

UNIVERSITÉ JOSEPH FOURIER  
MASTER 2 P : INGÉNIERIE STATISTIQUE

UFR IMA

COURS  
DE  
SÉRIES CHRONOLOGIQUES

Serge Dégerine

17 septembre 2007



# INTRODUCTION

Une série chronologique est constituée de l'ensemble des observations d'une grandeur effectuées à intervalles réguliers au cours du temps. Les exemples dans le monde économique et social sont donc nombreux : inflation, cours boursiers, chômage, productions, exportations, natalité, immigration, scolarisation, logement,... Ces grandeurs sont souvent mesurées à l'aide d'indices ou de taux publiés dans des revues spécialisées. Les exemples ne manquent pas non plus à l'intérieur même de l'entreprise, qu'elle soit à caractère industriel, commercial ou de services : chiffre d'affaires, stocks, ventes, prix, vie d'un produit, clientèle,... La plupart des disciplines scientifiques sont amenées à traiter des données temporelles : astronomie, météorologie, biologie, médecine, physique,... En particulier le traitement du signal, qui intervient dans ces domaines, constitue à lui seul une activité de recherche importante très liée à celle effectuée en séries chronologiques. L'étude des premiers exemples relève essentiellement du domaine temporel alors que le traitement du signal fait beaucoup plus appel à l'analyse spectrale. Les deux approches sont souvent complémentaires mais nous considérerons ici uniquement l'aspect temporel. Citons simplement quelques ouvrages spécifiques au traitement du signal : [KAY87], [MAR87], [OWY83].

La spécificité de l'analyse d'une série chronologique, qui la distingue d'autres analyses statistiques, est précisément dans l'importance accordée à l'ordre dans lequel sont effectuées les observations. Les méthodes statistiques classiques demandent souvent que les variables étudiées soient stochastiquement indépendantes et observées plusieurs fois. En séries chronologiques la dépendance temporelle entre les variables constitue la source principale d'information. Celle-ci peut être entièrement contenue dans la valeur moyenne des variables, qui sont alors supposées stochastiquement indépendantes. Cependant l'ordre demeure essentiel car cette moyenne représente l'évolution lente du phénomène (tendance) à laquelle s'ajoute parfois un effet périodique (mouvement saisonnier). Les contraintes sur la fonction moyenne sont nécessaires car chaque variable n'est observée qu'une seule fois. Lorsque la moyenne

est nulle ou constante, la dépendance temporelle se situe dans l'organisation des différentes corrélations entre les variables. Dans ce cas l'hypothèse de stationnarité au second ordre, qui stipule que cette organisation est invariante par translation du temps, sert à compenser l'absence de répétitions. Ces deux approches sont complémentaires et conduisent naturellement à l'idée d'introduire la dépendance temporelle simultanément sur les deux premiers moments des variables observées, c'est une façon élémentaire de sortir de la stationnarité. En effet la non stationnarité permet plus généralement de prendre en compte une tendance et/ou un effet saisonnier de façon stochastique. Dans le cas gaussien on sait que l'ordre deux suffit à caractériser la structure probabiliste des variables. Une extension consiste alors à considérer des propriétés qui font appel à des moments d'ordre supérieur et exige ainsi de sortir de l'hypothèse gaussienne.

La discussion précédente fait apparaître trois grandes options pour aborder l'étude d'une série chronologique. Cependant le nombre d'observations dont on dispose peut restreindre ce choix de façon importante. Une série trimestrielle observée sur cinq ans laisse peu de degrés de liberté. Dix ans d'une série mensuelle ne constitue pas non plus un grand champ d'investigation. Ainsi la plupart des données économiques à court ou moyen terme ne se prêtent pas à des traitements statistiques très sophistiqués. L'utilisation de modèles paramétriques adaptés et comportant un nombre restreint de paramètres permet de compenser la faiblesse du nombre d'observations. Le domaine financier fournit des exemples de séries composées d'un très grand nombre d'observations mais l'analyse d'une rentabilité pourra être différente selon que l'on se place à l'échelle journalière, hebdomadaire ou mensuelle.

Notons que le temps est discret, puisqu'il sert à indexer les observations, même si la grandeur étudiée varie continûment. Ainsi la modélisation en temps continu, qui peut être préférable dans certains cas où le temps constitue un élément dynamique du système considéré, même si l'analyse statistique passe nécessairement par une phase de discrétisation, ne sera pas abordée. D'autre part la distinction entre temps continu et temps discret résulte plus souvent d'un choix de modélisation que de considérations physiques sur le phénomène étudié. L'échelle de temps est également très importante car le pas de discrétisation peut être infinitésimal. Le traitement digital du son supplée actuellement le traitement analogique.

