



**Institut des Hautes Etudes Commerciales de Sousse**

# **Econométrie des séries temporelles**

Mr. Boubaker Heni

4<sup>ème</sup> Année :  
Actuariat et Finance

Année universitaire : 2006- 2007

# Plan du cours

## Chapitre 1 : Introduction

- 1 Préliminaires sur les séries temporelles
- 2 Exemples de séries temporelles
- 3 Objectifs de l'analyse d'une série temporelle
- 4 Tendances, Saisonnalité et Résidus

## Chapitre 2 : Méthodes sans modèle

- 2.1 Extrapolation déterministe des séries
  - 2.1.1 Tendances linéaires
  - 2.1.2 Tendances autorégressives
  - 2.1.3 Modèles non linéaires

2.2 Moyennes mobiles

2.3 Lissages exponentiels

2.3 1 Lissage exponentiel simple

2.3 2 Lissage exponentiel double

2.3 3 Lissage exponentiel multiples ou généralisé

3. Méthodes de Holt-Winters

## Chapitre 3 : Introduction aux modèles linéaires ARIMA

3.1 Rappels sur les espaces  $\mathcal{L}^2$

3.2 Polynômes d'opérateurs retard L et avance F

3.3 Compléments sur les séries stationnaires

3.3.1 Autocovariance et autocorrélation

3.3.2 Autocorrélations partielles

3.3.3 Densité spectrale

3.4 Les processus autorégressifs : AR (p)

- 3.4.1 Réécriture de la forme AR (p)
- 3.4.2 Propriétés des autocorrélations : les équations de Yule-Walker
- 3.4.3 Le processus AR (1)
- 3.4.4 Le processus AR (2)
- 3.5 Les processus moyenne-mobile : MA(q)
  - 3.5.1 Propriétés des autocorrélations
  - 3.5.2 Le processus MA(1)
  - 3.5.3 Le processus MA(2)
- 3.6 Les processus ARMA(p,q)
  - 3.6.1 Propriétés des autocorrélations
  - 3.6.2 Densité spectrale des processus ARMA(p,q)
  - 3.6.3 Les processus ARMA(1,1)
- 3.7. Introduction aux modèles linéaires non-stationnaires
  - 3.7.1 Les processus ARIMA(p; d; q)
  - 3.7.2 Les modèles SARIMA
- 3.8 Théorème de Wold

# Chapitre 4 : Estimation des modèles ARIMA : Box-Jenkins

- 4.1 Estimation du paramètre d'intégration  $d$ 
  - 4.1.1 Approche empirique par l'autocorrélogramme
  - 4.1.2 Tests de racine unité
  - 4.1.3 Tests de racines unitaires saisonnières
- 4.2 Estimation des ordres  $p$  et  $q$  d'un modèle ARMA( $p; q$ )
  - 4.2.1 Problèmes d'unicité de la représentation ARMA
  - 4.2.2 Comportement asymptotique des moments empiriques
  - 4.2.3 Méthode pratique d'estimation des ordres  $p$  et  $q$
- 4.3 Test de bruit blanc et de stationnarité
  - 4.3.1 Analyse des fonctions d'autocorrélation.
  - 4.3.2 Statistique de Box-Pierce, ou test de "portmanteau"
- 4.4 Estimation des paramètres d'un modèle ARMA ( $p; q$ )
  - 4.4.1 Vraisemblance d'un processus ARMA ( $p; q$ )

- 4.4.2 Résolution du programme d'optimisation
- 4.4.3 Tests statistiques de validation du modèle
- 4.5 Choix d'un modèle
  - 4.5.1 Critère de pouvoir prédictif
  - 4.5.2 Critère d'information
- 4.6 Prévisions à l'aide des modèles ARIMA

## Chapitre 5 : Les modèles ARCH - Autorégressifs Conditionnellement Hétéroscédastiques

- 5.1 Présentation des modèles ARCH
  - 5.1.1 Processus ARCH (1)
  - 5.1.2 Processus ARCH (p)
  - 5.1.3 Processus GARCH (p; q)
- 5.2 Modèles avec erreurs ARCH
  - 5.2.1 Erreurs ARCH (1)
  - 5.2.2 Erreurs ARCH (p)

5.2.3 Remarque : test de racine unité en présence d'erreurs ARCH

5.4 Estimation et prévision

---

---

## **Séances de TD/TP**

- Approfondissement théorique + exercices construits et corrigés ensembles
- Vous êtes mis face à des données qui impliquent un type de modélisation économétrique
- J'utiliserai comme logiciels : EViews et OxMetrics dans la mesure où cela n'implique pas des calculs trop compliqués.
- Vous travaillerez chez-vous sur EViews ou bien sur le logiciel gratuit JMulti.

## **Notes d'évaluation : 12,5%**

- Contrôle de connaissances d'1h, à la 8ème semaine,
- **ou**, dossier  $\leq$  10 pages par groupe de 3 au maximum : application de techniques économétriques du cours sur des données qui vous intéressent. A rendre au plus tard la deuxième semaine du moi de décembre.



# **Bibliographie**

- Bourbonnais R. *Econométrie*, 3<sup>ème</sup> éd., 2000, Dunod
- Bosq D. et Lecoutre J-P. *Modélisation des Séries chronologiques*, Masson
- Bresson G., Pirotte A. *Econométrie des séries temporelles théorie et applications*, 1995, 1<sup>ère</sup> éd, Presses Universitaires de France.
- Gourieroux C., Monfort A. *Séries temporelles et modèles dynamiques*, 1995, 2<sup>ème</sup> éd, Economica.
- Gourieroux C. *Modèles ARCH et applications financières*, 1992, Economica.
- Johnston J., Dinardo J. *Méthodes économétriques*, 1999, 4<sup>ème</sup> éd, Economica.
- Lardic S., Mignon V. *Econométrie des séries temporelles macroéconomiques et financières*, 2002, Economica.

# Introduction

Ce cours est une initiation aux méthodes statistiques de prévision.

- Pourquoi vouloir prévoir ?

Très souvent pour des raisons économiques (prévoir l'évolution des ventes d'un certain produit, prévoir la consommation d'électricité pour ajuster au mieux la production...).

- Comment prévoir ?

Toujours en s'appuyant sur ce qui s'est passé avant l'instant à partir duquel démarre la prévision. Exemple Dans la prévision des ventes pour l'année  $j+1$  on s'appuiera sur l'évolution des ventes durant les années  $j, j-1, \dots$ , mais on tiendra compte aussi, éventuellement,

d'indices exogènes (la conjoncture économique, certains événements récents inattendus etc....).

- Peut on prévoir parfaitement bien ?

Jamais. Il y a toujours une erreur de prévision, et les bonnes méthodes fournissent non pas une prévision, mais un intervalle de prévision. Il arrive souvent qu'une amélioration minime de la qualité de la prévision ait une grosse incidence sur les coûts.

- De quelles méthodes dispose-t-on pour prévoir ?

Elles sont nombreuses.

1- Vous connaissez la plus rudimentaire (commode), qui consiste à faire une régression et à se baser sur elle pour prévoir. Par exemple on ajuste sur la série  $X_1, \dots, X_T$  un modèle de type :

$$X_j = a_1 j^2 + a_2 j + a_3 + e_j \quad j = 1, \dots, T.$$

On estime les coefficients par une méthode de régression. On valide le résultat, puis on prévoit  $X_{T+1}$  par :

$$\hat{a}_1(T+1)^2 + \hat{a}_2(T+1) + \hat{a}_3,$$

et  $X_{T+2}$  ainsi que les suivants de façon analogue.

**2-** Les lissages exponentiels sont une batterie de procédures extrêmement simples à mettre en œuvre. Ils sont exposés au chapitre 2.

**3-** Les méthodes de prévision les plus populaires depuis une trentaine d'années sont celles liées à la modélisation ARMA. Elles consistent en gros à dégager les tendances évidentes dans le phénomène qu'on observe (croissance, périodicités etc...) et à se concentrer sur ce qui reste lorsqu'on les a supprimées. On procède

à une modélisation fine du résidu obtenu et on s'appuie sur cette modélisation pour obtenir la prévision.

Ces méthodes, plus sophistiquées mais plus performantes que la régression et que les lissages exponentiels, sont coûteuses en temps calcul. Leur utilisation systématique dans tous les domaines, que ce soit économique, industriel, physique etc..., est due au développement exponentiellement rapide de l'outil informatique. Elles seront l'objet principal de ce cours. Il y a bien d'autres méthodes dont on ne parlera pas ici.

Certaines données sont faciles et, sur elles, toutes les méthodes donnent de bonnes prévisions. Tout l'art du statisticien, face à des données difficiles, est de confronter plusieurs méthodes et de choisir celle qui convient le mieux.

# Préliminaires sur les séries temporelles

Une série temporelle (ou encore une série chronologique) est une suite finie  $(X_1, \dots, X_T)$  de données indexées par le temps. L'indice temps peut être selon les cas la minute, l'heure, le jour, l'année etc.... Le nombre  $T$  est appelé la longueur de la série. La date à laquelle l'observation est faite est une information importante sur le phénomène observé.

Il est la plupart du temps bien utile de représenter la série temporelle sur un graphe construit de la manière suivante : en abscisse le temps, en ordonnée la valeur de l'observation à chaque instant. Pour des questions de lisibilité, les points ainsi obtenus sont reliés par des segments de droite. Le graphe apparaît donc comme une ligne brisée. **Exemples** ....

## **Quelques précautions élémentaires**

Quel que soit le type de variable étudié, on s'assurera de respecter les points suivants.

### **1- Régularité des observations**

Elle est parfaite dans le cas de certains relevés météorologiques mais c'est déjà moins vrai pour beaucoup de variables économiques ou financières puisque les mois ne comportent pas le même nombre de jours, en particulier de jours ouvrables. Une correction par simple proportionnalité peut être envisagée mais elle change la signification concrète des valeurs manipulées.

### **2- Stabilité des structures conditionnant le phénomène étudié**

La plupart des chroniques étudiées concernent des grandeurs économiques et les techniques d'analyse cherchent à déterminer l'évolution lente du phénomène ainsi que ses variations

saisonniers (pour une meilleure compréhension ou à des fins de prévision). Cela suppose une certaine stabilité qui, lorsqu'elle n'est pas vérifiée, peut être obtenue en décomposant la chronique observée en plusieurs chroniques successives (empiriquement à l'aide d'une représentation graphique où à partir de la connaissance des modifications de l'environnement économique).

### **3- Permanence de la définition de la grandeur étudiée**

Cette condition, qui paraît évidente, n'est parfois pas respectée. C'est en particulier le cas de certains indices économiques (changement du mode de calcul de l'indice)

### **4- Aspect périodique d'une partie de la grandeur observée**

Cette condition est indispensable dans l'usage des techniques cherchant à déterminer des variations saisonnières. Elle suppose comparable deux observations relatives au même mois de deux



années différentes. Elle n'exclut pas l'existence d'une évolution lente. Elle indique qu'une part du phénomène (la composante saisonnière) se répète de façon plus ou moins identique d'une année à l'autre. Dans ce cas il est souvent commode d'indexer la chronique à l'aide de deux indices,  $y_{ij}$  représente l'observation du  $j^{\text{ème}}$  mois de la  $i^{\text{ème}}$  année, et les données sont listées dans une table à double entrée.

⇒ Contrairement à l'économétrie traditionnelle, le but de l'analyse des séries temporelles n'est pas de relier des variables entre elles, mais de s'intéresser à la dynamique d'une variable. Cette dernière est en effet essentielle pour deux raisons : les avancées de l'économétrie ont montré qu'on ne peut relier que des variables qui présentent des propriétés similaires, en particulier une même stabilité ou instabilité ; les propriétés mathématiques des modèles permettant d'estimer le lien entre deux variables dépendent de leur dynamique.

# Objectifs de l'analyse d'une série temporelle

Décrire Quand on s'intéresse à une série temporelle, la première étape, comme pour toutes données est de décrire la série. On utilise pour ce faire un certain nombre de graphiques. Sur les graphiques on peut repérer les valeurs atypiques ou aberrantes. On calcule aussi les statistiques descriptives usuelles : moyenne, variance, coefficients d'aplatissement et d'asymétrie. Une même série temporelle peut être analysée de différentes façons suivant l'objectif poursuivi.

Modéliser Expliquer le niveau ou parfois la variance du niveau, par des modèles à peu de paramètres.

**Prédire** La prévision de valeurs à des dates futures connaissant le présent et le passé de la série peut être basée sur un modèle ou bien être construite sans référence à un modèle.

## **Résumé**

Série chronologique : données mesurées sur une période de temps. Exemples: valeur d'un titre, taux d'inflation, taux de chômage, ventes.

Si possible, les données sont recueillies à des intervalles de temps réguliers.

Le but est de détecter des comportements réguliers qui aideront à prévoir des valeurs futures.

# Tendance. Saisonnalité. Résidus

Il est classique de décomposer une série temporelle :

## **Tendance à long terme**

Croissance ou décroissance régulière sur une longue période.

## **Effet cyclique**

comportement régulier se répétant périodiquement sur une longue période.

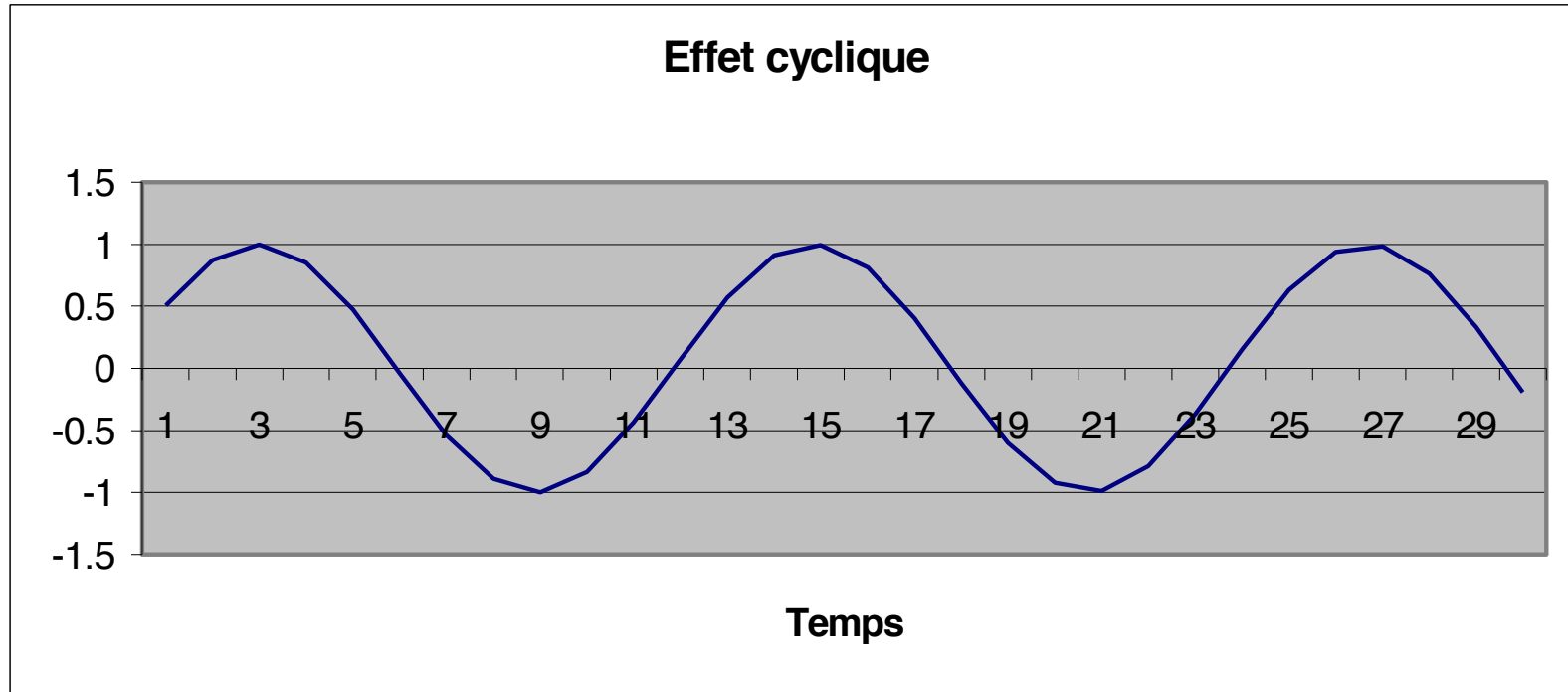
## **Effet saisonnier**

comportement régulier se répétant périodiquement sur une courte période.

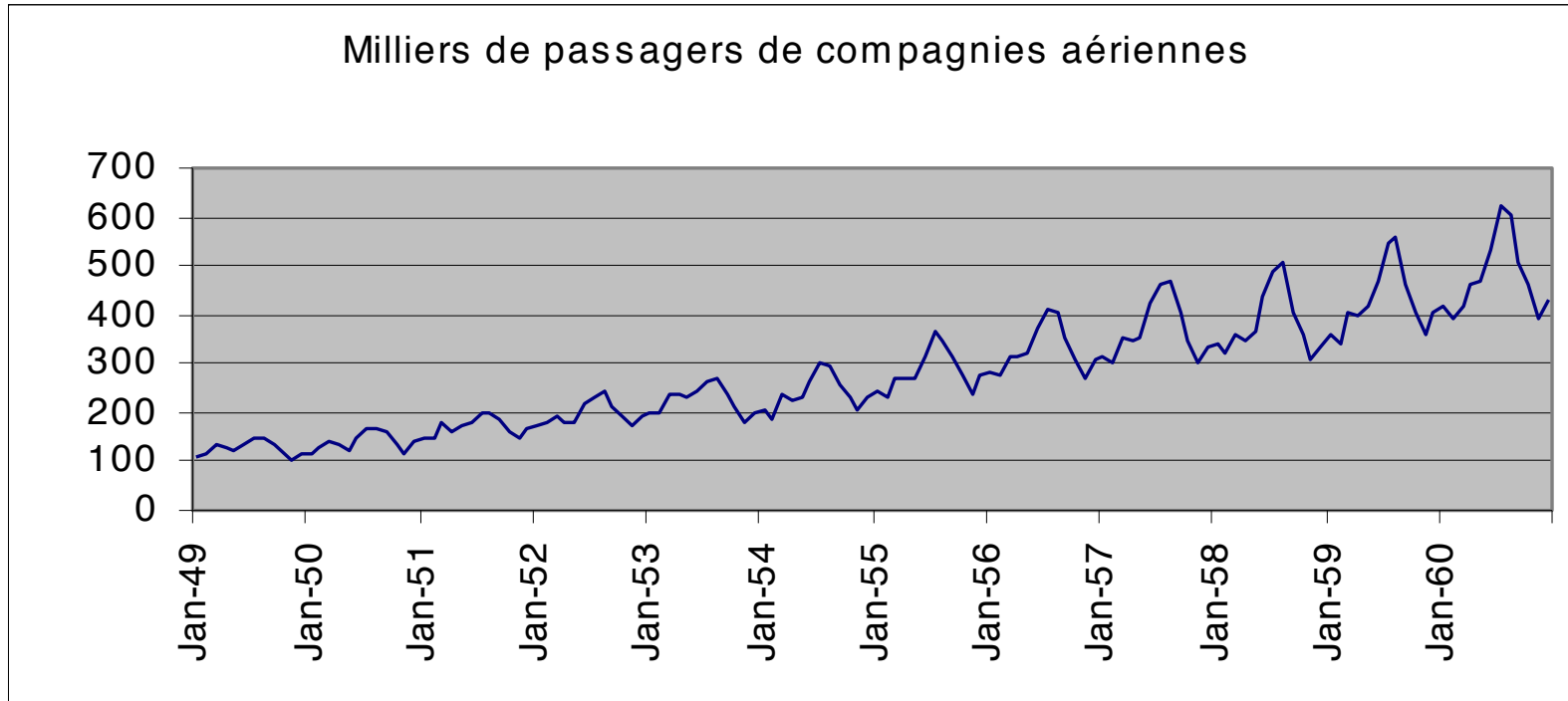
## **Variation aléatoire**

Variations irrégulières et imprévisibles.

