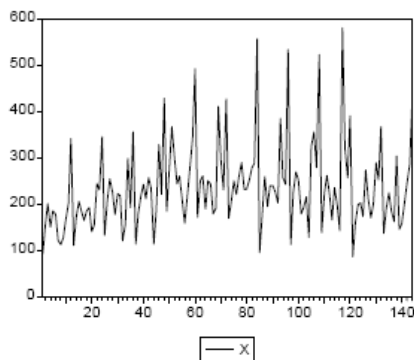
#### Validation des hypothèses

- Vérifier si les racines des polynômes caractéristiques sont de module strictement inférieur à 1 et que les polynômes n'ont pas de racine commune
- Tests sur les résidus

## 4.4. Exemple

<http://docs.ufrmd.dauphine.fr/st/part/partie3.pdf>

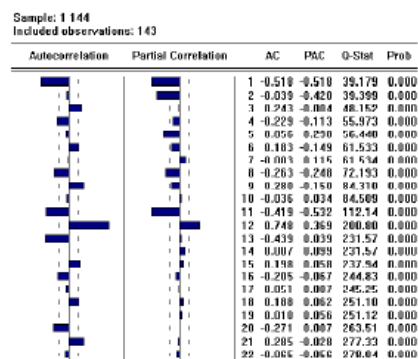
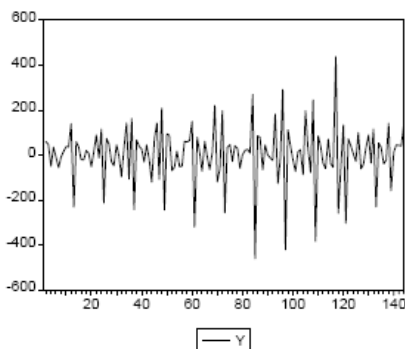
Nous allons considérer ici le nombre de contrats dans un portefeuille, exprimé en nombre de contrats mensuels, à l'aide de données mensuelles, sur une dizaine d'années (source :Droesbeke, Fichet, Tassi (1992))



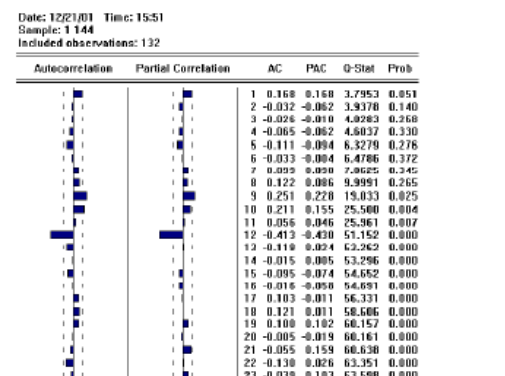
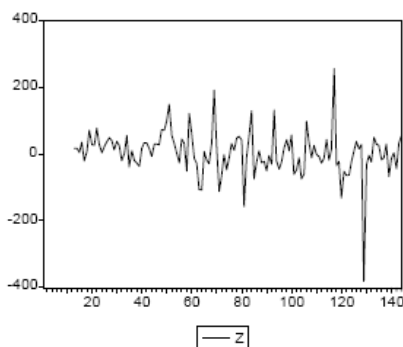
La série brute ( $X_t$ ) présente une forte saisonnalité, correspondant aux mois de septembre. En particulier 4 points se distinguent, correspondant à 4 fortes campagnes de publicité faites à partir de fin août.

### 9.1.1 Modélisation de la série

Afin d'enlever une éventuelle tendance, considérons  $Y_t = (1 - L) X_t = \Delta X_t$



Cette approche n'apportant rien de plus à l'étude de la série, nous pouvons noter que la série brute  $X_t$  possède de fortes autocorrélations  $\hat{\rho}(h)$  quand  $h$  est un multiple de 12. Il est donc possible d'étudier la série 'désaisonnalisée'  $Z_t = (1 - L^{12}) X_t$



Nous allons modéliser cette série ( $Z_t$ ) à l'aide d'un processus *ARMA*.