Un cycle est un comportement régulier se répétant périodiquement sur une longue période (plus d'une année).



L'effet saisonnier est similaire à l'effet cyclique sauf que la période est plus courte (moins d'une année).



## Effet aléatoire

Les variations aléatoires (aussi appelées bruit) incluent tous les changements irréguliers qui ne sont pas dus aux autres effets (tendance, cycle, saisonnalité).

Le bruit est comme un brouillard nous empêchant de voir les autres composantes.

Un des objectifs est de tenter de se débarrasser de cet effet (en utilisant le lissage par exemple).

## Modèles

modèle additif

$$y_t = T_t + C_t + S_t + R_t$$

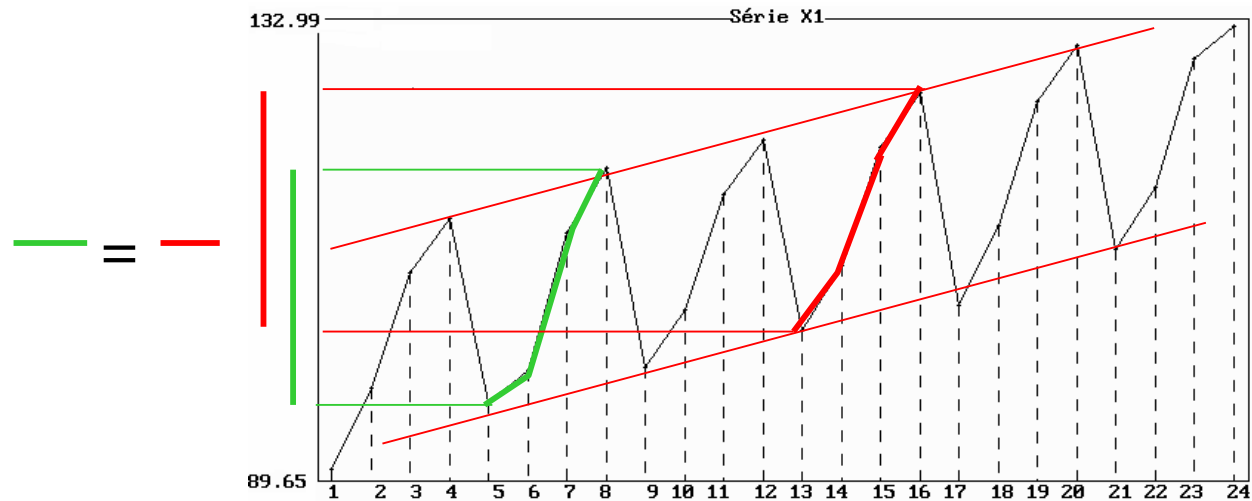
modèle multiplicatif

$$y_t = T_t \cdot C_t \cdot S_t \cdot R_t$$

Par exemple, les séries montrant une saisonnalité qui a de plus en plus d'ampleur, sont souvent mieux ajustées par un modèle multiplicatif que par un modèle additif.

## Cas d'un modèle additif

- Présence de fluctuations saisonnières
- Variations d'amplitude stables
- Les fluctuations saisonnières se compensent sur l'ensemble du cycle
- + tendance
- + variations accidentelles



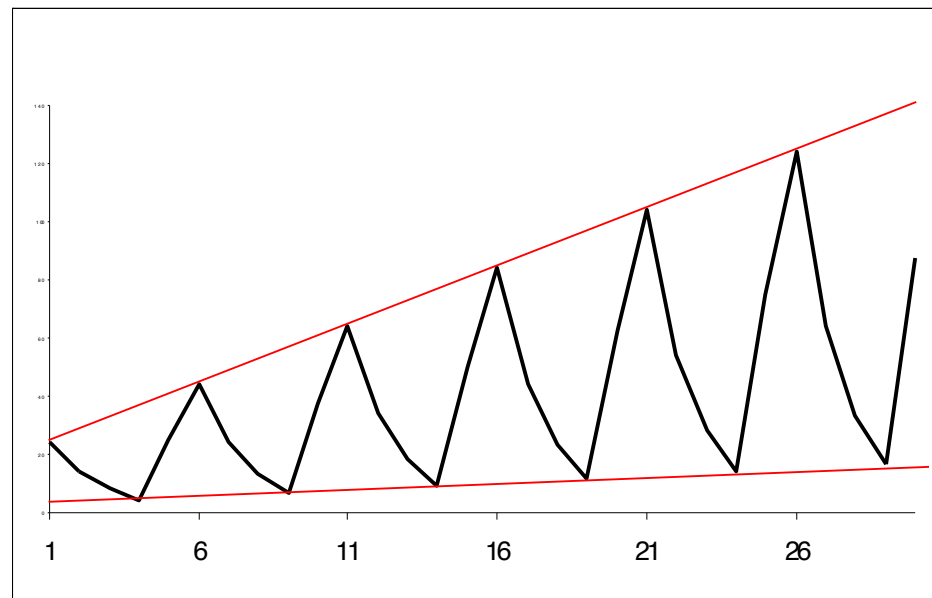
Saisonnalité additive



## Cas d'un modèle multiplicatif

- Présence de fluctuations saisonnières
- Différentes variations d'amplitude
- + tendance
- + variations accidentelles

Transformation logarithmique  $\Rightarrow$  Modèle additif



Saisonnalité multiplicative

# Méthodes sans modèle

## Lissage exponentiel :

Le lissage exponentiel est utilisé pour effectuer une prévision à l'horizon  $h \geq 1$ , notée  $\hat{Y}_T(h)$ , au vu des observations  $Y_1, Y_2, \dots, Y_T$ . Cependant la suite des prévisions à un pas,  $\hat{Y}_T(1)$ ,  $T = 1, 2, \dots$ , constitue un lissage de la série. La particularité consiste à accorder aux valeurs passées une importance qui décroît de manière exponentielle avec le temps, on parle de *facteur d'oubli*.

L'autre point important est que la mise à jour de  $\hat{Y}_T(h)$ , lors de l'acquisition d'une nouvelle observation  $Y_{T+1}$ , est réalisée de façon simple.

Enfin la convention  $Y_t = 0$  pour  $t \leq 0$  facilite la présentation des résultats.

## 1 Lissage exponentiel simple :

La prévision  $\hat{Y}_T(h)$  fourni par la méthode du *lissage exponentiel simple*, avec la constante de lissage satisfaisant  $0 \leq \gamma \leq 1$ , est définie par :

$$\hat{Y}_T(h) = (1 - \gamma) \sum_{k=0}^{\infty} \gamma^k Y_{T-k}, \quad h = 1, 2, \dots$$

Ainsi  $\hat{Y}_T(h)$  est une moyenne pondérée des valeurs passées puisque  $(1 - \gamma) \sum_{k=0}^{\infty} \gamma^k = 1$ . Notons que  $\hat{Y}_T(h)$  ne dépend pas de  $h$ . Cette constante  $\hat{Y}_T$  s'interprète aussi comme une *valeur lissée* de la grandeur observée à l'instant  $T$ , on parle de *filtrage* de la série. Les relations,

$$\hat{Y}_{T+1} = \hat{Y}_T + (1 - \alpha) Y_{T+1} = \hat{Y}_T + (1 - \alpha) (Y_{T+1} - \hat{Y}_T),$$

montrent que  $(1 - \alpha)$ , appelé *facteur d'oubli*, représente le poids accordé à la nouvelle acquisition. Elles permettent de réaliser le filtrage de manière récursive. On utilise  $\hat{Y}_0 = Y_1$  comme valeur initiale plutôt que  $\hat{Y}_0 = 0$ , la différence est rapidement négligeable et les deux séries  $\hat{Y}_{T-1}$  et  $Y_T$ ,  $T \geq 1$ , partent ainsi de la même valeur  $Y_1$  ou encore la prévision de  $Y_2$  au vu de  $Y_1$  est  $Y_1$ .

La prévision  $\hat{Y}_T$  est égale à l'estimateur  $\hat{\beta}_T$  d'une moyenne constante ajoutée à la série observée par la méthode des moindres carrés pondérés :

$$\min_b \sum_{k=0}^{\infty} \alpha^k (Y_{T-k} - b)^2 \Rightarrow \hat{Y}_T \approx \hat{b} = \hat{\beta}_T.$$

- ◆ La méthode n'est donc pas adaptée, sous cette forme simple, à une chronique présentant une tendance variant fortement et / ou un effet saisonnier très marqués.
- ◆ Elle permet de suivre une tendance à variation très lente, en l'absence d'effet saisonnier, que l'on ne veut pas paramétrer sous une forme rigide mais comme une constante pouvant évoluer au cours du temps.
- ◆ Les généralisations qui suivent sont obtenues selon le même principe des moindres carrés pondérés et permettent de prendre en compte une tendance polynomiale ainsi qu'un mouvement saisonnier ajustés localement (les coefficients dépendent de  $T$ ) par la présence du facteur d'oubli.
- ◆ L'intérêt de la méthode est d'aboutir à des formules de mise à jour des paramètres sans avoir à recalculer à chaque instant la solution du modèle linéaire.

- ◆ La simplicité des formules obtenues tient à la nature exponentielle du facteur d'oubli qui permet d'utiliser des approximations.
- ◆ Le choix de la constante de lissage est évidemment important. Les valeurs proches de 0 correspondent à un lissage rigide, car le passé intervient peu, alors que les valeurs proches de 1 donnent un lissage souple où le passé conserve son influence longtemps.
- ◆ Un choix plus objectif, basé sur la variance de l'erreur de prévision, est possible dans l'ajustement de modèles. L'exemple du modèle autorégressif d'ordre 1 est traité dans [GM90].
- ◆ Les valeurs 0,7 ou 0,8 généralement recommandées.

## 2 Lissage exponentiel double :

*Le lissage exponentiel double* résulte, comme nous l'avons annoncé plus haut, de l'ajustement d'une tendance linéaire au voisinage de  $T$  par la méthode des moindres carrés pondérés par un facteur d'oubli de nature

exponentielle. La prévision  $\hat{Y}_T(h)$  d'horizon  $h$  est donc linéaire par rapport à  $\beta$ ,  $\hat{Y}_T(h|\beta) = \hat{Y}_T(h) + \beta h$ , et les coefficients sont solution de :

$$\min_{a,b} \sum_{k=0}^{\infty} \gamma^k (Y_{T-k} + ak - b)^2.$$

En utilisant les sommes de séries,

$$\sum_{k=0}^{\infty} \gamma^k = \frac{1}{1-\gamma}, \quad \sum_{k=0}^{\infty} k \gamma^k = \frac{\gamma}{(1-\gamma)^2}, \quad \sum_{k=0}^{\infty} k^2 \gamma^k = \frac{\gamma(1+\gamma)}{(1-\gamma)^3},$$

le système obtenu par annulation des dérivées partielles s'écrit :

$$\begin{aligned} (1-\gamma) \sum_{k=0}^{\infty} \gamma^k Y_{T-k} + a \frac{\gamma}{(1-\gamma)^2} - b &= 0 \\ (1-\gamma)^2 \sum_{k=0}^{\infty} k \gamma^k Y_{T-k} + \gamma a \frac{1+\gamma}{(1-\gamma)^3} - b &= 0. \end{aligned}$$

On note  $S_1(T) = \hat{Y}_T$  la série lissée introduite à la section 1 et  $S_2(T)$  la série doublement lissée définie par :

$$\gamma \quad S_2(T) = (1 - \gamma) \sum_{k=0}^{\infty} \gamma^k S_1(T - k) = (1 - \gamma)^2 \sum_{k=0}^{\infty} \gamma^k Y_{T-k} + (1 - \gamma) S_1(T).$$

La solution du système est alors donnée par :

$$\alpha \quad \hat{Y}_T = \frac{1 - \gamma}{1 - \beta} [S_1(T) - \beta S_2(T)], \quad \hat{Y}_T = 2S_1(T) - S_2(T).$$

Les formules de mise à jour de  $S_1(T)$  et  $S_2(T)$  sont immédiates :

$$\begin{aligned} S_1(T+1) &= (1 - \gamma) Y_{T+1} + S_1(T), \\ S_2(T+1) &= (1 - \gamma)^2 Y_{T+1} + S_2(T) + (1 - \gamma) S_1(T). \end{aligned}$$

Elles conduisent ([GM90]) à la mise à jour des coefficients :



$$\begin{bmatrix} \hat{Y}_{T+1} \\ \hat{Y}_{T+1} \end{bmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{bmatrix} \hat{Y}_T \\ \hat{Y}_T \end{bmatrix} + \begin{bmatrix} Y_{T+1} - \hat{Y}_T \\ \gamma \end{bmatrix} \begin{bmatrix} 1 - \lambda \\ (1 - \lambda)^2 \end{bmatrix}.$$

La prévision de  $Y_{T+2}$  obtenue à l'instant  $T + 1$ ,

$$\gamma \quad \hat{Y}_{T+1}(1) = \hat{Y}_{T+1} + \hat{Y}_{T+1} = \hat{Y}_T(2) + 2(1 - \lambda) [Y_{T+1} - \hat{Y}_T(1)]$$

est égale à la celle déjà obtenue à l'instant précédent  $T$  corrigée par un facteur proportionnel à la dernière erreur de prévision observée. Les valeurs initiales  $\alpha_2 = Y_2$  et  $\beta_2 = Y_2 - Y_1$  consistent à prévoir  $Y_3$  à l'aide de la droite définie par les points  $(1, Y_1)$  et  $(2, Y_2)$ .

### 3 Lissage exponentiel généralisé :

*Le lissage exponentiel généralisé* construit, à chaque instant  $T$ , une prévision des valeurs futures de la chronique au vu du passé en ajustant

un modèle de régression linéaire généralisé dont l'ensemble des paramètres, et en particulier la matrice de covariance du bruit, change avec  $T$ . Il ne s'agit donc pas de la résolution récursive en temps d'un modèle global. De plus la matrice des régresseurs doit satisfaire certaines hypothèses afin d'obtenir des formules simples pour la mise à jour des estimateurs.

**On considère à l'instant  $T$  le modèle linéaire généralisé,**

$$\sigma \quad Y(T) = X(T)\theta(T) + \varepsilon(T), \quad E\{\varepsilon(T)\} = 0, \quad V\{\varepsilon(T)\} = \sigma^2 \Gamma(T),$$

où  $Y(T) = [Y_T, \dots, Y_1]$  est le vecteur des observations ordonnées dans le sens inverse du sens habituel,  $\theta(T) = [\theta_1(T), \dots, \theta_p(T)]^t$  est le vecteur des paramètres et  $\Gamma(T)$  est une matrice diagonale avec

$$\text{Diag}\{\Gamma(T)\} = (\sigma_1^2, \dots, \sigma_p^2).$$

La matrice des régresseurs  $X(T)$  est également particulière,

$$X(T) = \begin{pmatrix} x_1(0) & \dots & x_r(0) \\ \vdots & & \vdots \\ x_1(-T+1) & & x_r(-T+1) \end{pmatrix} = \begin{bmatrix} x(0) \\ \vdots \\ x(-T+1) \end{bmatrix},$$

Puisque ses lignes sont définies par une fonction vectorielle  $x(t)$ ,  $t \in \mathbb{Z}$ , satisfaisant la récurrence  $x(t) = Ax(t-1)$  ou  $A$  est une matrice  $r \times r$  régulière. On dit que  $x(t)$ , est un vecteur d'état à transition fixe. Sur la période  $1, \dots, T$ , l'observation  $Y_t$  est représentée comme une version bruitée d'une forme linéaire de l'état  $x(t-T)$ :

$$Y_t = (T) x(t-T) + (T), \quad t = 1, \dots, T.$$

L'état d'origine  $x(0)$ , est systématiquement recalé sur  $Y_t$  et les matrices des régresseurs satisfont

$$X(T+1) = \begin{bmatrix} X(T) \\ x(-T) \end{bmatrix} = \begin{bmatrix} x(0) \\ X(T)_{A^t} \end{bmatrix}.$$

Ainsi, l'espace des moyennes  $\Psi\{X(T)_{A^t}\}$  du sous modèle de dimension  $T$  obtenu à l'étape  $T+1$  est identique au précédent  $\Psi\{X(T)\}$ . Par contre la matrice de variance des erreurs associées est divisée par et les modèles successifs ne sont donc pas "emboîtés". Le recalage de l'origine permet d'obtenir une mise à jour simple de l'estimateur  $\hat{\theta}(T)$ . En effet, si la prévision  $\hat{Y}_T(1)$  est égale à l'observation  $Y_{T+1}$ , l'expression de  $X(T+1)$  ci-dessus montre que l'on a  $\hat{\theta}(T+1) = A^t \hat{\theta}(T)$ . **La solution est en fait :**

$$\hat{\theta}(T+1) = A^t \hat{\theta}(T) + [Y_{T+1} - \hat{Y}_T(1)] M^{-1} x(0),$$

où

$$M = \sum_{k=0}^{\infty} x^{(-k)} x^{(-k)}.$$

Pour cela considérons le problème des moindres carrés pondérés,

$$\min_{\theta} \sum_{k=0}^{\infty} \left[ Y_{T-k} - \sum_{j=1}^{\infty} x_j^{(-k)} \right]^2,$$

avec la convention  $Y_t = 0$  pour  $t \leq 0$ . La solution est donnée par :

$$\theta \hat{Y}(T) = M^{-1} X(\infty) \Gamma(\infty)^{-1} Y(T), \quad M = X(\infty) \Gamma(\infty)^{-1} X(\infty).$$

À l'étape  $T+1$ , seul  $Y(T)$  change et son écriture sous la forme,

$$Y(T+1) = \begin{bmatrix} Y_{T+1} \\ Y(T) \end{bmatrix} = \begin{bmatrix} \hat{Y}_T(1) \\ Y(T) \end{bmatrix} + \begin{bmatrix} Y_{T+1} - \hat{Y}_T(1) \\ 0 \end{bmatrix},$$

conduit au résultat.