9.1.2 Estimation de modèles ARMA

Si la série paraît stationnaire, une autocorrélation reste significativement non nulle, pour  $h = 12$ . Il peut être intéressant de considérer un modèle de type MA(12) sur cette série,

$$[1] \quad (1 - L^{12}) X_t = \square + (1 - \theta_1 L^{12}) \varepsilon_t$$

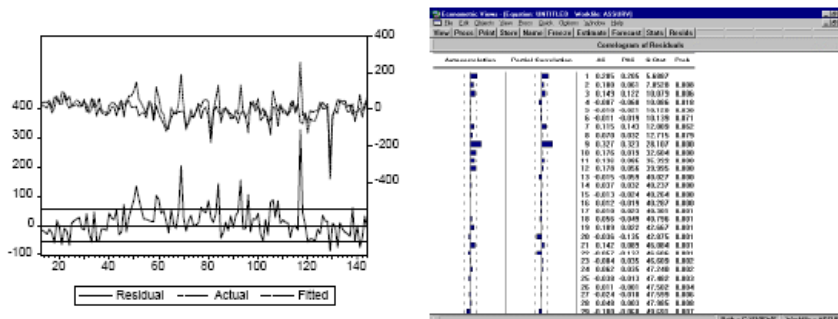
étant donné que les autocorrélations entre 1 et 11 peuvent sembler négligeables. Une méthode basée sur des moindres carrés conditionnels, donne les estimations suivantes (respectivement avec et sans constante)

LS // Dependent Variable is Z				LS // Dependent Variable is Z					
Sample: 13 144				Sample: 13 144					
Included observations: 132 after adjusting endpoints				Included observations: 132 after adjusting endpoints					
Convergence achieved after 17 iterations				Convergence achieved after 14 iterations					
Variable	Coefficient	Std. Error	T-Statistic	Prob.	Variable	Coefficient	Std. Error	T-Statistic	Prob.
C	2.096042	1.845862	1.130118	0.2606	MA(12)	-0.899009	0.032780	-27.42595	0.0000
MA(12)	-0.881818	0.043054	-20.48182	0.0000					
R-squared	0.258916		Mean dependent var	4.750000	R-squared	0.250655		Mean dependent var	4.750000
Adjusted R-squared	0.253216		S.D. dependent var	65.83328	Adjusted R-squared	0.250655		S.D. dependent var	65.83328
S.E. of regression	56.89094		Akaike info criterion	8.097308	S.E. of regression	56.98639		Akaike info criterion	8.093242
Sum squared resid	420755.3		Schwartz criterion	8.140987	Sum squared resid	425445.6		Schwartz criterion	8.115081
Log likelihood	-719.7222		F-statistic	45.41878	Log likelihood	-720.4538		Durbin-Watson stat	1.502360
Durbin-Watson stat	1.517747		Prob(F-statistic)	0.000000					
Inverted MA Roots	.99	.86 - .49i	.86 + .49i	.49 - .86i	Inverted MA Roots	.99	.86 + .50i	.86 - .50i	.50 + .86i
	.49 + .86i	.00 + .99i	-.00 - .99i	-.49 + .86i		.50 - .86i	-.00 - .99i	-.00 + .99i	-.50 - .86i
	-.49 - .86i	-.86 + .49i	-.86 - .49i	-.99		-.50 + .86i	-.86 + .50i	-.86 - .50i	-.99

Pour ce premier modèle, nous obtenons l'écriture suivante

$$[1] \quad (1 - L^{12}) X_t = \left( 1 + \frac{0.89901L^{12}}{(0.0328)} \right) \varepsilon_t$$

où la volatilité  $\square$  de l'erreur ( $\varepsilon_t$ ) est estimée par 57. Les résidus, représentés ci-dessous à gauche, ont l'autocorrélogramme suivant



Le modèle présenté est rejeté : les erreurs ne suivent pas un bruit blanc. En particulier, si l'on considère les autocorrélations de ( $Z_t$ ) la première semble significativement non nulle : ceci pousse à tester un modèle ARMA(1, 12) de la forme  $(1 - \phi L) Z_t = (1 - \theta L^{12}) \varepsilon_t$ , soit

LS // Dependent Variable is Z				LS // Dependent Variable is Z					
Sample: 14 144				Sample: 14 144					
Included observations: 131 after adjusting endpoints				Included observations: 131 after adjusting endpoints					
Convergence achieved after 13 iterations				Convergence achieved after 10 iterations					
Variable	Coefficient	Std. Error	T-Statistic	Prob.	Variable	Coefficient	Std. Error	T-Statistic	Prob.
C	1.969250	2.238375	0.875300	0.3831	AR(1)	0.247243	0.084927	2.911269	0.0042
AR(1)	0.238275	0.086711	2.779982	0.0063	MA(12)	-0.890945	0.034460	-25.85917	0.0000
MA(12)	-0.882966	0.039380	-22.42171	0.0000					
R-squared	0.301621		Mean dependent var	4.656489	R-squared	0.297269		Mean dependent var	4.656489
Adjusted R-squared	0.290709		S.D. dependent var	66.07720	Adjusted R-squared	0.291822		S.D. dependent var	66.07720
S.E. of regression	55.64882		Akaike info criterion	8.060792	S.E. of regression	55.60616		Akaike info criterion	8.051737
Sum squared resid	395403.6		Schwartz criterion	8.126636	Sum squared resid	398973.8		Schwartz criterion	8.095633
Log likelihood	-710.8628		F-statistic	27.64084	Log likelihood	-711.2697		F-statistic	54.58963
Durbin-Watson stat	2.040467		Prob(F-statistic)	0.000000	Durbin-Watson stat	2.049872		Prob(F-statistic)	0.000000
Inverted AR Roots	.24				Inverted AR Roots	.25			
Inverted MA Roots	.99	.86 + .49i	.86 - .49i	.49 - .86i	Inverted MA Roots	.99	.86 + .50i	.86 - .50i	.50 + .86i
	.49 + .86i	.00 + .99i	-.00 + .99i	-.49 - .86i		.50 - .86i	-.00 - .99i	-.00 + .99i	-.50 - .86i
	-.49 - .86i	-.86 + .49i	-.86 - .49i	-.99		-.50 + .86i	-.86 + .50i	-.86 - .50i	-.99



LS // Dependent Variable is Z  
Sample: 14 144  
Included observations: 131 after adjusting endpoints  
Convergence achieved after 13 iterations

Variable	Coefficient	Std. Error	T-Statistic	Prob.
C	3.715450	3.659678	1.012745	0.3131
AR(1)	0.208215	0.088354	2.356593	0.0200
MA(9)	0.226980	0.058097	3.906922	0.0002
MA(12)	-0.684472	0.060959	-11.22845	0.0000

	R-squared	Adjusted R-squared	S.E. of regression	Sum squared resid	Log likelihood	Durbin-Watson stat
	0.338724	0.323103	54.36419	375344.1	-707.2872	1.995098

	Mean dependent var	S.D. dependent var	Akaike info criterion	Schwarz criterion	F-statistic	Prob(F-statistic)
	4.656489	66.07720	8.021470	8.109262	21.68428	0.000000

	Inverted AR Roots	Inverted MA Roots			
	.21	.94	.85 +.46i	.85 -.46i	.50 +.86i
		.50 -.86i	-.02 +.97i	-.02 -.97i	-.47 +.82i
		-.47 -.82i	-.83 +.51i	-.83 -.51i	-.99

LS // Dependent Variable is Z  
Sample: 14 144  
Included observations: 131 after adjusting endpoints  
Convergence achieved after 13 iterations

Variable	Coefficient	Std. Error	T-Statistic	Prob.
AR(1)	0.214089	0.087981	2.433366	0.0163
MA(12)	-0.674685	0.060046	-11.23631	0.0000
MA(9)	0.234509	0.057176	4.101524	0.0001

	R-squared	Adjusted R-squared	S.E. of regression	Sum squared resid	Log likelihood	Durbin-Watson stat
	0.333420	0.323004	54.36816	378354.7	-707.8105	1.997797

	Mean dependent var	S.D. dependent var	Akaike info criterion	Schwarz criterion	F-statistic	Prob(F-statistic)
	4.656489	66.07720	8.014192	8.080036	32.01242	0.000000