Rappelons que la prévision est  $\hat{Y}_T(h) = (T)_x(h)$ . L'initialisation qui conduit aux solutions des moindres carrés ci-dessus est

$$\theta \quad \hat{(\cdot)}_r = M^{-1} X(\infty) \Gamma(\infty)^{-1} Y(\cdot)_r.$$

◆ Une forme approchée consiste à résoudre le problème des moindres carrés pondérés ou non à l'étape  $r$  sans la convention  $Y_t = 0$  pour  $t \leq 0$ ,

$$\theta \quad \hat{(\cdot)}_r = X(\cdot)_r^{-1} \Gamma(\cdot)_r^{-1} Y(\cdot)_r \quad \text{ou} \quad \hat{(\cdot)}_r = X(\cdot)_r^{-1} Y(\cdot)_r$$

◆ La condition  $x(t) = A x(t-1)$  n'est pas très restrictive car on peut construire des systèmes complexes par regroupement de systèmes élémentaires :

$$\begin{bmatrix} x(t) \\ x(t) \end{bmatrix} = \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix} \begin{bmatrix} A 0 \\ 0 A \end{bmatrix} = \begin{bmatrix} x(t-1) \\ x(t-1) \end{bmatrix}$$

Pour un polynôme de degré  $q$  on utilise

$$x_0(t) = 1, \quad x_j(t) = \frac{t(t-1)\dots(t-j+1)}{j!} \quad j = 1, \dots, q,$$

et la matrice  $A$  est triangulaire inférieure avec  $A_{ij} = 1$  si  $i = j$  ou  $i = j + 1$ , 0 sinon. Les fonctions sinusoidales  $\lambda_1 \sin t + \lambda_2 \cos t$  sont associées à la rotation d'angle  $\theta$  et la fonction exponentielle  $e^{at}$  à  $e^{\alpha t}$  avec  $\alpha \geq 0$  pour l'existence de  $M$ .

◆ Un problème pratique important est celui du choix de  $x$ . La nature de la série (tendance et saisonnalité) peut donner des indications.

#### 4 L'approche de Holt et Winters

La mise en œuvre rigoureuse du lissage exponentiel généralisé reste complexe. Sur le plan pratique, Holt et Winters ont proposé des modèles voisins beaucoup plus accessibles.

## 4.1 Modèle de Holt

Posant  $\alpha = \beta(1-\gamma)$  et  $\beta = \frac{\gamma}{1+\gamma}$ , la mise à jour des coefficients du lissage exponentiel double s'écrit :

$$\hat{Y}_{T+1} = Y_{T+1} + (1-\alpha)(\hat{Y}_T + \hat{Y}_T) = Y_{T+1} + (1-\alpha)\hat{Y}_T(1),$$

$$\hat{\beta}_{T+1} = \beta(1-\alpha)\hat{Y}_T + (\hat{Y}_{T+1} - \hat{Y}_T).$$

Le modèle de Holt consiste à appliquer ces relations avec  $0 \leq \alpha \leq 1$  et  $0 \leq \beta \leq 1$ , sans tenir compte de la liaison entre  $\alpha$  et  $\beta$  due à  $\gamma$ . Le modèle gagne en souplesse par l'utilisation des deux constantes  $\alpha$  et  $\beta$ , mais perd sa justification par le critère des moindres carrés. Notons que l'on a :

$$\hat{Y}_{T+1}(1) = \hat{Y}_{T+1} + \hat{Y}_{T+1} = \hat{Y}_T(2) + (1+\beta)[Y_{T+1} - \hat{Y}_T(1)]$$



Les valeurs initiales  $\hat{\beta}_2 = Y_2$  et  $\hat{\alpha}_2 = Y_2 - Y_1$  consistent à utiliser la droite passant par  $(1, Y_1)$  et  $(2, Y_2)$  pour prévoir  $Y_3$  (comme dans le lissage exponentiel double).

## 4.2 Modèle de Holt-Winter additif avec saisonnalité

Le modèle de Holt-Winter additif avec saisonnalité d'ordre  $p$  propose une prévision à l'horizon  $h$ , au vu de  $Y_1, \dots, Y_T$  sous la forme :

$$\hat{Y}_T(h) = \hat{\alpha}_T h + \hat{\beta}_T + \hat{S}_{T+h-p}, \quad h = 1, \dots, p,$$

en utilisant les formules de mise à jour suivantes :

$$\begin{aligned} \hat{\beta}_{T+1} &= \left[ Y_{T+1} - \hat{S}_{T+1-p} \right] + (1 - \beta) \left[ \hat{\beta}_T + \hat{\alpha}_T \right], \\ \hat{\alpha}_{T+1} &= \left[ \hat{\beta}_{T+1} - \hat{\beta}_T \right] + (1 - \alpha) \hat{\alpha}_T, \\ \hat{S}_{T+1} &= \left[ Y_{T+1} - \hat{\beta}_{T+1} \right] + (1 - \beta) \hat{S}_{T+1-p}, \end{aligned}$$

$\beta$  ou  $\delta$ , et sont trois constantes à choisir dans  $[0,1]$ .

◆ Pour initialiser les paramètres  $\hat{\alpha}_p$  et  $\hat{\beta}_p$  de la tendance (l'origine est à l'instant  $p$ !), ainsi que les coefficients saisonniers  $\hat{S}_j$ ,  $j = 1, \dots, p$ , on peut utiliser le critère des moindres carrés sur les points  $(t, Y_t)$ ,  $t = 1, \dots, p + h$  avec  $h \geq 1$  (pour  $h = p$ , on retrouve **Buys-Ballot** voir polycopie), avec ou sans facteur d'oubli.

◆ Plus simplement, on prendra  $\hat{S}_j = Y_j - \hat{\alpha}_p (j - p) - \hat{\beta}_p$ ,  $j = 1, \dots, p$  où  $\hat{\alpha}_p$  et  $\hat{\beta}_p$  sont les paramètres de la droite des moindres carrés ajustée sur les points  $(t, Y_t)$ ,  $t = 1, \dots, p$  :

$$\hat{\alpha}_p = \frac{12}{p(p^2 - 1)} \left[ \sum_{k=1}^p k Y_k - \frac{p(p+1)}{2} \bar{Y}_p \right], \quad \hat{\beta}_p = \bar{Y}_p + \frac{p-1}{2} \hat{\alpha}_p,$$

$$\text{avec } \bar{Y}_p = \frac{1}{p} \sum_{k=1}^p Y_k.$$

### 4.3 Modèle de Holt-Winter multiplicatif avec saisonnalité

Le modèle de Holt-Winter multiplicatif avec saisonnalité d'ordre  $p$  ne consiste pas à **appliquer le précédent sur le logarithme de la série** (La tendance locale serait alors de nature exponentielle), mais à utiliser la formule de prévision :

$$\hat{Y}_T(h) = \left[ \hat{a}_T h + \hat{b}_T \right] \hat{S}_{T+h-p}, h = 1, \dots, p,$$

avec les formules de mise à jour suivantes :

$$\begin{aligned} \hat{a}_{T+1} &= \frac{Y_{T+1}}{\hat{S}_{T+1-p}} + (1 - \alpha) \left[ \hat{a}_T + \hat{b}_T \right] \\ \hat{b}_{T+1} &= \left[ \hat{a}_{T+1} - \hat{a}_T \right] + (1 - \beta) \hat{b}_T \\ \hat{S}_{T+1} &= \frac{Y_{T+1}}{\hat{a}_{T+1}} + (1 - \beta) \hat{S}_{T+1-p} \end{aligned}$$

◆ Pour les valeurs initiales des paramètres, on utilise  $\hat{S}_j = \frac{Y_j}{\hat{\alpha}_p (j - p)^{-\hat{\alpha}_p}}$ ,  $j = 1, \dots, p$  avec les mêmes valeurs pour  $\hat{\alpha}_p$  et  $\hat{\beta}_p$  que précédemment.

◆ Une difficulté, dans l'utilisation de ces modèles, est le choix des constantes.

◆ Dans les écritures retenues, on constatera que les valeurs proches de 1 correspondent ici à un lissage rigide.

## Conclusion :

**Toutes les méthodes que l'on a vu à l'exception de la dernière sont en fait des cas particuliers du système de prévision de Box et Jenkins.**

## Série corrigée des variations saisonnières (Série CVS) :

Les moyennes mobiles permettent de construire des séries désaisonnalisées sans avoir à faire d'hypothèses contraignantes sur la série brute. On se place dans le cadre du modèle additif :

$$Y_t = f_t + S_t + \varepsilon_t, \quad t = 1, \dots, T, \quad \varepsilon_t \text{ i.i.d. } (0, \sigma^2),$$

où le mouvement saisonnier est rigoureusement périodique de période  $p$  et défini par les coefficients saisonniers  $S_j, j = 1, \dots, p$  vérifiant  $\sum_{j=1}^p S_j = 0$ . Appliquons la moyenne mobile d'ordre  $p$  :

$$M_p(Y_t) = M_p(f_t) + M_p(S_t) + M_p(\varepsilon_t) = M_p(f_t) + M_p(\varepsilon_t).$$

En effet la moyenne mobile d'ordre  $p$  est clairement égale à la moyenne arithmétique des valeurs d'une période, elle est donc constante et  $M_p(S_t) = 0$  compte tenu que les coefficients saisonniers "sont centrés" (de moyenne nulle). La tendance  $f_t$  étant une fonction à variation lente,

$\varepsilon$  la série  $M_p(f_t)$  est peu différente de  $f_t$ . Enfin la partie résiduelle  $M_p(\varepsilon_t)$  doit être voisine de 0 puisqu'elle représente la moyenne de  $p$  variables centrées non corrélées. En résumé la série  $M_p(Y_t)$  constitue une estimation de la tendance,  $\hat{f}_t = M_p(Y_t)$ , et l'estimateur est sans biais lorsque la tendance est linéaire. Pour estimer l'effet saisonnier, on retire la tendance ainsi estimée à la série brute,  $\Delta_t = Y_t - M_p(Y_t)$ ,  $t = \lfloor p/2 \rfloor + 1, \dots, T - \lfloor p/2 \rfloor$ . L'estimation définitive est obtenue en accord avec la contrainte  $\sum_{j=1}^p S_j = S = 0$  :

$$\hat{S}_j = S_j - \bar{S} \quad \text{avec} \quad S_j = \frac{1}{p} \sum_{k=1}^p S'_k, \quad j = 1, \dots, p.$$

La série corrigée des variations saisonnières, notée  $Y_t^c$  est définie sur toute la période d'observation,  $t = 1, \dots, T$ , en retranchant à la série brute l'estimation de la composante saisonnière :

$$Y_{ij}^c = Y_{ij} - \hat{S}_j \quad j = 1, \dots, p, \quad i = 1, \dots, n.$$

## Moyennes mobiles :

Les moyennes mobiles utilisées dans les opérations de désaisonnalisation sont symétriques et d'ordre fini :

$$Y_t = M(Y_t) = \sum_{j=-k}^k \gamma_{-j} Y_{t-j}, \quad \gamma_{-j} = \gamma_j, \quad j = 1, \dots, k, \quad t = k+1, \dots, T-k.$$

Les coefficients cependant peuvent être négatifs (*filtre*) et l'étude est très liée à la notion d'équation de récurrence linéaires. Le lissage exponentiel traité au paragraphe précédent fait intervenir des moyennes mobiles

d'ordre infini,  $\hat{Y}_T(1) = (1 - \alpha) \sum_{k=0}^{\infty} \alpha^k Y_{T-k}$ .

### 1 Définitions et propriétés :

Il est commode, dans l'étude des propriétés générales des moyennes mobiles, de considérer des séries doublement infinies  $Y = \{Y_t; t \in \mathbb{Z}\}$  afin d'éliminer les problèmes de bord. Introduisons l'opérateur de retard (ou de décalage arrière)  $B$  (backward) ou  $L$  (lag), qui à la série décalée

dans le passé d'une unité de temps  $BY = \{Y_{t-1}; t \in \mathbb{Z}\}$ . On utilisera souvent l'élément générique  $(BY)_t$ , noté  $B(Y_t)$  ou  $BY_t$ , pour désigner la série  $BY$ .

Ainsi une moyenne mobile d'ordre  $m$  est une transformation formée d'une combinaison linéaire réelle des puissances de l'opérateur de retard,

$$M = \sum_{j=-l}^k a_j B^j, \quad \sum_{j=-l}^k a_j \neq 0, \quad m = l + k + 1.$$

Elle est dite centrée lorsque  $l = k$  et symétrique si  $a_{-j} = a_j, j = 1, \dots, k$ .

Dans les méthodes de lissage on impose que **la somme des coefficients soit égale à 1**. Cette condition est naturelle lorsque les coefficients sont positifs ou nuls puisqu'elle signifie que  $M$  réalise une moyenne pondérée au sens usuelle du terme. Lorsqu'ils sont de signe quelconque, **elle équivaut à ce que  $M$  conserve les constantes. Si de plus elle est symétrique, elle conserve les polynômes de degré 1.**



♦ Une moyenne mobile est un opérateur linéaire:  
 $\alpha \quad \alpha_M (Y_t) = M(Y_t), \quad M(X_t + Y_t) = M(X_t) + M(Y_t),$   
 qui permute avec l'opérateur de retard,  $M(BY_t) = B(MY_t).$

♦ La composition (ou produit) des moyennes mobiles,  
 $NM(Y_t) = N\{M(Y_t)\},$  est très utile dans la construction de moyennes complexes à partir d'éléments simples, comme par exemple les **moyennes arithmétiques**. L'opération est clairement associative et commutative. Elle préserve le caractère centré, la symétrie ainsi que la contrainte de sommation à 1 des coefficients. La somme  $N + M$ , et la multiplication par un scalaire,  $\alpha M$ , ne sont pas utilisées en tant que telles, d'ailleurs elles ne préservent pas la contrainte ci-dessus. Elles interviennent dans la définition à partir de l'opérateur retard.

♦ En ce sens l'ensemble des moyennes mobiles d'abord fini est l'espace vectoriel réel de dimension infinie engendré par les puissances

$\{B^j; j \in \mathbb{Z}\}$  de cet opérateur. Les moyennes mobiles centrées forment un sous-espace vectoriel stable par composition. De même les moyennes mobiles symétrique forment un sous-espace vectoriel du précédent également stable par composition. Cet espace est engendré par  $\{(B + F)^j; j \in \mathbb{Z}\}$  où  $F = B^{-1}$  est l'opérateur de décalage avant (forward). Il l'est également par les puissances de l'opérateur de différence symétrique  $\{(B - 2I + F)^j; j \in \mathbb{Z}\}$ . Notons que ce dernier s'écrit  $B - 2I + F = \Delta\Lambda$  où  $\Delta = I - B$  est l'opérateur de différence arrière et  $\Lambda$  est l'opérateur de différence avant. Par la suite on se restreindra à l'utilisation des opérateurs  $B$  et  $\Delta$ .

## 2 Noyau et invariants :

Les composantes éliminées (noyau) et celles conservées (invariants) par une moyenne mobile forment des sous-espaces vectoriels de l'ensemble

des chroniques et se décrivent très simplement à partir des solutions d'équations de récurrence linéaires.

Considérons l'équation de récurrence linéaire d'ordre  $k$  à coefficients réels :

$$\alpha \quad x_t + x_{t-1} + \dots + x_{t-k} = 0, \quad t \in \mathbb{N}, \quad x_0, \dots, x_{k-1} \neq 0,$$

à laquelle on associe le polynôme caractéristique,

$$\alpha \quad P(z) = z^k + z^{k-1} + \dots + z + 1.$$

Si  $x$  est racine réelle d'ordre  $q$  de l'équation caractéristique  $P(z) = 0$ ,

$$x_t = (c_0 + c_1 t + \dots + c_{q-1} t^{q-1}) x^t, \quad t \in \mathbb{N},$$

est solution de l'équation de récurrence pour tout choix des constantes réelles  $c_0, \dots, c_{q-1}$ . Si  $e^{\lambda i}$  est racine complexe d'ordre  $s$  de  $P(z) = 0$ , il en est de même de la racine conjuguée  $e^{-\lambda i}$  car les coefficients sont réels et

$$x_t = \left[ (a_0 + a_1 t + \dots + a_{s-1} t^{s-1}) \cos t + (b_0 + \dots + b_{s-1} t^{s-1}) \sin t \right] t, \quad t \in \mathbb{N},$$

est solution de l'équation de récurrence pour tout choix de constantes réelles  $a_i, b_i, i = 0, \dots, s-1$ .

◆ L'ensemble des solutions ainsi associées aux racines de l'équation caractéristique engendrent un espace vectoriel réel de dimension  $k$ . Lorsque toutes les racines de l'équation caractéristique sont sur le cercle unité, le polynôme est symétrique,  $a_j = a_{k-j}, j = 0, \dots, k$ , et les solutions  $x_t$  ne présentent plus de formes explosives dues aux facteurs  $x^t$  ou  $t$ .

◆ Le noyau d'une moyenne mobile  $M$ , noté  $Ker(M)$ , est l'ensemble des chroniques  $Y$  annulées par cette moyenne. La condition  $Y \in Ker(M)$ , s'écrit :

$$M(Y_t) = \sum_{j=-l}^k Y_{t-j} = 0, \quad t \in \mathbb{Z}.$$

Ce sont donc les chroniques solution de l'équation de récurrence linéaire d'ordre  $m-1$ :

$$\gamma \quad -l Y_t + -l+1 Y_{t-1} + \dots + k Y_{t-l-k} = 0, \quad t \in \mathbb{Z},$$

construites à partir des racines du polynôme caractéristique associé,

$$g(z) = -l z^{k+l} + -l+1 z^{k+l-1} + \dots + k z + k.$$

On a  $\dim \{Ker(M)\} = m-1$  et  $Ker(NM) \supseteq Ker(N) + Ker(M)$  avec égalité si et seulement si  $Ker(N) \cap Ker(M) = \{0\}$ . L'ensemble des polynômes de degré au plus  $q$  est le noyau de  $\Delta^{q+1}$ . En d'autres termes  $f_t$  est un polynôme de degré au plus  $q$  si et seulement si  $\Delta^{q+1}(f_t) = 0$ .

◆ L'élimination d'un mouvement saisonnier d'ordre  $p$  se traduit par la factorisation de  $(1 - z^p)$  dans le polynôme  $g(z)$ . Par contre lorsque sa moyenne est nulle sur la période (prise en compte dans la composante fondamentale) il suffit d'avoir :

$$(1 - z^p) g(z) = (1 - z^p) r(z) \Leftrightarrow g(z) = (1 + z + \dots + z^{p-1}) r(z).$$

Ceci justifie l'utilisation des **moyennes arithmétique**  $\left( r(z) \equiv 1 \right)$  qui de plus conservent les polynômes de degré 1 car elles sont symétrique et la somme de leurs coefficients est égale à 1.