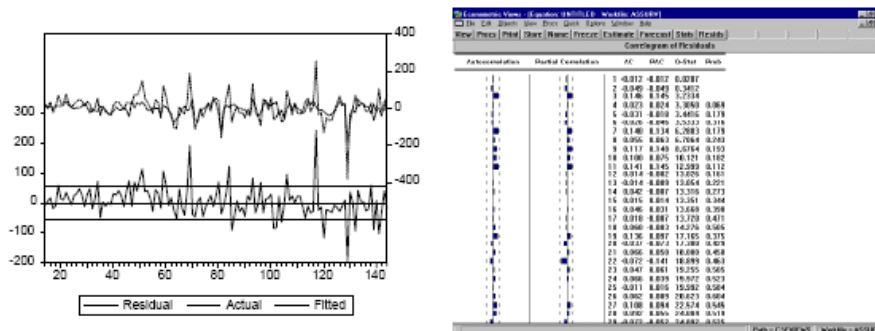
  

	Inverted AR Roots	Inverted MA Roots			
	.21	.94	.85 +.46i	.85 -.46i	.50 +.86i
		.50 +.86i	-.03 +.97i	-.03 -.97i	-.47 -.81i
		-.47 +.81i	-.83 +.51i	-.83 -.51i	-.99

d'où le modèle

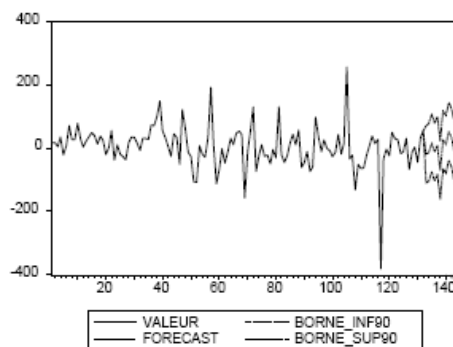
$$[4] \quad \left(1 - \frac{0.21409L}{(0.088)}\right) (1 - L^{12}) X_t = \left(1 - \frac{0.23450L^9 + 0.67498L^{12}}{(0.060)}\right) \varepsilon_t$$

Là encore, le modèle est significatif, et surtout, l'hypothèse de bruit blanc des résidus est validée :



Ce modèle peut alors être utilisé pour faire de la prévision. Sous SAS, nous obtenons les prévisions suivantes, sur 10 mois, pour  $Z_t$

$h$	1	2	3	4	5	6	7	8	9	10
$T Z_{T+h}^*$	-19.92	-12.88	16.18	-12.52	9.18	-69.64	27.40	9.86	52.36	32.02
volatilité	55.20	56.08	56.11	56.11	56.11	56.11	56.11	56.11	56.11	56.11



On peut alors noter que  ${}_T X_{T+h}^* = {}_T X_{T+h-12}^* + {}_T Z_{T+h}^*$ . De plus, la variable  $X_t$  peut alors être prédite en utilisant la forme  $MA(\infty)$  du processus  $Z_t$  :  $(1 - \phi L) Z_t = (1 - \alpha L^9 - \beta L^{12}) \varepsilon_t$ . Ainsi,

$$\begin{aligned} Z_t &= (1 - \phi L)^{-1} (1 - \alpha L^9 - \beta L^{12}) \varepsilon_t = (1 - 0.21409.L)^{-1} (1 - 0.23450.L^9 + 0.67498.L^{12}) \varepsilon_t \\ &= (1 - 0.23450.L^9 + 0.67498.L^{12}) (1 + 0.2141.L + 0.2141^2.L^2 + 0.2141^3.L^3 + \dots) \varepsilon_t \end{aligned}$$

le polynôme  $MA B(L)$  s'écrit alors

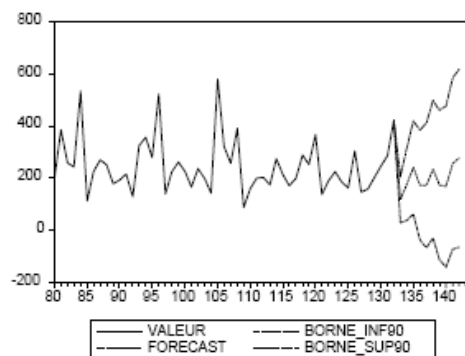
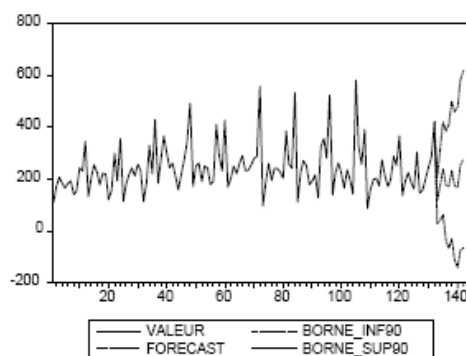
$$B(L) = \sum_{i=0}^{\infty} b_i L^i \text{ où } b_i = \begin{cases} 1 & \text{pour } i = 0 \\ 0.2141^i & \text{pour } 1 \leq i \leq 8 \\ 0.2141^i - 0.2141^{i-9} \times 0.2345 & \text{pour } 9 \leq i \leq 11 \\ 0.2141^i - 0.2141^{i-9} \times 0.2345 + 0.2141^{i-12} \times 0.6745 & \text{pour } 11 \leq i \end{cases}$$