♦ Une chronique  $Y$  est *invariante* par une moyenne mobile  $M$  lorsque

$$M(Y_t) = \sum_{j=-l}^k Y_{t-j} = Y_t, \quad t \in \mathbb{Z}.$$

Il s'agit alors des chroniques solution de l'équation de récurrence linéaire d'ordre  $m-1$  :

$$Y_t + \binom{m-1}{1} Y_{t-1} + \dots + \binom{m-1}{l-1} Y_{t-l+1} + \dots + Y_{t-l} = 0, \quad t \in \mathbb{Z}.$$

Les solutions forment un sous-espace vectoriel  $I(M)$  de dimension  $m-1$  obtenues à partir des solutions de l'équation  $g(z)z^k = 0$ . Par composition on a l'inclusion  $I(M) \cap I(N) \subseteq I(MN)$ . La conservation des polynômes constitue un cas particulier intéressant. La moyenne mobile conserve les polynômes de degré au plus  $q$  si et seulement si :

$z = 1$  est racine d'ordre  $q + 1$  de l'équation  $g(z) - z^k = 0$ .

### 3 Moyennes de Spencer d'ordre 15 :

La moyenne de Spencer d'ordre 15 est définie par :

$$M = \frac{1}{320} (-3, -6, -5, 3, 21, 46, 67, \underline{74}) = \frac{1}{320} [4] \circ [5] (-3, 3, \underline{4}).$$

Dans la première notation, on donne les coefficients du filtre précédant le terme central indiqué en caractère souligné pour préciser qu'il s'agit d'une moyenne symétrique. Dans la deuxième écriture, on fait apparaître que la moyenne est obtenue par composition de moyennes simples,

$$\frac{1}{5} [5] = \frac{1}{5} (1, 1, 1, 1, 1), \quad \frac{1}{16} [4] = \frac{1}{16} (1, 2, 3, \underline{4}), \quad \frac{1}{4} (-3, 3, \underline{4}).$$

Les deux moyennes d'ordre 4 sont choisies de la forme

$$\frac{1}{4} (1, 1, 1, 1, 0), \quad \frac{1}{4} (0, 1, 1, 1, 1),$$

De sorte à obtenir la symétrie par composition. Cette moyenne élimine les saisonnalités (centrées) de période 4 dont l'amplitude varie linéairement,  $\frac{1}{16}[4]$ , celle de période 5,  $\frac{1}{5}[5]$  et conserve les polynômes de degré 3 (effet conjugué des moyennes précédentes et de la moyenne  $\frac{1}{4}(-3,3,4)$ ). Ce dernier point peut être vérifié en s'assurant que  $320[g(z) - z^7]$  ainsi que ses 3 premières dérivées s'annulent en  $z = 1$ . S'agissant de l'élimination des saisonnalités, le fait qu'elles soient centrées se traduit par :

$$(1-z)^3 g(z) = (1-z^4)^2 (1-z^5) r(z).$$

**Voir TD 2 pour les démonstrations (transformée d'un Bruit ainsi que l'exemple de moyenne arithmétique), les explications et les applications.**



## Estimation des composantes d'une série

- **Filtres linéaires : moyennes mobiles centrées**
- **Cas sans saisonnalité :**
  - **Estimation de la tendance**
- **Cas avec saisonnalité :**
  - **Estimation de la tendance**
  - **Estimation de la composante saisonnière**
    - . **modèle additif**
    - . **modèle multiplicatif**

# Les filtres linéaires

- Décomposition d'une chronique en ses diverses composantes et de prévisions de ces composantes
- Transformation d'une chronique  $X_t$  en une autre chronique  $Y_t$



$$y_t = \sum_{i=1}^s a_i \cdot x_{t-r+i}$$

- $a_i$  : coefficients de pondération
- $r$  : décalage temporel
- $s$  : nombre de termes consécutifs de la chronique initiale nécessaires pour calculer un terme de la nouvelle chronique.

## Moyennes mobiles centrées : cas sans saisonnalité

### Estimation de la composante tendancielle (linéaire)

→ pour faire **apparaître plus clairement la tendance**, il faut atténuer la composante accidentelle.

- on appelle **moyenne mobile centrée** de longueur impaire  $(2k + 1)$  à l'instant  $t$  la valeur moyenne  $\tilde{x}_t$  des observations  $x_{t-k}, x_{t-k+1}, \dots, x_t, x_{t+1}, \dots, x_{t+k}$  :

$$\tilde{x}_t = (x_{t-k} + \dots + x_{t-1} + x_t + x_{t+1} + \dots + x_{t+k}) / (2k + 1)$$

$$\tilde{x}_t = \frac{1}{2k + 1} \sum_{i=-k}^{+k} x_{t+i}$$

- on appelle moyenne mobile centrée de longueur paire  $(2k)$  à l'instant  $t$  la valeur moyenne  $\tilde{x}_t$  des observations  $x_{t-k}, x_{t-k+1}, \dots, x_t, x_{t+1}, \dots, x_{t+k}$ , la première et la dernière étant pondérées par 0.5 :

$$\tilde{x}_t = (0.5 x_{t-k} + \dots + x_{t-1} + x_t + x_{t+1} + \dots + 0.5 x_{t+k}) / (2k)$$

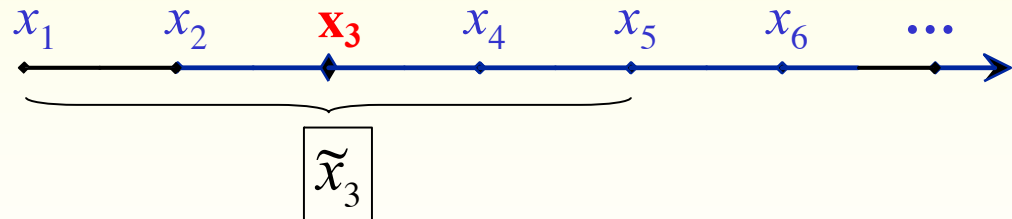
$$\tilde{x}_t = \frac{1}{2k} \left( \frac{1}{2} x_{t-k} + \sum_{i=-(k-1)}^{+k-1} x_{t+i} + \frac{1}{2} x_{t+k} \right)$$

## Exemple : moyennes mobiles centrées

La première valeur d'une moyenne mobile de longueur 4 ( $= 2 \times 2$ ) ou 5 ( $= 2 \times 2 + 1$ ) que l'on peut calculer, est à l'instant  $t = 3$ , puisque la première observation connue est  $x_1$  :

$$(l = 4) \quad \tilde{x}_3 = (0.5 x_1 + x_2 + x_3 + x_4 + 0.5 x_5) / 4$$

$$(l = 5) \quad \tilde{x}_3 = (x_1 + x_2 + x_3 + x_4 + x_5) / 5$$



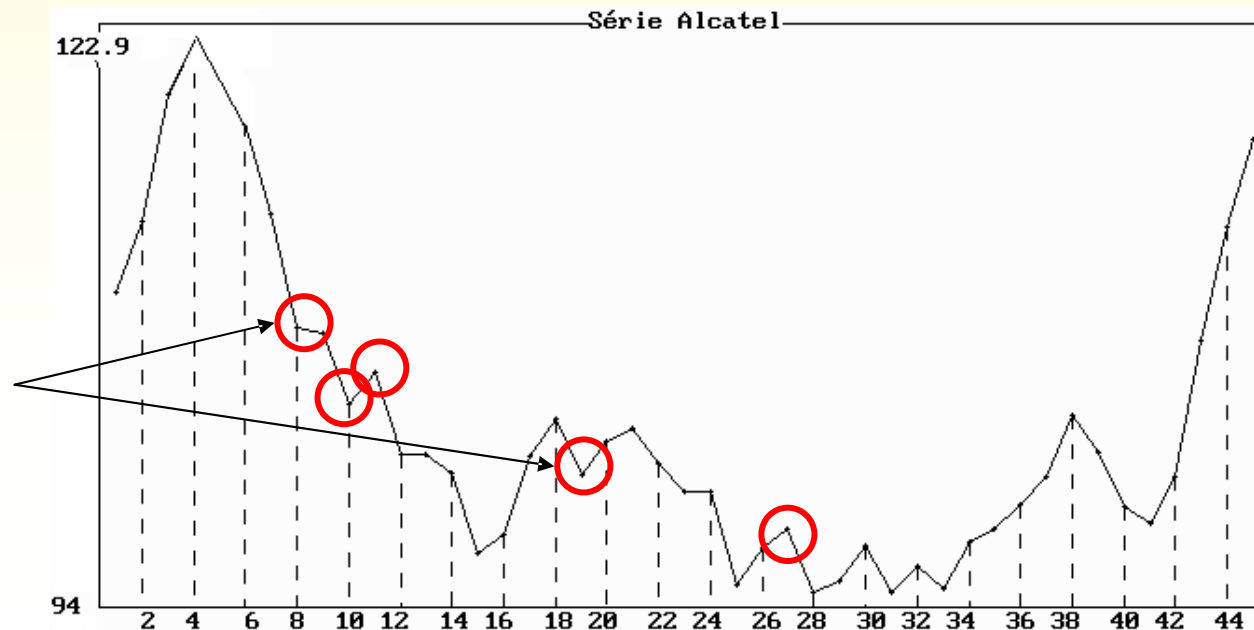
**Remarque :** l'avantage des moyennes mobiles est d'atténuer la composante accidentelle tout en conservant les tendances linéaires : la série est dite « lissée », et est d'autant plus lissée que la longueur de la moyenne mobile est élevée.

# Exemple : moyennes mobiles centrées

1	109.500	10	103.750	19	100.100	28	94.000	37	99.950
2	113.200	11	105.400	20	101.800	29	94.600	38	103.150
3	119.700	12	101.175	21	102.450	30	96.425	39	101.250
4	122.350	13	101.100	22	100.600	31	94.025	40	98.450
5	122.900	14	100.150	23	99.200	32	95.350	41	97.550
6	118.250	15	96.050	24	99.200	33	94.175	42	100.000
7	113.550	16	96.950	25	94.375	34	96.600	43	107.050
8	107.700	17	101.000	26	96.350	35	97.250	44	112.900
9	107.400	18	103.000	27	97.250	36	98.500	45	117.400

Cours du titre Alcatel du 4 janvier 1999  
au 5 mars 1999

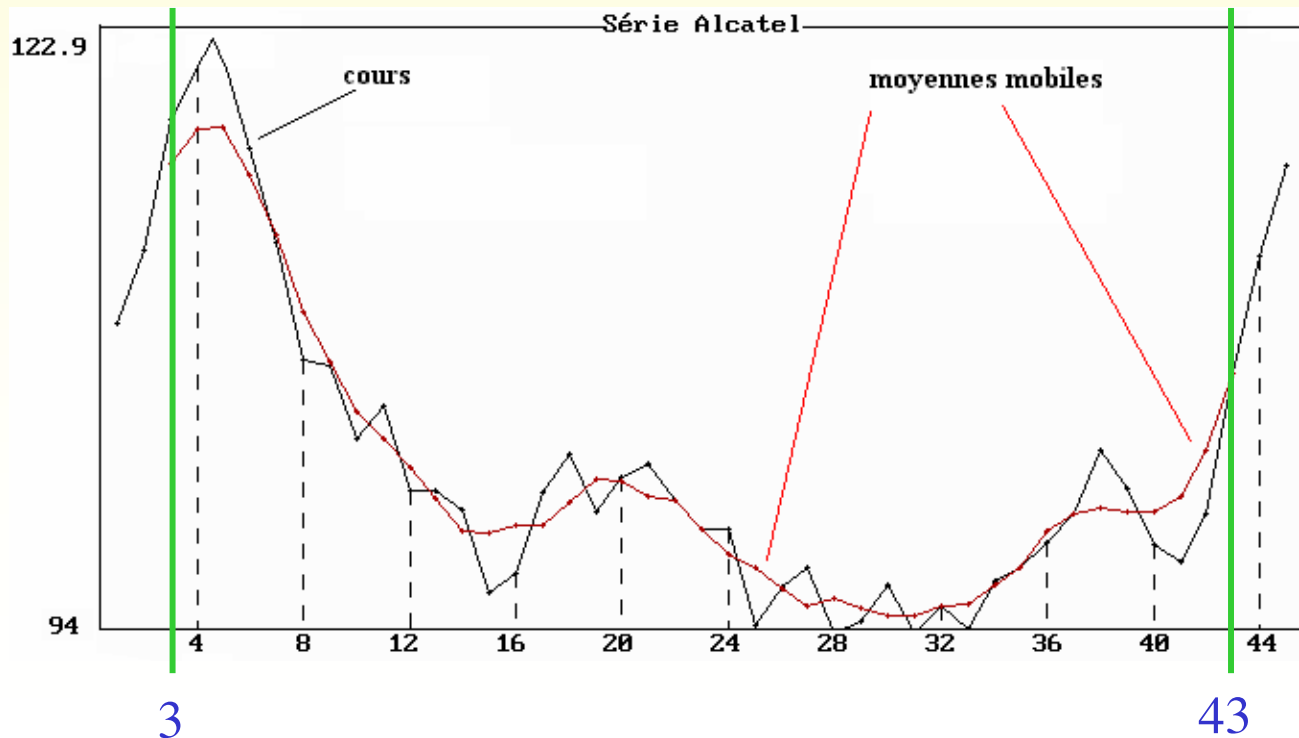
variations  
accidentelles



# Exemple : moyennes mobiles centrées<sub>s</sub>

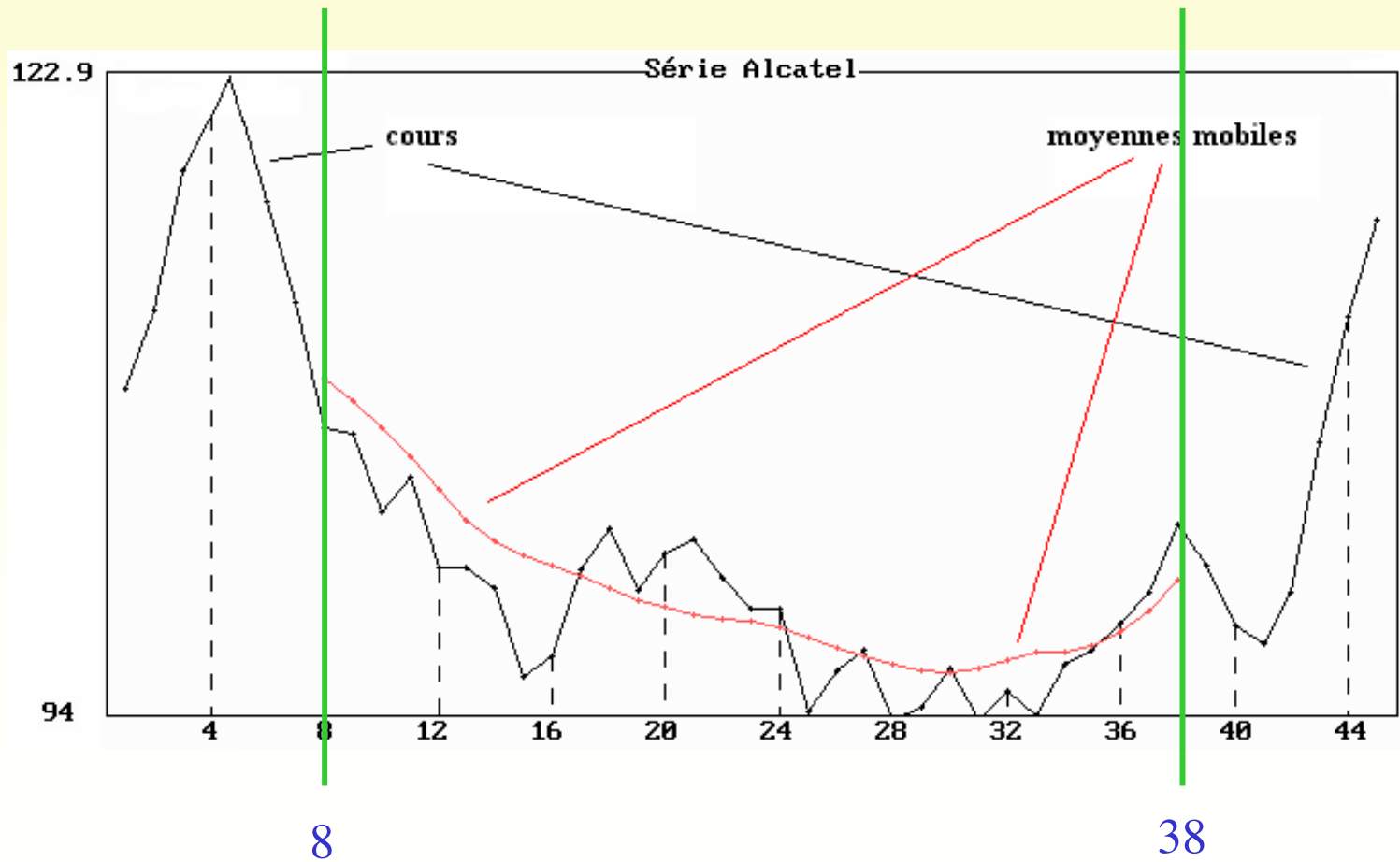
Moyennes mobiles de longueur 5 (extrait).

<i>t</i>	<i>cours</i> (€)	<i>moyenne mobile</i>	<i>t</i>	<i>cours</i> (€)	<i>moyenne mobile</i>
1	109.50000		42	100.00000	103.19000
2	113.20000		43	107.05000	106.98000
3	119.70000	117.53000	44	112.90000	
4	122.35000	119.28000	45	117.40000	



# Exemple : moyennes mobiles centrées

Moyennes mobiles de longueur 14.