La variance de la prévision  ${}_T X_{T+h}^*$  est alors donnée par

$$V({}_T X_{T+h}^*) = \sigma^2 \sum_{j=0}^{h-1} [1 + b_1 + \dots + b_j]^2$$

soit ici, puisque  $\sigma^2$  est estimé par  $s^2 = 54.368^2$ ,

$h$	1	2	3	4	5	6	7	8	9	10
${}_T X_{T+h}^*$	117.08	177.12	241.18	173.48	172.18	234.36	173.4	167.86	255.36	277.02
volatilité	54.37	85.52	109.57	129.50	146.81	162.29	176.41	189.49	201.73	209.47



## 5. PREVISION DANS UN PROCESSUS ARMA

Etant donnée une série stationnaire  $\{X_t\}$ , observée entre les instants 1 et  $T$ , on cherche à faire la prévision à l'horizon  $h$ , et donc de prévoir  $X_{T+1}, \dots, X_{T+h}$ . On sait alors que la meilleure approximation de  $X_{T+h}$  par une fonction des  $X_t$  avec  $t < T$ , est l'espérance conditionnelle

$$\hat{X}_{T+h} = E[X_{T+h} | X_T, X_{T-1}, \dots, X_1].$$

La forme de l'espérance conditionnelle va dépendre du modèle utilisé.

### 5.1. Prévision avec un processus AR

Supposons que  $\{X_t\}$  soit modélisé par un modèle  $AR(p)$  de coefficients connus et nous allons procéder de façon récurrente.

- A l'horizon 1, on a le modèle

$$X_{T+1} = a_1 X_T + a_2 X_{T-1} + \dots + a_p X_{T-1-p} + \varepsilon_{T+1}.$$

Or  $\varepsilon_{T+1}$  est orthogonal à  $X_T, X_{T-1}, \dots$ , donc

$$\hat{X}_{T+1} = E[X_{T+1} | X_T, X_{T-1}, \dots, X_1] = a_1 X_T + a_2 X_{T-1} + \dots + a_p X_{T-1-p}$$

- A l'horizon 2, on a

$$X_{T+2} = a_1 X_{T+1} + a_2 X_T + \dots + a_p X_{T-p} + \varepsilon_{T+2}$$

d'où

$$\hat{X}_{T+2} = a_1 X_{T+1} + a_2 X_T + \dots + a_p X_{T-p}$$

or  $X_{T+1}$  n'a pas été observé mais est estimé à l'horizon 1 par  $\hat{X}_{T+1}$

$$\hat{X}_{T+2} = a_1 \hat{X}_{T+1} + a_2 X_T + \dots + a_p X_{T-p}$$

- A l'horizon h, on procède de façon récurrente pour obtenir

$$\hat{X}_{T+h} = \begin{cases} (a_1 \hat{X}_{T+h-1} + \dots + a_{h-1} \hat{X}_{T+1}) + (a_h X_T + \dots + a_p X_{T+h-p}) & \text{si } h \leq p \\ a_1 \hat{X}_{T+h-1} + \dots + a_p \hat{X}_{T+h-p} & \text{si } h > p \end{cases}$$

### Exemple

Prenons le cas très simple d'un AR(1),  $X_t = aX_{t-1} + \varepsilon_t$ . On a alors

$$X_{T+1} = aX_T + \varepsilon_{T+1} \Rightarrow \hat{X}_{T+1} = aX_T$$

$$X_{T+2} = aX_{T+1} + \varepsilon_{T+2} \Rightarrow \hat{X}_{T+2} = aX_{T+1} = a^2 X_T$$

$$X_{T+h} = aX_{T+h-1} + \varepsilon_{T+h} \Rightarrow \hat{X}_{T+h} = aX_{T+h-1} = a^h X_T$$

## 5.2. Prédiction dans les processus MA

Supposons que  $\{X_t\}$  soit modélisé par un modèle MA(q) de coefficients connus,

$$X_t = \varepsilon_t + b_1 \varepsilon_{t-1} + \dots + b_q \varepsilon_{t-q}.$$

Si on procède de façon récurrente comme précédemment, on obtient

$$X_{T+1} = \varepsilon_{T+1} + b_1 \varepsilon_T + \dots + b_q \varepsilon_{T+1-q}$$

$$\Rightarrow \hat{X}_{T+1} = b_1 \varepsilon_T + \dots + b_q \varepsilon_{T+1-q} \text{ car } \varepsilon_{T+1} \text{ est orthogonal à } X_T, X_{T-1}, \dots$$

$$X_{T+2} = \varepsilon_{T+2} + b_1 \varepsilon_{T+1} + b_2 \varepsilon_T + \dots + b_q \varepsilon_{T+2-q}$$

$$\Rightarrow \hat{X}_{T+2} = b_2 \varepsilon_T + \dots + b_q \varepsilon_{T+2-q} \text{ car } \varepsilon_{T+2} \text{ et } \varepsilon_{T+1} \text{ sont orthogonaux à } X_T, X_{T-1}, \dots$$

De proche en proche, on obtient  $\hat{X}_{T+h}$  et on remarque qu'il est nul pour  $h > q$ . Cette méthode présente le désavantage d'estimer  $X_{T+h}$  à partir des résidus passés (*a priori* non observables) et non pas à partir du passé de la variable. Pour contrer ce problème, nous allons utiliser l'écriture AR( $\infty$ ) du processus MA(q),

$$X_t = -\sum_{n \geq 1} \beta_n X_{t-n} + \varepsilon_t.$$

$$\begin{aligned}
X_{T+1} &= -\sum_{n \geq 1} \beta_n X_{T+1-n} + \varepsilon_{T+1} \Rightarrow \hat{X}_{T+1} = -\sum_{n \geq 1} \beta_n X_{T+1-n} \\
X_{T+2} &= -\sum_{n \geq 1} \beta_n X_{T+2-n} + \varepsilon_{T+2} \Rightarrow \hat{X}_{T+2} = -\beta_1 X_{T+1} - \sum_{n \geq 2} \beta_n X_{T+2-n} \\
&\Rightarrow \hat{X}_{T+2} = -\beta_1 \hat{X}_{T+1} - \sum_{n \geq 2} \beta_n X_{T+2-n} \\
X_{T+h} &= -\sum_{n \geq 1} \beta_n X_{T+h-n} + \varepsilon_{T+h} \Rightarrow \hat{X}_{T+h} = -\sum_{n=1}^{h-1} \beta_n X_{T+h-n} - \sum_{n \geq h} \beta_n X_{T+h-n} \\
&\Rightarrow \hat{X}_{T+h} = -\sum_{n=1}^{h-1} \beta_n \hat{X}_{T+h-n} - \sum_{n \geq h} \beta_n X_{T+h-n}
\end{aligned}$$

Le problème dans l'expression ci-dessus est que les  $X_t$  ne sont pas observés pour  $t < 0$ . On va donc tronquer la somme infinie pour ne garder que les indices positifs de  $X_t$  et négliger le reste de la somme,

$$\hat{X}_{T+h} = -\sum_{n=1}^{h-1} \beta_n \hat{X}_{T+h-n} - \sum_{n=h}^{T+h} \beta_n X_{T+h-n} .$$

### 5.3. Prédiction dans les processus ARMA

Supposons que  $\{X_t\}$  soit modélisé par un modèle ARMA(p,q) de coefficients connus,

$$X_t - a_1 X_{t-1} - \dots - a_p X_{t-p} = \varepsilon_t + b_1 \varepsilon_{t-1} + \dots + b_q \varepsilon_{t-q} .$$

On procède comme pour les processus MA, c'est-à-dire qu'on utilise l'écriture sous la forme d'un processus AR( $\infty$ ), puis on tronque la somme pour ne garder que les indices positifs de  $X_t$ .