## Moyennes mobiles centrées : cas avec saisonnalité

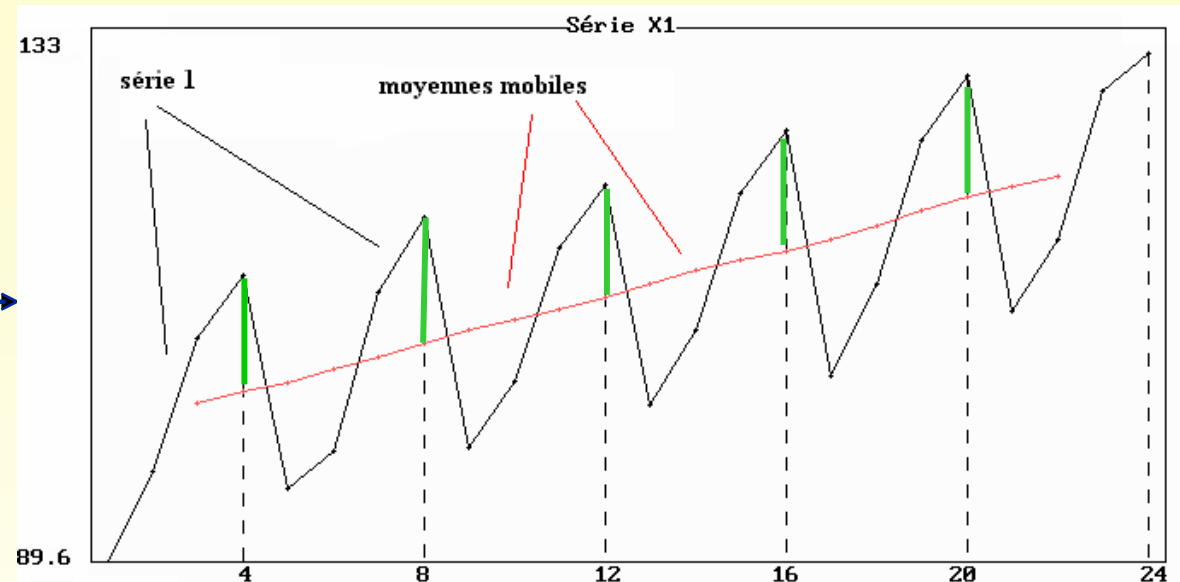
### Principe :

- Calcul des moyennes mobiles en choisissant comme longueur la période  $p$  des variations saisonnières (et plus généralement si leur longueur est un multiple de la période), de façon à les faire disparaître.
- Si la moyenne mobile choisie est de longueur différente, les variations saisonnières ne sont pas toujours éliminées.
- Les moyennes mobiles permettent d'atténuer les variations accidentelles.

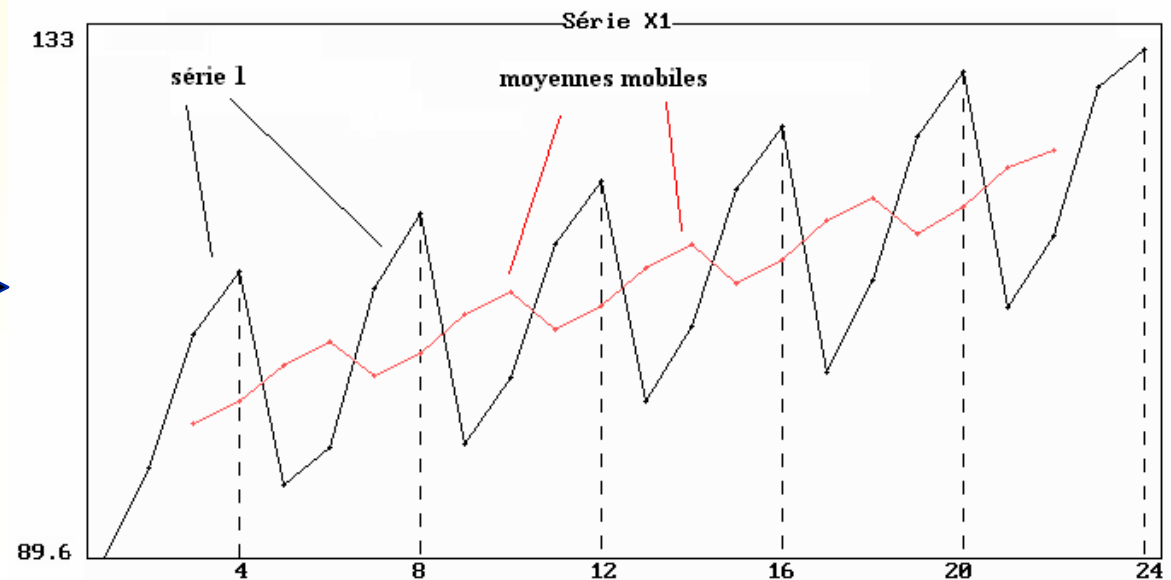


# Moyennes mobiles centrées : cas avec saisonnalité

Moyennes mobiles avec :  
**longueur = période**  
Ex :  $l = p = 4$



Moyennes mobiles avec :  
**longueur  $\neq$  période**  
Ex :  $l = 5 \neq p = 4$



# Cas 1 : Modèle additif – coefficients saisonniers

		$j = 1$	$j = 2$	$j = 3$	$j = 4$	
Ex : années	Saison 1	$i = 1$	$\mathbf{x}_{1,1}$	$\mathbf{x}_{1,2}$	$\mathbf{x}_{1,3}$	$\mathbf{x}_{1,4}$
	Saison 2	$i = 2$	$\mathbf{x}_{2,1}$	$\mathbf{x}_{2,2}$	$\mathbf{x}_{2,3}$	$\mathbf{x}_{2,4}$
	Saison 3	$i = 3$	$\mathbf{x}_{3,1}$	$\mathbf{x}_{3,2}$	$\mathbf{x}_{3,3}$	$\mathbf{x}_{3,4}$
	...					
		Période 1	Période 2	Période 3	Période 4	Ex : mois

$$\forall t = 1, \dots, T \quad \mathbf{x}_t = \mathbf{c}_t + \mathbf{v}_t + \boldsymbol{\varepsilon}_t$$

pour  $n$  saisons et  $p$  périodes :

$$\forall i = 1, \dots, n$$

$$\forall j = 1, \dots, p$$

$$\mathbf{x}_{i,j} = \mathbf{c}_{i,j} + \mathbf{v}_j + \boldsymbol{\varepsilon}_{i,j}$$

Variation saisonnière  
= coefficients saisonniers

$$\mathbf{x}_{i,j} - \mathbf{c}_{i,j} = \mathbf{v}_j + \boldsymbol{\varepsilon}_{i,j}$$

Pour une même période  $j$ , la différence entre l'observation et la tendance est à peu près constante et égale à  $\mathbf{v}_j$

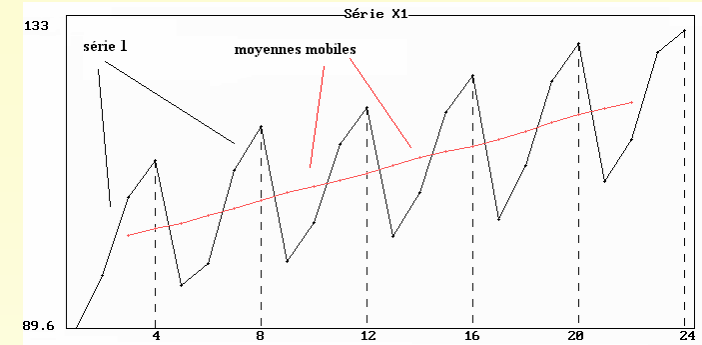
$$\mathbf{x}_{i,j} - (\mathbf{c}_{i,j} + \boldsymbol{\varepsilon}_{i,j}) = \mathbf{v}_j$$

$$\mathbf{x}_{i,j} - \text{mm}_{i,j} \approx \mathbf{v}_j$$

# Exemple

- Moyennes mobiles de longueur 4

	1 trimestre	2 trimestre	3 trimestre	4 trimestre
An 1			103.39678	104.44080
An 2	105.14860	106.25917	107.31233	108.46116
An 3	109.65950	110.47448	111.27404	112.28937
An 4	113.42748	114.58360	115.45379	116.22236
An 5	117.24943	118.38337	119.64839	120.80691
An 6	121.79719	122.54573		



$$v'_j = \frac{1}{n} \sum_{i=1}^n (x_{i,j} - mm_{i,j})$$

- Différences entre les observations et les moyennes mobiles

	1 trimestre	2 trimestre	3 trimestre	4 trimestre
An 1			5.50932	9.71580
An 2	-8.94393	-6.86050	5.45047	10.72334
An 3	-10.05746	-5.28258	5.28226	9.62153
An 4	-10.15538	-4.93910	5.75471	10.28534
An 5	-11.61263	-4.95497	5.99271	10.33979
An 6	-10.67929	-5.33023		
	$v_1 = -10.2897$	$v_2 = -5.4735$	$v_3 = 5.5979$	$v_4 = 10.1371$
	$v_1 = -10.2827$	$v_2 = -5.4664$	$v_3 = 5.6049$	$v_4 = 10.1442$

Pour le principe de conservation des aires

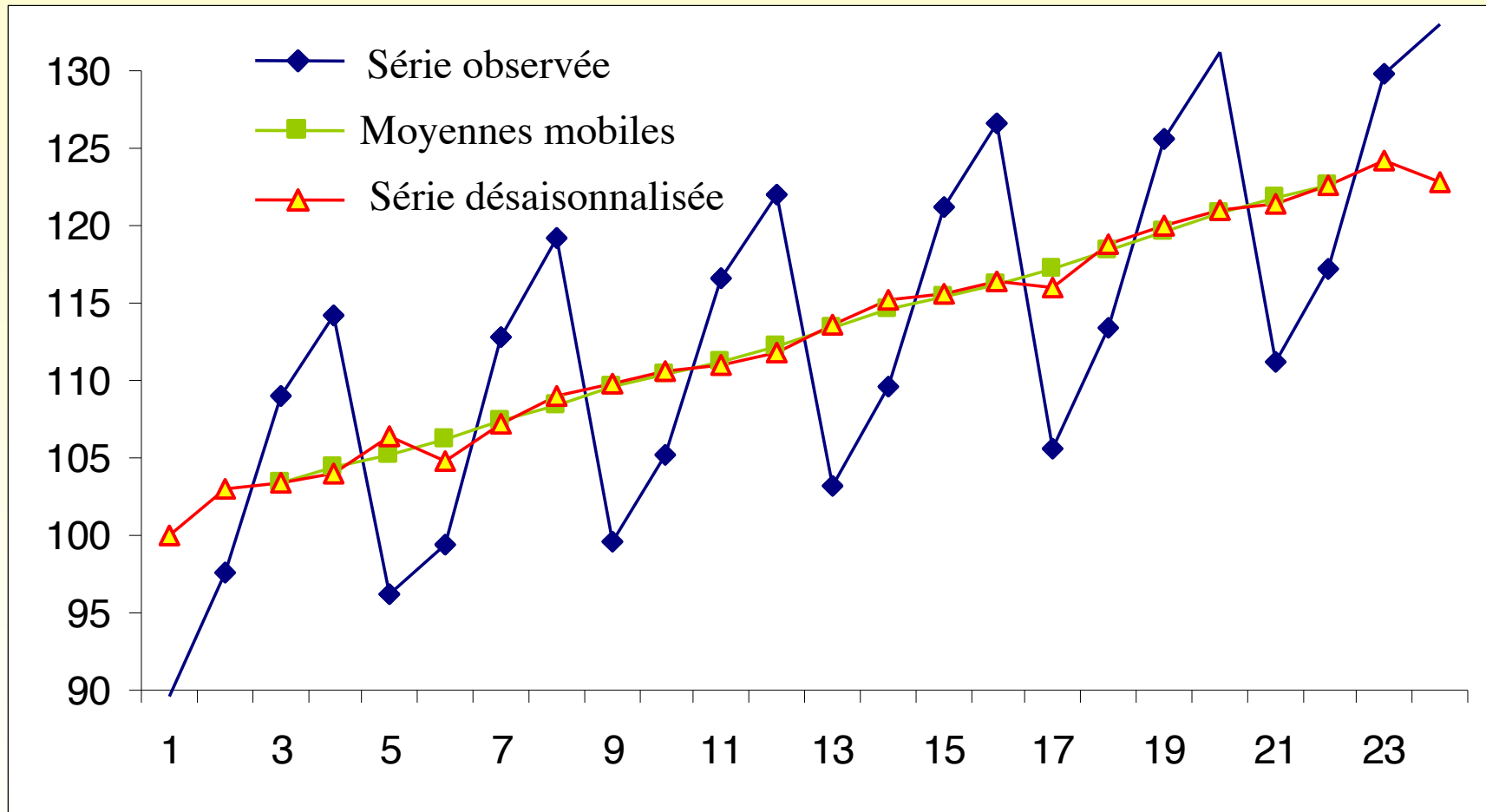
$$v_{moy} = \frac{1}{p} (v'_1 + v'_2 + \dots + v'_p)$$

$$v_j = v'_j - v_{moy}$$

$$v_{moy} = -0.00704$$

Pour que :  $\sum_{j=1}^p v_j = 0$

# Exemple



Observation corrigée d'un modèle additif :  $x'_{i,j} = x_{i,j} - v_j$

Série désaisonnalisée (série des observations corrigées) caractérise à la fois la tendance et la variation accidentelle.

# Résumé

## Règle de calcul des estimations des coefficients saisonniers du modèle additif

- Calcul des différences entre les observations  $x_{i,j}$  et les moyennes mobiles  $mm_{i,j}$  ;
- Calcul de la moyenne ou la médiane  $v'_j$  des différences de chaque colonne du tableau ;
- Calcul de la moyenne  $v_{moy}$  de ces valeurs  $v'_j$  ;
- Obtention des estimations  $v_j$  en centrant les valeurs  $v'_j$  :  $v_j = v'_j - v_{moy}$
- Calcul des observations corrigées par :  $x'_{i,j} = x_{i,j} - v_j$

## Cas 2 : Modèle multiplicatif – coefficients saisonniers

		$j = 1$	$j = 2$	$j = 3$	$j = 4$		
Ex : années	Saison 1	$i = 1$	$x_{1,1}$	$x_{1,2}$	$x_{1,3}$	$x_{1,4}$	
	Saison 2	$i = 2$	$x_{2,1}$	$x_{2,2}$	$x_{2,3}$	$x_{2,4}$	
	Saison 3	$i = 3$	$x_{3,1}$	$x_{3,2}$	$x_{3,3}$	$x_{3,4}$	
	...						
		Période 1	Période 2	Période 3	Période 4	Ex : mois	

$$\forall t = 1, \dots, T$$

$$x_t = c_t (1 + v_t) + \varepsilon_t$$

$$\forall i = 1, \dots, n$$

$$\forall j = 1, \dots, p$$

$$x_{i,j} = c_{i,j} (1 + v_j) + \varepsilon_{i,j}$$

$$x_{i,j} - c_{i,j} = c_{i,j} \cdot v_j + \varepsilon_{i,j}$$

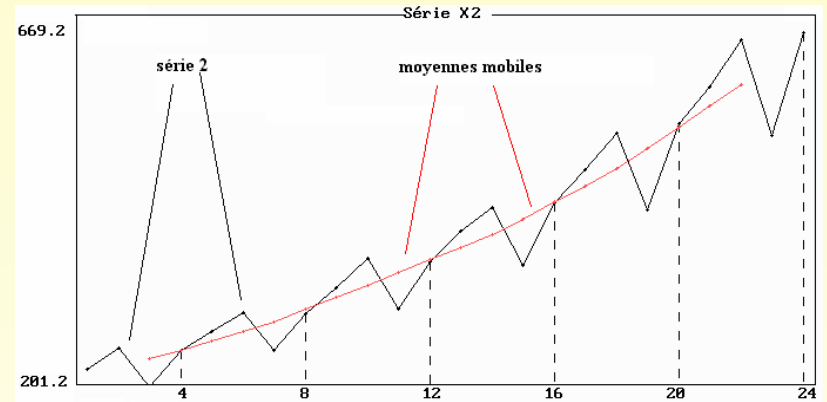
→ Pour une même période  $j$ , la différence entre l'observation et la tendance est proportionnelle à la tendance

- Lorsque cette tendance est croissante, la différence augmente, et lorsqu'elle est décroissante, elle diminue.
- Les différences permettent ainsi de déterminer si la série chronologique étudiée suit un modèle multiplicatif.

# Exemple

- Moyennes mobiles de longueur 4

	1 <sup>er</sup> trimestre	2 <sup>e</sup> trimestre	3 <sup>e</sup> trimestre	4 <sup>e</sup> trimestre
Année 1			238.210	250.322
Année 2	262.140	274.284	287.670	303.773
Année 3	319.494	334.743	352.509	370.167
Année 4	385.759	402.895	422.986	445.525
Année 5	467.090	489.404	516.167	545.210
Année 6	572.761	599.955		



- Rapports des observations aux moyennes mobiles

	1 <sup>er</sup> trimestre	2 <sup>e</sup> trimestre	3 <sup>e</sup> trimestre	4 <sup>e</sup> trimestre
Année 1			0.84481	0.99449
Année 2	1.04670	1.09435	0.86524	0.98244
Année 3	1.03904	1.10952	0.86078	0.98848
Année 4	1.05411	1.08712	0.85482	0.99847
Année 5	1.04566	1.09629	0.84385	1.00761
Année 6	1.04407	1.09890		
	$V'_1 = 1.0459$	$V'_2 = 1.0972$	$V'_3 = 0.8539$	$V'_4 = 0.9943$
	$V_1 = 1.0482$	$V_2 = 1.0996$	$V_3 = 0.8557$	$V_4 = 0.9964$

Les rapports  $x_{i,j} / mm_{i,j}$  sont donc des approximations des termes ( $V_j = 1 + v_j$ ) appelés coefficients saisonniers.

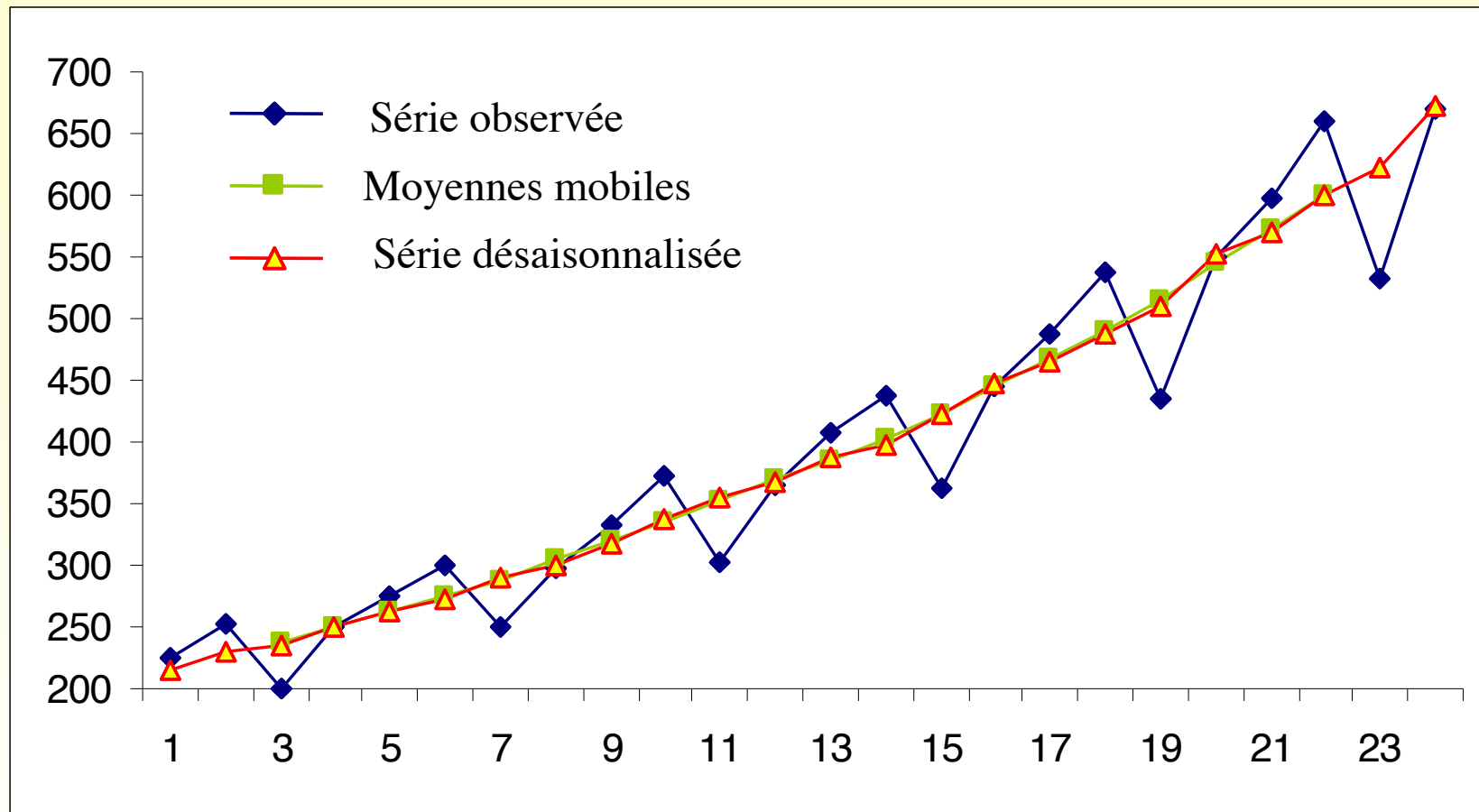
$$V'_j = \frac{1}{n} \sum_{i=1}^n (x_{i,j} / mm_{i,j})$$

$$V_{moy} = \frac{1}{p} (V'_1 + V'_2 + \dots + V'_p)$$

$$V_j = V'_j / V_{moy}$$

$V_{moy} = 0.9978$

# Exemple



Observation corrigée d'un modèle multiplicatif :  $x'_{i,j} = x_{i,j} / V_j$

Série désaisonnalisée (série des observations corrigées) caractérise à la fois la tendance et la variation accidentelle.



# Résumé

## Règle de calcul des estimations des coefficients saisonniers du modèle multiplicatif

- Calcul des rapports des observations  $x_{i,j}$  aux moyennes mobiles  $mm_{i,j}$  ;
- Calcul de la moyenne ou la médiane  $V'_j$  des rapports de chaque colonne du tableau ;
- Calcul de la moyenne  $V_{moy}$  de ces valeurs  $V'_j$  ;
- Obtention des estimations  $V_j$  en réduisant les valeurs  $V'_j$  :  $V_j = V'_j / V_{moy}$
- Calcul des observations corrigées par :  $x'_{i,j} = x_{i,j} / V_j$

# Validation des prévisions

Validation d'une prévision



# Qualité des prévisions

- **Ecart algébrique moyen (biais)** : cette mesure donne la tendance d'un modèle à fournir des prévisions supérieures ou inférieures aux réalisations.

$$M.E = \frac{1}{T} \sum_{t=1}^T (x_t - \tilde{x}_t)$$

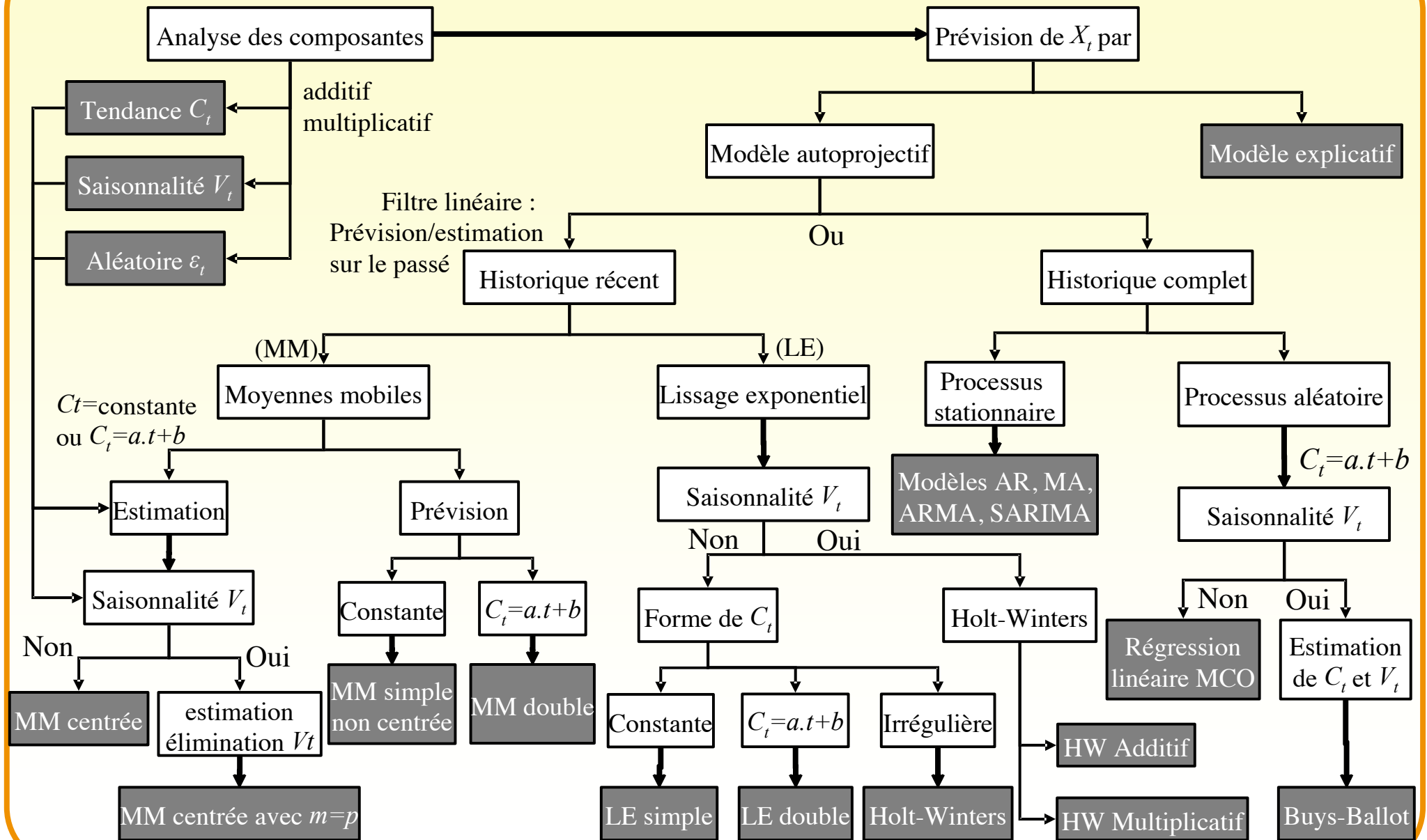
- **Ecart absolu moyen** : cette mesure considère l'importance plutôt que le sens des erreurs de prévision.

$$E.A.M = \frac{1}{T} \sum_{t=1}^T |x_t - \tilde{x}_t|$$

- **Ecart carré moyen** : déceler une méthode qui fournit plus souvent que les autres des écarts de prévisions importants.

$$M.S.E = \frac{1}{T} \sum_{t=1}^T (x_t - \tilde{x}_t)^2$$

# Conclusion sur les méthodes



# Concepts des séries temporelles

## 1 Variables aléatoires :

Soit  $(\Omega, M, P)$  un espace de probabilité, où  $\Omega$  est l'espace des événements,  $M$  est une tribu adaptée à  $\Omega$  (c'est l'ensemble qui contient les combinaisons possibles d'événements) et  $P$  est une mesure de probabilité définie sur  $M$ .

### Définition :

Une variable aléatoire réelle (*v.a.r*) est une fonction à valeurs réelles  $y: \Omega \rightarrow \mathbb{R}$  telle que tout réel  $c$ ,  $A_c = \left\{ \omega \in \Omega \mid y(\omega) \leq c \right\} \in M$ .

En d'autres termes,  $A_c$  est un événement dont la probabilité est définie en terme de  $P$ . La fonction  $F: \mathbb{R} \rightarrow [0,1]$  définie par  $F(c) = P(A_c)$  est la *fonction de distribution* de  $y$ .

## 2 Processus stochastiques :

Soit  $T$  un ensemble d'indexation dénombrable contenu dans l'ensemble des entiers naturels ou dans celui des entiers relatifs.

### Définition :

Un processus stochastique (discret) est une fonction à valeurs réelles

$$y: T \cdot \Omega \rightarrow \mathbb{R},$$

Telle que pour tout  $t \in T$  donné,  $y_t(\cdot)$  soit une variable aléatoire.

En d'autres termes, un processus stochastique est une suite ordonnée de variables aléatoires  $\{y_t(\cdot), \omega \in \Omega, t \in T\}$ , telle que tout  $t \in T$ ,  $y_t$  soit une variable aléatoire sur  $\Omega$  et que pour tout  $\omega \in \Omega$  soit une réalisation du processus stochastique sur l'ensemble d'indexation  $T$ .

$\omega$	$\dots$	$0$	$\omega$	$\dots$	$J$	$\omega$	
$\omega t_0$	$\dots$	$y_{t_0} \begin{pmatrix} \phantom{0} \\ 0 \end{pmatrix}$	$\omega$	$\dots$	$y_{t_0} \begin{pmatrix} \phantom{0} \\ j \end{pmatrix}$	$\omega$	$y_{t_0} \begin{pmatrix} \phantom{0} \\ m \end{pmatrix}$
$\omega t_i$	$\dots$	$y_{t_i} \begin{pmatrix} \phantom{0} \\ 0 \end{pmatrix}$	$\omega$	$\dots$	$y_{t_i} \begin{pmatrix} \phantom{0} \\ j \end{pmatrix}$	$\omega$	$y_{t_i} \begin{pmatrix} \phantom{0} \\ m \end{pmatrix}$
$\omega t_n$	$\dots$	$y_{t_n} \begin{pmatrix} \phantom{0} \\ 0 \end{pmatrix}$	$\omega$	$\dots$	$y_{t_n} \begin{pmatrix} \phantom{0} \\ j \end{pmatrix}$	$\omega$	$y_{t_n} \begin{pmatrix} \phantom{0} \\ m \end{pmatrix}$

**Définition :**

Une série temporelle  $\{y_t\}_{t=1}^T$  est (la partie de dimension finie d') une réalisation d'un processus stochastique  $\{y_t\}$ .

$\omega$  La réalisation d'un processus stochastique est une fonction  $T \rightarrow \mathbb{R}$  où  $t \rightarrow y_t(\omega)$ . Le processus sous-jacent est dit avoir généré la série temporelle. La série temporelle  $y_1(\omega), \dots, y_T(\omega)$  est généralement notée  $y_1, \dots, y_T$  ou simplement  $y_t$ .

◆ Un processus stochastique peut être décrit par la fonction de distribution commune des toutes les sous-collections de dimension finie de  $y_t, t \in S \subset T$ . En pratique le système complet de distributions est souvent inconnu et on se cantonne aux premiers et seconds moments.

◆ La distribution jointe de  $(y_t, y_{t-1}, \dots, y_{t-h})$  est généralement caractérisée par sa fonction d'autocovariance qui représente le lien entre les valeurs à des dates différentes :



$$\begin{aligned}
\gamma_t(h) &= \text{Cov}(y_t, y_{t-h}) \\
&= E[(y_t - \mu_t)(y_{t-h} - \mu_{t-h})] \\
&= \int \int (y_t - \mu_t)(y_{t-h} - \mu_{t-h}) f(y_t, y_{t-h}) dy_t dy_{t-h}, \\
\mu_t &= E[y_t] = \int y_t f(y_t) dy_t, \quad \text{l'espérance (ou moyenne)} \\
&\text{inconditionnelle de } y_t.
\end{aligned}$$

La fonction d'autocorrelation est donnée par :

$$\rho_t(h) = \frac{\gamma_t(h)}{\sqrt{\gamma_t(0) \gamma_{t-h}(0)}}.$$

### 3 Stationnarité :

#### Définition :

Le processus  $\{y_t\}_{t=1}^T$  est dit stationnaire au sens faible, ou stationnaire au second ordre si les premier (moyenne ou espérance mathématique) et

second (variance et autocovariances) moments du processus existent et sont indépendants de  $t$  :

$$E[y_t] = \mu < \infty, \text{ Pour tout } t \in T,$$

$$E[(y_t - \mu)(y_{t-h} - \mu)] = \gamma(h), \text{ Pour tout } h \text{ et } t.$$

Lorsqu'  $y_t$  est stationnaire  $\gamma(h) = \gamma(-h) = \gamma(h)$ . La stationnarité est une propriété de stabilité, la distribution de  $y_t$  est identique à celle de  $y_{t-1}$ . La série oscille autour de sa moyenne avec une variance constante ; le lien entre  $y_t$  et  $y_{t-h}$  ne dépend alors que de l'intervalle  $h$  et non de la date  $t$  :

**Définition :**

Le processus  $\{y_t\}_{t=1}^T$  est dit strictement ou fortement stationnaire si pour tous  $h_1, \dots, h_n$ , la distribution jointe de  $(y_t, y_{t+h_1}, \dots, y_{t+h_n})$  dépend uniquement des intervalles de  $h_1, \dots, h_n$ , et non de  $t$  :

$$f(y_t, y_{t+h_1}, \dots, y_{t+h_n}) = f(y_t, y_{t+h_1}, \dots, y_{t+h_n}), \quad \forall (t).$$

La stationnarité stricte implique que tous les moments soient indépendants du temps.

### **Définition :**

Le processus  $\{y_t\}$  est appelé Gaussien si la distribution de  $f(y_t, y_{t+h_1}, \dots, y_{t+h_n})$  suit une loi Normale multivariée pour tous  $h_1, \dots, h_n$ .

En pratique, pour les séries suivant une distribution Gaussienne, la stationnarité au sens faible est équivalente à la stationnarité au sens strict.

### **4 Ergodicité :**

Le théorème d'ergodicité statistique concerne l'information qui peut être obtenue à partir d'une moyenne sur le temps concernant la moyenne

commune à tout instant. Remarquons que la loi faible des grands nombres ne s'applique pas car la série temporelle observée correspond à une seule observation du processus stochastique.

**Définition :**

Soit  $\{y_t(\omega), \omega \in \Omega, t \in T\}$  un processus stationnaire au sens faible, tel que  $E[y_t(\omega)] = \mu < \infty$  et  $E[(y_t - \mu)^2] = \sigma_y^2 < \infty$  pour tous  $t$ .

Soit  $\bar{y}_t = T^{-1} \sum_{t=1}^T y_t$  la moyenne temporelle. Si  $\bar{y}_t$  converge en probabilité vers  $\mu$  quand  $T \rightarrow \infty$ , alors  $\{y_t(\omega)\}$  est ergodique pour la moyenne.

**5. Rappels sur les espaces  $L^2$  :**

On considère le processus  $\{y_t\}$  défini sur l'espace de probabilité  $(\Omega, M, P)$ , à valeur dans  $\mathbb{R}$ .

### **Définition :**

L'espace  $L^2(\Omega, \mathcal{M}, P)$  est l'espace des variables de carrée intégrable (variance-covariance finies).

De façon plus générale et plus formelle, on désigne par  $L^p$  l'espace de Banach des classes d'équivalence (pour l'égalité  $P$ -presque sûre) des fonctions mesurables telles que  $\|f\|_p = \left[ \int |f|^p dP \right]^{1/p}$  soit finie.

### **Propriété :**

$L^2$  est un espace de Hilbert, muni du produit scalaire  $\langle \cdot, \cdot \rangle$  et la norme associée  $\|\cdot\|$

$$\begin{cases} \langle X, Y \rangle = E(XY) \\ \|X\|^2 = \langle X, X \rangle = E(X^2) = V(X) + E(X)^2. \end{cases}$$

Par définition de la covariance, on peut noter que, si  $X$  et  $Y$  sont centrées,  $\langle X, Y \rangle = E(XY) = cov(X, Y)$ .

### Théorème :(de projection):

Si  $H$  est un sous espace fermé de  $L^2$ , pour toute variable  $Y \in L^2$ , il existe une unique variable aléatoire  $\hat{Y} \in H$  tel que

$$\|Y - \hat{Y}\| = \min_{h \in H} \|Y - h\|,$$

caractérisé par  $\hat{Y} \in H$  et  $Y - \hat{Y} \in H^\perp$ .

### Remarque :


$X_n$  converge vers  $X$  au sens de  $L^2$  si  $\lim_{n \rightarrow \infty} \|X_n - X\| = 0$ , c'est-à-dire :

$$\begin{cases} \lim E(X_n) = E(X) \\ \lim V(X_n - X) = 0. \end{cases}$$

On peut alors définir la variable aléatoire  $Y = \sum_{n \in \mathbb{N}} a_n X_n$  comme limite,

dans  $L^2$  de  $Y_{p,q}$  tel que  $Y_{p,q} = \sum_{n=p}^q a_n X_n$  et  $Y = \lim_{p,q \rightarrow \infty} Y_{p,q} = \sum_{n \in \mathbb{N}} a_n X_n$ .

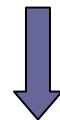
# Modèles Mathématiques

 Définition : Le but poursuivi est la formulation d'un modèle statistique qui soit une représentation **congruente** du **processus stochastique** qui génère la série observée.