Appliquons cette méthode au cas d'un ARMA(1,1),

$$X_t - aX_{t-1} = \varepsilon_t + b_1 \varepsilon_{t-1},$$

tel que  $0 < |a| < 1$  et  $0 < |b| < 1$ . Si on l'écrit avec un opérateur de retard, on a

$$(1 - aB)X_t = (1 + bB)\varepsilon_t .$$

$(1 - bB)^{-1}$  peut se développer en série puisque  $|b| < 1$ , d'où

$$\begin{aligned}
\varepsilon_t &= (1 - aB)(1 + bB)^{-1} \varepsilon_t \\
&= (1 - aB)(1 - bB + b^2 B^2 - b^3 B^3 + \dots) X_t \\
&= X_t + (a + b) \sum_{n \geq 1} (-1)^n b^{n-1} X_{t-n} .
\end{aligned}$$

$$\Leftrightarrow X_t = (a + b) \sum_{n \geq 1} (-1)^n b^{n-1} X_{t-n} + \varepsilon_t$$

$$X_{T+1} = (a + b) \sum_{n \geq 1} (-1)^n b^{n-1} X_{T+1-n} + \varepsilon_{T+1} \Rightarrow \hat{X}_{T+1} = (a + b) \sum_{n \geq 1} (-1)^n b^{n-1} X_{T+1-n}$$

$$\begin{aligned}
X_{T+2} &= (a + b) \sum_{n \geq 1} (-1)^n b^{n-1} X_{T+2-n} + \varepsilon_{T+2} \\
&\Rightarrow \hat{X}_{T+2} = -(a + b) X_{T+1} + (a + b) \sum_{n \geq 2} (-1)^n b^{n-1} X_{T+2-n}
\end{aligned}$$

$$\begin{aligned} \Rightarrow \hat{X}_{T+2} &\approx -(a+b)\hat{X}_{T+1} + (a+b)\sum_{n=2}^{T+2} (-1)^n b^{n-1} X_{T+2-n} \\ X_{T+3} &= (a+b)\sum_{n \geq 1} (-1)^n b^{n-1} X_{T+3-n} + \varepsilon_{T+3} \\ \Rightarrow \hat{X}_{T+3} &= -(a+b)X_{T+2} + (a+b)bX_{T+1} + (a+b)\sum_{n \geq 3}^{T+3} (-1)^n b^{n-1} X_{T+3-n} \\ \Rightarrow \hat{X}_{T+3} &\approx -(a+b)\hat{X}_{T+2} + (a+b)b\hat{X}_{T+1} + (a+b)\sum_{n=3}^{T+3} (-1)^n b^{n-1} X_{T+3-n} \end{aligned}$$

Par récurrence, on obtient

$$\hat{X}_{T+h} \approx (a+b)\sum_{n=1}^{h-1} (-1)^n b^{n-1} \hat{X}_{T+h-n} + (a+b)\sum_{n=h}^{T+h} (-1)^n b^{n-1} X_{T+h-n}.$$

#### 5.4. Intervalle de confiance

Nous avons vu précédemment que les processus AR, MA et ARMA peuvent s'écrire sous la forme MA( $\infty$ )

$$X_t = \varepsilon_t + \sum_{n \geq 1} \beta_n \varepsilon_{t-n}.$$

Pour un modèle MA(q), il suffit de poser  $b_n=0$  pour  $n>q$ .

On en déduit que

$$X_{T+h} = \varepsilon_{T+h} + \sum_{n \geq 1} \beta_n \varepsilon_{T+h-n} = \varepsilon_{T+h} + \underbrace{\sum_{n=1}^{h-1} \beta_n \varepsilon_{T+h-n}}_{\text{indices } > T} + \underbrace{\sum_{n=h}^{\infty} \beta_n \varepsilon_{T+h-n}}_{\text{indices } \leq T}$$

$$\Rightarrow \hat{X}_{T+h} = \sum_{n=1}^{\infty} \beta_n \varepsilon_{T+h-n}$$

D'où l'erreur de prévision

$$e_h = X_{T+h} - \hat{X}_{T+h} = \varepsilon_{T+h} + \sum_{n=1}^{h-1} \beta_n \varepsilon_{T+h-n}$$

Si on suppose de plus que le bruit blanc est gaussien alors on peut en déduire que l'erreur de prévision suit une loi normale

$$e_h \sim N(0, \sigma^2 (1 + \sum_{n=1}^{h-1} \beta_n^2))$$

D'où l'intervalle de confiance de niveau  $1-\alpha$ ,

$$\hat{X}_{T+h} \pm z_{1-\alpha/2} \sigma \sqrt{1 + \sum_{n=1}^{h-1} \beta_n^2}.$$