## Approche :

Il est en pratique impossible de connaître la distribution d'une série temporelle  $\{y_t\}_{t \geq 0}$ , on s'intéresse par conséquent à la **modélisation** de la distribution conditionnelle de  $\{y_t\}$  via sa densité :

$$f(y_t | Y_{t-1})$$



Conditionnée sur l'historique du processus

$$Y_{t-1} = (y_{t-1}, y_{t-2}, \dots, y_0)$$



Il s'agit donc d'exprimer  $y_t$  en fonction de son passé

# Modèles Mathématiques

Résultat :

L'approche conditionnelle fournit une **Décomposition Prévission Erreur** selon laquelle :

$$Y_t = E[y_t | Y_{t-1}] + \varepsilon_t$$

où

$E[y_t   Y_{t-1}]$	est la composante de $y_t$ qui peut donner lieu à une prévision, quand l'historique du processus $Y_{t-1}$ est connu
$\varepsilon_t$	représente les informations imprévisibles



## Modèle de séries temporelles

### 1. Processus autorégressifs d'ordre 1, AR(1) :

$$\begin{cases} y_t = \alpha y_{t-1} + \varepsilon_t \\ \varepsilon_t \sim \text{WN}(0, \sigma^2) \text{ (bruit blanc)} \end{cases}$$

La valeur de  $y_t$  ne dépend que de son prédécesseur. Ses propriétés sont fonction de  $\alpha$  qui est facteur d'inertie :

$\alpha = 0$  :  $y_t$  est imprévisible et ne dépend pas de son passé, on parle de bruit blanc

$\alpha \in ]-1, 1[$  :  $y_t$  est stable autour de zéro

$|\alpha| = 1$  :  $y_t$  est instable et ses variations sont imprévisibles

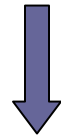
$|\alpha| < 1$  :  $y_t$  est explosif



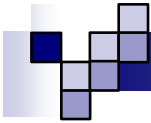
# Équations de Yule-Walker

En notant  $r_{xx}(i) = E[x_n x_{n-i}]$  l'autocorrélation de  $x_n$  on obtient

$$\sum_{i=1}^P r_{xx}(i-j)a_i = -r_{xx}(j) \quad j = 1 \dots P$$



éq de Yule Walker



Écriture sous forme matricielle

$$\begin{bmatrix} r_{xx}(0) & r_{xx}(P-1) \\ r_{xx}(P-1) & r_{xx}(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_P \end{bmatrix} = \begin{bmatrix} -r_{xx}(1) \\ -r_{xx}(P) \end{bmatrix}$$

$\mathbf{R} \quad \underline{a} \quad \underline{r}$

$$\underline{r} = \mathbf{R} \underline{a} \quad \Rightarrow \quad \underline{a} = \mathbf{R}^{-1} \underline{r} \quad \text{coefficients du filtre blanchisseur}$$

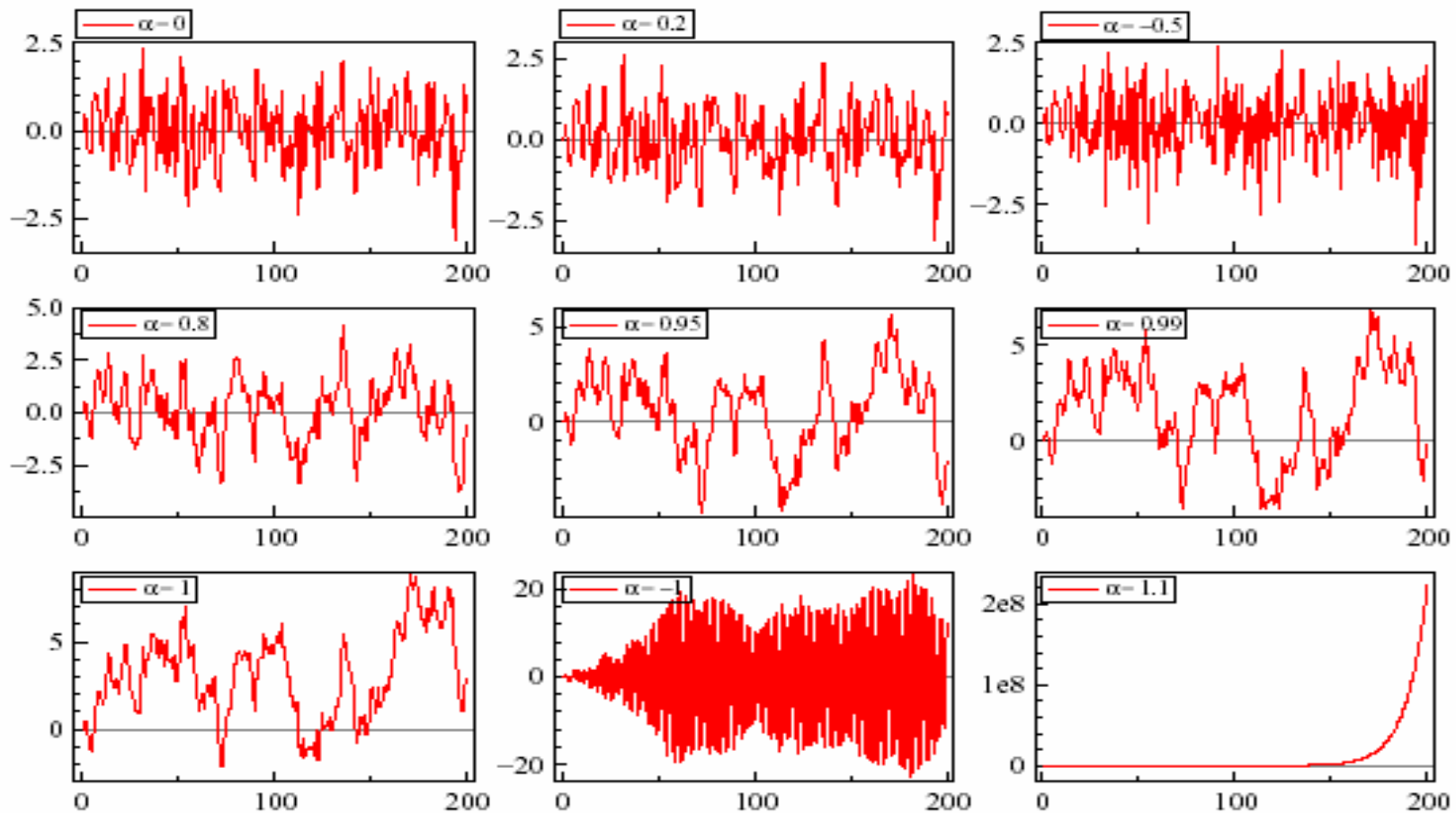
*Pb : inversion de la matrice  $\mathbf{R}$   $\rightarrow$  algorithme rapide d'inversion*

Par ailleurs,  $E[e_n^2] = \sigma^2 = r_{xx}(0) + \sum_{i=1}^P a_i r_{xx}(i)$  et l'ensemble peut s'écrire

$$\left[ \begin{array}{ccc|c} r_{xx}(0) & r_{xx}(1) & r_{xx}(P) & 1 \\ \hline r_{xx}(1) & r_{xx}(0) & r_{xx}(P-1) & a_1 \\ r_{xx}(P) & r_{xx}(P-1) & r_{xx}(0) & a_P \end{array} \right] = \begin{bmatrix} \sigma^2 \\ 0 \\ 0 \end{bmatrix}$$

# Modèles Mathématiques

## 1. Processus autorégressifs d'ordre 1, AR(1) :





## *Modélisation de séries chronologiques avec la méthodologie de Box-Jenkins*

Un modèle ARIMA est étiqueté comme modèle ARIMA (p,d,q) dans lequel :

p est le nombre de terme autorégressifs,

d est le nombre de différence,

q est le nombre de moyennes mobiles.



# Identification des modèles ARIMA

On désire ajuster un modèle ARMA( $p, q$ ):

$$\theta \quad \theta \quad \left( B \right) Z_t = \theta_0 + \left( B \right) a_t, \quad \forall t.$$

Un autre type de modèles est le modèle  
ARIMA( $p, d, q$ ):

$$\theta \quad \theta \quad \left( B \right) \left( 1 - B \right)^d Z_t = \theta_0 + \left( B \right) a_t, \quad \forall t.$$

Un modèle ARIMA est simplement un modèle dont la  $d^{\text{ième}}$  différence est ARMA.

Modèle	FAC	FACP
<b>Un paramètre autorégressif (p)</b>	décomposition exponentielle	pic à la période 1, pas de corrélation pour les autres périodes.
<b>Deux paramètres autorégressifs (p)</b>	une composante de forme sinusoïdale ou un ensemble de décompositions exponentielles	pics aux périodes 1 et 2, Aucune corrélation pour les autres périodes.
<b>Un paramètre de moyenne mobile (q) :</b>	pic à la période 1, aucune corrélation pour les autres périodes	exponentielle amortie.
<b>Deux paramètres de moyenne mobile (q) :</b>	pics aux périodes 1 et 2, Aucune corrélation pour les autres périodes	une composante de forme sinusoïdale ou un ensemble de décompositions exponentielles.
<b>Un paramètre autorégressif (p) et un de moyenne mobile (q) :</b>	Décomposition exponentielle commençant à la période 1	décomposition exponentielle commençant à la période 1.



# Méthode de Box et Jenkins

Box et Jenkins ont popularisé l'utilisation des modèles ARMA, en insistant sur les étapes nécessaires à la modélisation d'une série chronologique quelconque:

Les trois étapes sont:

- Identification des modèles (choix des ordres  $p$ ,  $d$  et  $q$ ).

- Estimation des paramètres; moindres carrés conditionnels, inconditionnels, maximum de vraisemblance.

- Validation du modèle (statistiques portmanteau; prévisions).





# Transformation de Box-Cox

La transformation a la forme:

$$y^{(\lambda)} = \begin{cases} y^{-1/\lambda}, & \lambda \neq 0, \\ \log(y), & \lambda = 0. \end{cases}$$

Le choix de  $\hat{\lambda}$  se fait souvent en effectuant un graphique de la vraisemblance en fonction de  $\lambda$ .



## Identification (suite)

Choix du niveau de différentiation. Ici on veut une série stationnaire. En particulier la moyenne ne doit pas dépendre du temps.

Si la série est stationnaire : Si  $\{z_t\}$  est ARMA( $p, q$ ), alors les autocorrélations  $\rho(k)$  sont en nombre infini et décroissent vers 0 *plutôt rapidement*.

Un estimateur de  $\rho(k)$  est donné par  $r(k)$ .



# Identification (suite)

De plus, on devrait avoir que:

$$\rho \quad r(k) \xrightarrow{P} \quad (k)$$

Cependant, la série  $\{z_t\}$  est non-stationnaire, que se passe-t-il? On sait que  $cov(z_t, z_{t-k})$  va dépendre de  $k$ , mais aussi de  $t$ . Ainsi les auto-corrélations  $(k)$  ne sont pas définies. Mais que sera le comportement des statistiques  $r(k)$ ?



# Théorème

Soit une série chronologique  $Z_1, Z_2, \dots, Z_n$  générée d'un processus  $ARIMA(p, d, q)$ , où  $d \geq 1$  on présume que le terme d'erreur  $\{a_t\}$  est un bruit blanc Gaussien.

Alors pour chaque délai  $k$  fixé, on a que:

$$r(k) \xrightarrow{P} 1, \quad 1 \leq k \leq K < n$$



# Cas stationnaire versus cas non-stationnaire, comportement des $r(k)$ .

Cas stationnaire:

$$\rho \quad r(k) \xrightarrow{P} (k)$$

Cas non-stationnaire:

$$r(k) \xrightarrow{P} 1, \quad 1 \leq k \leq K < n$$




# Autres éléments d'information

**Cas stationnaire:** Les  $r(k)$  comme fonction de  $k$  décroissent à 0 de façon exponentielle (décroissance vers 0 très rapide).

**Cas non-saisonnier:** à partir de  $k = 20$ , toutes les autocorrélations devraient être très près de 0.

**Cas non-stationnaire:** Décroissance des  $r(k)$  de manière linéaire vers 0? Décroissante très lente? Les  $r(k)$  sont toutes de même signes? Ce sont tous des indices d'un problème de stationnarité.



# Identification: choix de $p$ et de $q$

Ayant identifié  $d$ , on a maintenant comme modèle  $\left. \begin{array}{l} \\ \\ \end{array} \right\} W_t = \nabla^d Z_t$ .

Important éléments d'information:

Pour un *autorégressif*, toutes les autocorrélations partielles s'annulent après un certain délai.

Pour un *moyenne-mobile*, toutes les autocorrélations s'annulent après un certain délai.

# Propositions


sous l'hypothèse où  $(X_t) \sim MA(q)$  et que  $(\epsilon_t)$  est stationnaire

à l'ordre  $q$ , alors  $\sqrt{T} \frac{\hat{\rho}_T(h) - \rho}{\sqrt{1 + 2 \sum_{k=1}^q \rho^2(k)}} \xrightarrow{L} N(0,1)$  pour  $h > q$ .

En particulier l'intervalle de confiance à 95% des autocorrélations,

$$\rho \left[ \hat{\rho}_T(h) \pm 1.96 \sqrt{\frac{1 + 2 \sum_{k=1}^q \rho^2(k)}{T}} \right].$$





sous l'hypothèse où  $(X_t) \sim AR^{\mathcal{E}}(p)$  et que  $(\varepsilon_t)$  est stationnaire à l'ordre 4, alors  $\sqrt{T} [\hat{a}_T(h) - a(h)] \xrightarrow{L} N(0,1)$  pour  $h > p$ .

En particulier l'intervalle de confiance à 95% des autocorrélations,

$$\left[ \hat{a}_T(h) \pm 1.96 \frac{1}{\sqrt{T}} \right].$$

Résultat de Quenouille (1949)



# Modélisation des résidus

Supposons que le processus est:

$$\theta (B) (1 - B)^d Z_t = (B) b_t, \quad \forall t.$$

Cependant, le processus  $\left. \begin{matrix} \{ \\ b_t \} \end{matrix} \right\}$  n'est pas un bruit blanc mais un ARMA  $(\tilde{p}, \tilde{q})$ . Donc:

$$\phi \quad \theta \quad \sim (B) b_t = \sim (B) a_t, \quad \forall t.$$

$$\phi \quad \theta \quad b_t = \sim^{-1} (B) \sim (B) a_t$$

# Estimation

Dans le modèle:

$$\theta \left( (1 - B)^d \right) Z_t = \theta_0 + (B) a_t, \quad \forall t.$$

On doit procéder à l'estimation de:

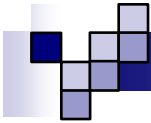
Paramètres autorégressifs:  $\phi_1, \dots, \phi_p$

Paramètres moyenne-mobiles:  $\theta_1, \dots, \theta_q$

Paramètre

$\sigma^2$  Paramètre

$\theta_0$   
 $a$



# Estimation des paramètres

Plusieurs techniques sont possibles:

Méthode du maximum de vraisemblance,

Méthode par moindres carrés conditionnels,

Méthode par moindres carrés inconditionnels.



# Validation du modèle

Quand on cherche à valider un modèle de régression, une étape consiste habituellement à analyser les résidus.

La situation est similaire en séries chronologiques.

Considérons le modèle:  $(\theta)(1 - \theta B)^d Z_t = \epsilon_t + (B)^{a_t}$

Les résidus sont:  $\hat{a}_t = Z_t - \hat{Z}_t$

Les  $\hat{Z}_t$  sont obtenus en estimant les divers paramètres.



# Test approximatif: test de bruit blanc sur les résidus

On se rappelle que si  $\{a_t\}$  est un bruit blanc, alors les  $r_a(k)$  sont asymptotiquement indépendants,

admettant pour un  $k$  donné une loi normale:

$$r_a(k) \approx N\left(0, \frac{1}{n}\right)$$

Un test de l'hypothèse  $H_0: r(k) = 0$  peut reposer sur le test:  $n^{1/2} |r(k)|$  et la règle de décision consiste à rejeter la nulle si:

$$n^{1/2} |r(k)| > 2$$



# Introduction aux tests de type portemanteau

Puisque les  $n^{1/2} |r(k)|$  sont approximativement de lois  $N(0,1)$ , et utilisant l'indépendance, lorsque  $\{a_t\}$  est bruit blanc fort, on trouve que:

$$Q(K) \approx n \sum_{k=1}^K r_a^2(k) \xrightarrow{L} \frac{2}{K}$$

Pour l'hypothèse nulle d'adéquation, on rejette pour de grandes valeurs, i.e.

$$\chi \quad Q(K) >_{\alpha} \frac{2}{K, 1-\alpha}$$

# Test de Box-Pierce et de Ljung-Box

En suivant un raisonnement similaire au test de bruit blanc, mais tenant compte du fait que l'on construit une statistique de test basée sur des résidus, Box et Pierce, ainsi que Ljung et Box ont montré que pour tester l'adéquation d'un modèle ARMA( $p, q$ ):

$$Q_{BP}(K) = n \sum_{k=1}^K r_{\hat{a}}^2(k) \xrightarrow{L} \chi^2_{K-p-q}$$

$$Q_{LB}(K) = (n+2) \sum_{k=1}^K \left( \frac{n-k}{n} \right) r_{\hat{a}}^2(k) \xrightarrow{L} \chi^2_{K-p-q}$$

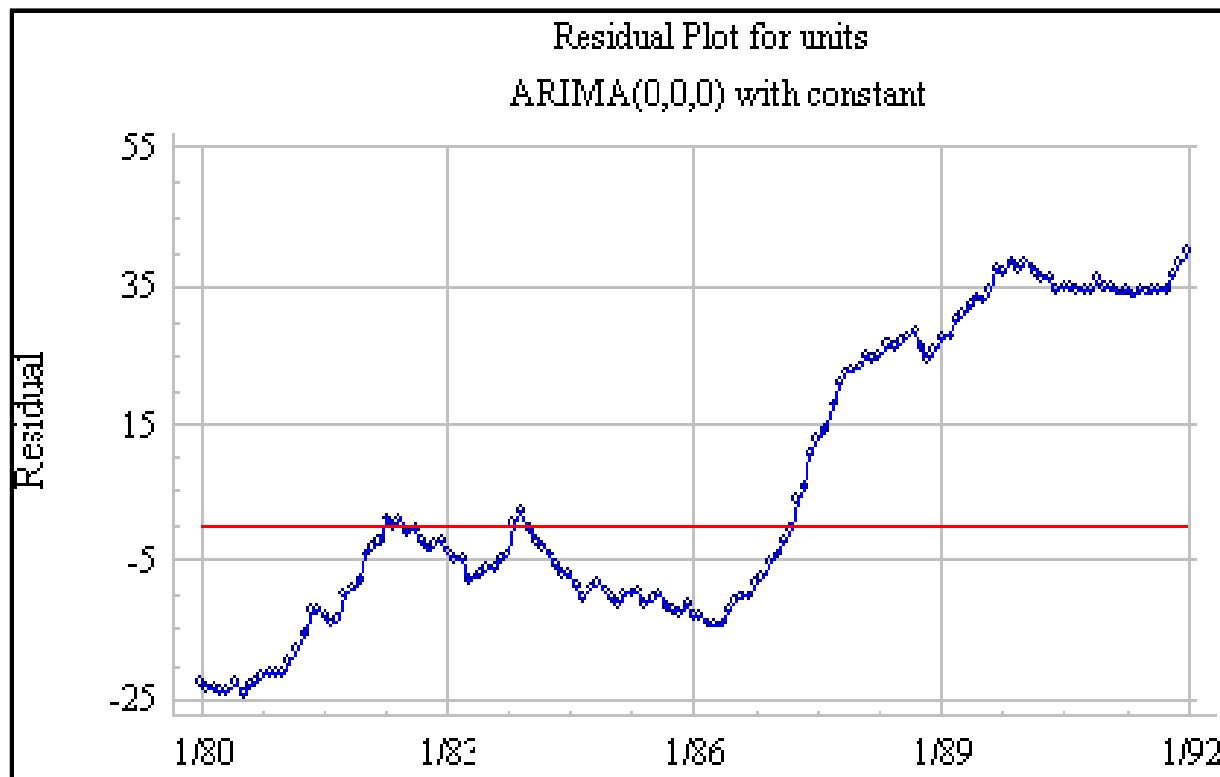
on rejette pour de grandes  $K$  l'adéquation si

$$Q_{LB}(K) > \chi^2_{K-p-q, 1-\alpha}$$



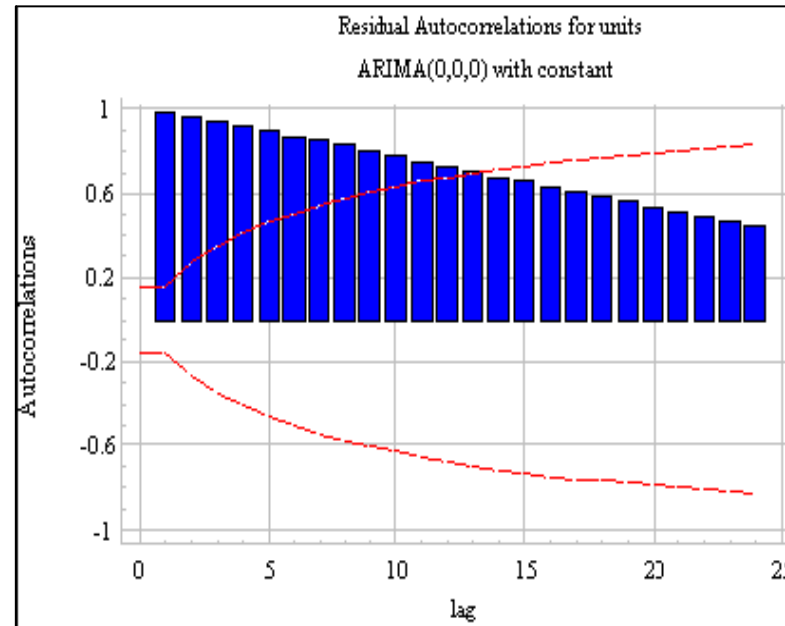
## 2. Les différentes étapes :

On part de la série temporelle originale de l'évolution des unités de ventes suivant :



## Etape 1 : détermination de l'ordre de différenciation

Le graphique de la fonction d'auto-corrélation présente une régression lente et linéaire typique de séries non stationnaires :



Or la méthode ARIMA suppose que l'on travaille sur une série stationnaire, c'est-à-dire que la moyenne et la variance soient constantes dans le temps.

On va donc remplacer la série originale par une série de différences adjacentes.

Pour corriger la non-stationnarité des valeurs, on pourra utiliser une transformation logarithmique ou exponentielle.

On a un écart type important 17.56. Cette série nécessite donc d'être différenciée.

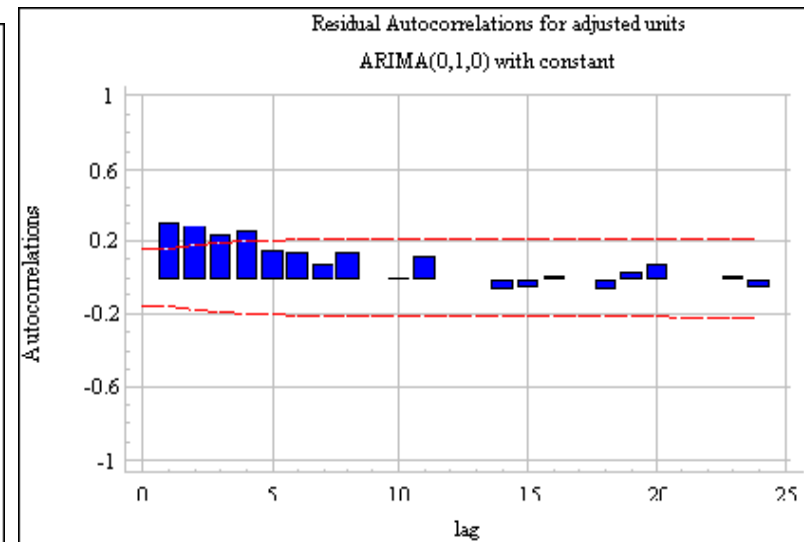
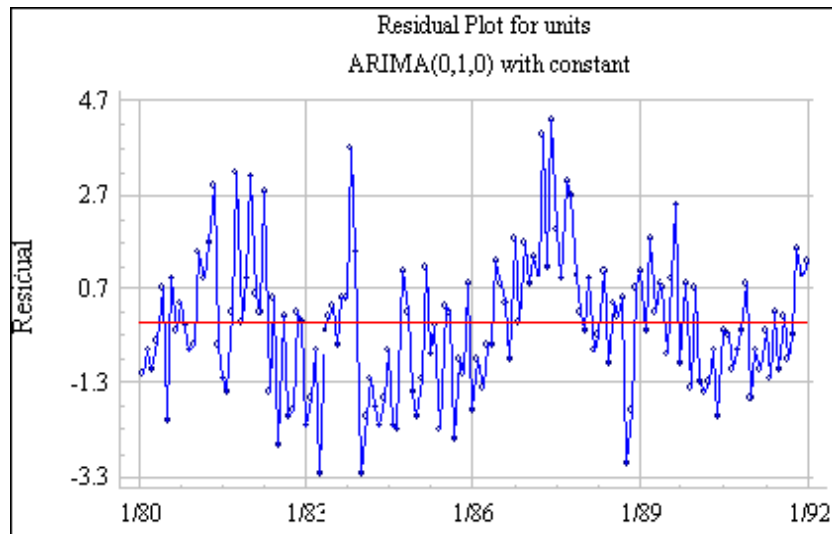
Une différenciation d'ordre 1 suppose que la différence entre 2 valeurs successives de  $y$  est constante. On utilise donc la fonction suivante :

$$y_t - y_{t-1} = \mu + \varepsilon_t$$

où  $\mu$  est la constante du modèle et représente la différence moyenne en  $y$ .

Si  $\mu = 0$ , la série est stationnaire.

Une première différenciation avec l'application du modèle ARIMA(0,1,0) donne les résidus suivants :

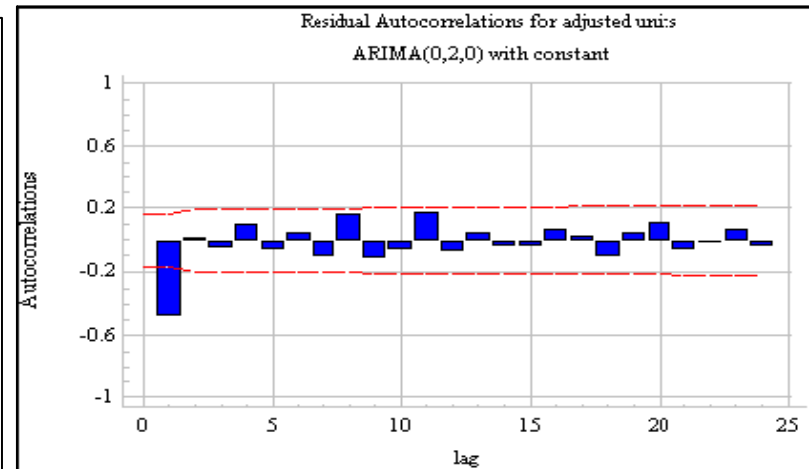
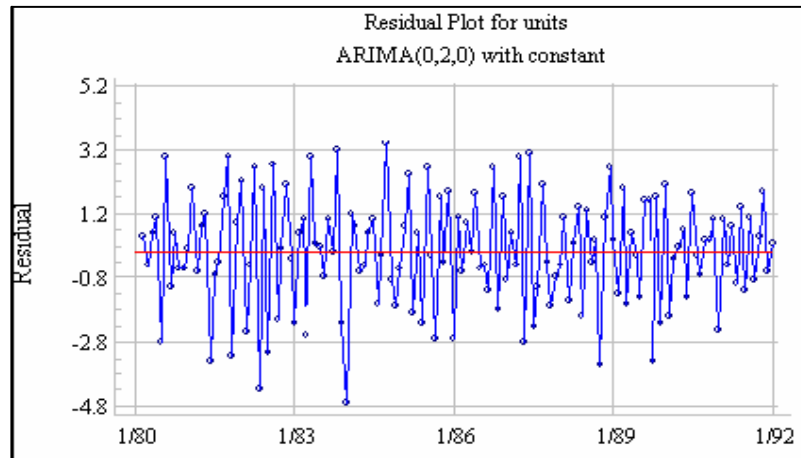


La série semble a peu près stationnaire et l'écart type a été réduit de manière importante : 1.54 au lieu de 17.56.

Si on essaie une seconde différenciation en appliquant un modèle ARIMA(0,2,0). Les modèles d'ordre 2 ne travaillent plus sur des différences mais sur les différences de différence. On utilisera alors l'équation de prédiction suivante :

$$y_t - 2y_{t-1} + y_{t-2} = \mu + \xi_t \text{ ou encore } y_t = \mu + 2y_{t-1} - y_{t-2} + \xi_t$$

on obtient les résultats suivants :



Cette série montre des signes clairs de sur-différenciation et l'écart type a augmenté de 1.54 à 1.81. Ceci semble indiqué que l'ordre optimal de différenciation pour cette série est de 1.

Toute fois ce modèle devra être optimisé par l'ajout des termes AR ou MA.

## Conclusion intermédiaire :

Un modèle sans différenciation suppose que la série originale est stationnaire.

Un modèle avec une différenciation d'ordre 1 suppose que la série originale présente une tendance constante.

Un modèle avec une différenciation d'ordre 2 suppose que la série originale présente une tendance variant dans le temps.

## Etape 2 : détermination des termes AR

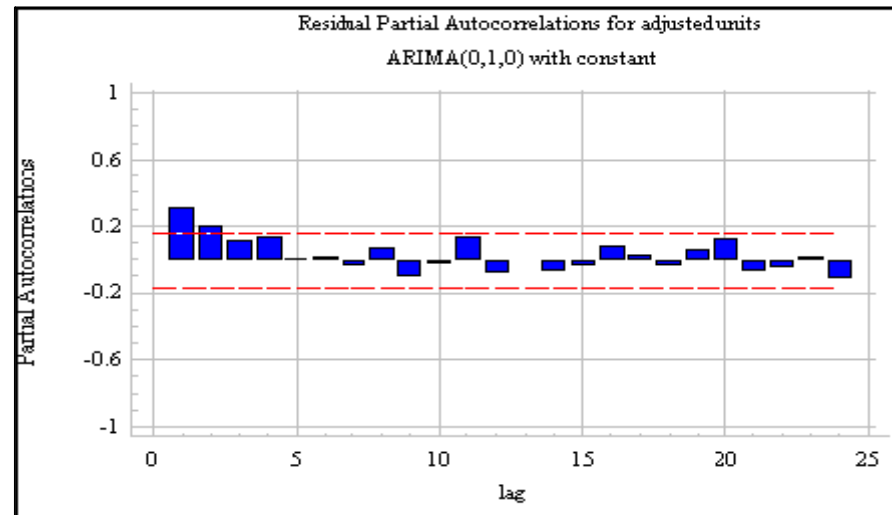
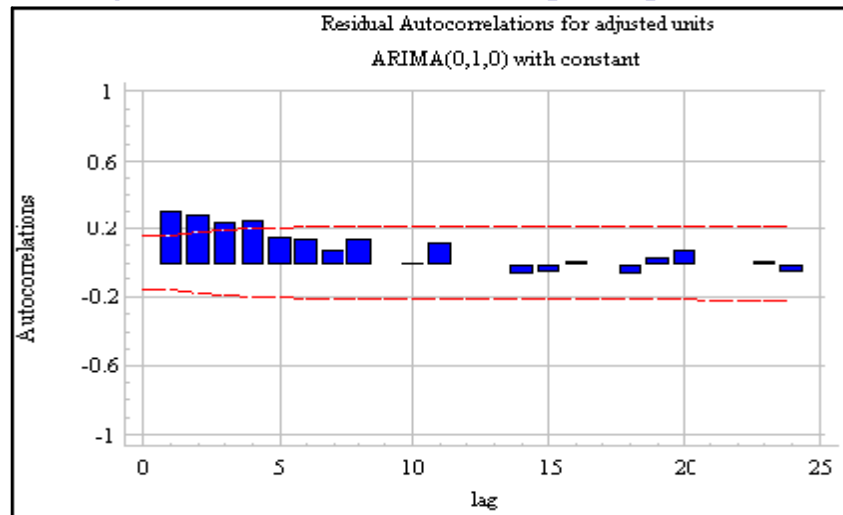
Analyse basée sur l'examen des fonctions d'auto-corrélation (ACF) et d'auto-corrélations partielles (PACF).

Auto-corrélation est la corrélation d'une série avec elle-même selon un décalage défini.

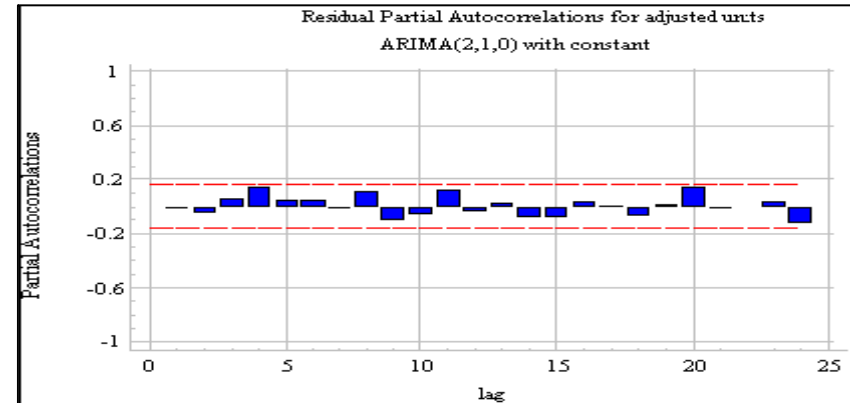
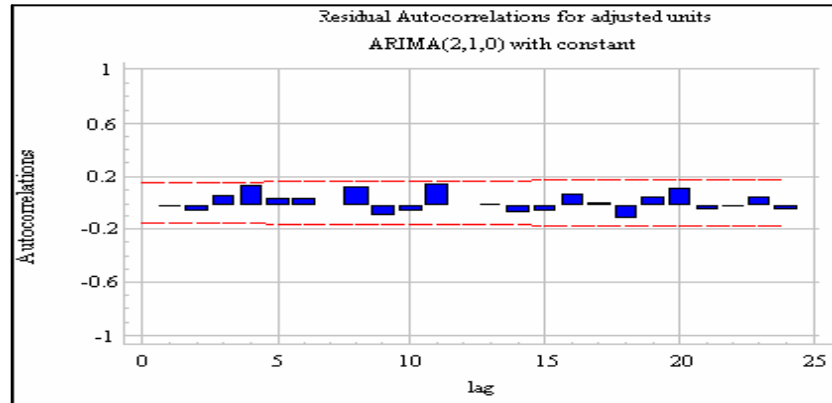
Les modèles autorégressifs supposent que  $y_t$  est une fonction linéaire des fonctions précédentes

$$y_t = \mu + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \Phi_3 y_{t-3} + \xi_t$$

où  $\xi$  est le choc aléatoire et  $\Phi_1$ ,  $\Phi_2$  et  $\Phi_3$  sont les coefficients d'auto-régression compris dans l'intervalle  $]-1,1[$

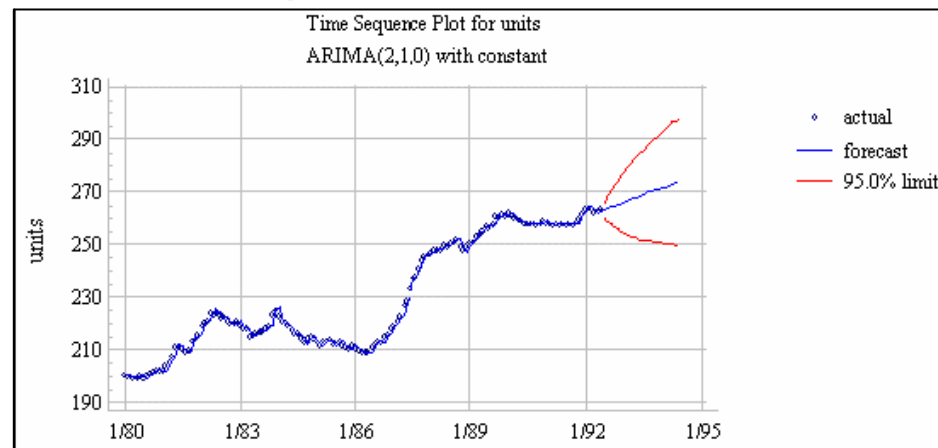


Si on ajuste cette série avec un modèle ARIMA(2,1,0) on obtient les fonctions ACF ET PACF suivantes :



L'analyse montre que les coefficients AR sont significativement différents de 0 et que l'écart type a été réduit de 10% (1.42 au lieu de 1.54). L'équation de prédiction a donc la forme suivante :  $y_t = \mu + y_{t-1} + \Phi_1(y_{t-1} - y_{t-2}) + \Phi_2(y_{t-2} - y_{t-3})$  avec  $\mu = 0.258178$ ,  $\Phi_1 = 0.2524$  et  $\Phi_2 = 0.195572$

Cette équation permet d'établir le graphique de prédictions suivant :



### Etape 3 : détermination des termes MA

Analyse également basée sur l'examen des fonctions d'auto-corrélation (ACF) et d'auto-corrélations partielles (PACF).

Les modèles à moyenne mobile suggèrent que la série présente des fluctuations autour d'une valeur moyenne.

$$y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \varepsilon_t$$

où  $\theta_1$ ,  $\theta_2$  et  $\theta_3$  sont les coefficients de moyenne mobile.

L'analyse des différents résultats va montrer que le modèle le plus pertinent serait un ARIMA(0,2,1) dont l'équation de prédiction serait la suivante :

$$y_t = 2y_{t-1} - y_{t-2} - \theta_1 \varepsilon_{t-1}$$



## Conclusion :

Ces deux modèles peuvent ajuster de manière alternative la série de départ.

Le choix d'un ou l'autre modèle peut reposer sur des présupposé théoriques liés au phénomène observé.

La décision n'est pas simple et les cas les plus atypiques requièrent, outre l'expérience, de nombreuses expérimentations avec des modèles différents (avec divers paramètres ARIMA).

Puisque le nombre de paramètres (à estimer) de chaque type dépasse rarement 2, il est souvent judicieux d'essayer des modèles alternatifs sur les mêmes données.