Une simplification importante est de ne considérer que des variables scalaires réelles. Les méthodes vectorielles sont largement développées ([HAN70], [PRI81]) et il existe des approches spécifiques aux variables à valeurs entières.

Nous ne les étudierons pas ici.

Certaines données historiques célèbres sont utilisées par de nombreux auteurs afin de comparer les différentes approches dans l'analyse d'une série chronologique. C'est en particulier le cas de l'activité solaire depuis 1700 jusqu'à nos jours (*cf.* Figures 1 et 2). Marple [MAR87] donne une présentation très complète de ces données, dont l'origine est attribuée à Rudolph Wolf, et reproduit les moyennes mensuelles de 1749 à 1984. Les moyennes annuelles sur certaines périodes figurent dans les ouvrages d'Anderson [AND71] et de Box, Jenkins & Reinsel [BJR94] sous le nom de Wölfer's Sunspot Numbers. Selon Izenman [IZE83] cette erreur de nom serait due à Yule. Anderson [AND71] analyse l'indice annuel du prix du blé en Europe de 1500 à 1869 construit par Beveridge (*cf.* Figure 3). Box, Jenkins & Reinsel [BJR94] considèrent le nombre mensuel de passagers aériens internationaux aux Etats Unis de 1949 à 1960 (*cf.* Figure 4). On a également représenté un bruit blanc gaussien de variance unité (*cf.* Figure 5), c'est-à-dire une suite de variables aléatoires gaussiennes, centrées, de variance 1 et indépendantes. La marche aléatoire, obtenue en cumulant les valeurs de ce bruit, prend l'aspect d'un phénomène naturel (*cf.* Figure 6) alors qu'elle résulte uniquement du hasard.

L'analyse statistique d'une série chronologique peut être simplement de nature descriptive. Il s'agit alors de dégager des éléments de synthèse, en général sous forme de nouvelles séries, qui résument au mieux la grandeur observée ou d'exhiber certaines caractéristiques. Cette phase descriptive utilise principalement des méthodes non paramétriques, qui permettent de travailler avec peu d'hypothèses préalables, mais peut faire appel à des techniques paramétriques importantes. Elle est souvent une étape préalable à une analyse plus complète dans laquelle on cherche à expliquer le mécanisme générateur de la série. À cette fin on a recours à des modèles probabilistes qui, le plus souvent, sont de nature paramétrique. C'est en particulier le cadre idéal pour faire de la prévision, qui reste un des objectifs principaux de l'étude d'une série chronologique. Les aspects filtrage et contrôle, très liés à l'étude des systèmes dynamiques, ne seront pas développés ici.

Le cours est consacré aux méthodes dans lesquelles la partie structurée de la grandeur étudiée est déterministe et représente la moyenne des variables aléatoires observées ou d'une transformation simple de celles-ci (*cf.* [CAL65]). La partie stochastique est alors une erreur additive et forme une séquence de variables aléatoires centrées, non corrélées et de même variance. L'hypothèse gaussienne est retenue pour la construction d'intervalles de confiance ou de tests. La régression linéaire et le lissage par moyennes mobiles constituent

l'essentiel de cette approche (*cf.* [GM90], [MEL90], ...).

L'étude des séries chronologiques ne se limite évidemment pas à cette première approche. La deuxième étape consiste à supposer que les observations, après transformation éventuelle, sont issues d'une série stationnaire au second ordre. La structure du phénomène se traduit donc par les corrélations entre les variables. Les modèles autorégressifs moyennes mobiles (ARMA) constituent la base de cette seconde approche ([BRI81], [BD91], [BJR94], ...). Certaines méthodes permettent d'échapper aux hypothèses de stationnarité ou de linéarité. Les modèles autorégressifs moyennes mobiles intégrés (ARIMA) et les modèles autorégressifs conditionnellement hétéroscédastiques (ARCH) en sont les exemples principaux (*cf.* [DFT89], [DFT94], [GUE94], ...).

Les connaissances statistiques préalables pour aborder ce cours se résument aux méthodes inférentielles classiques en univers gaussien (*cf.* [ABC92], [GM89], [TAS89], ...).

On note  $y_1, y_2, \dots, y_T$  la séquence temporelle des données constituant la série chronologique étudiée. Sauf exception il s'agira toujours d'une suite de valeurs numériques réelles. Ces valeurs sont considérées comme les observations d'une suite de variables aléatoires,

$$Y_t = g(t) + \varepsilon_t, \quad t = 1, \dots, T,$$

dont la moyenne  $g(t) = \mathbb{E}(Y_t)$  représente à elle seule la partie structurée de la grandeur observée. La séquence des erreurs  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T$  est donc une suite de variables aléatoires centrées, non corrélées et de même variance  $\sigma^2$  :

$$\mathbb{E}(\varepsilon_t) = 0, \quad \text{Cov}(\varepsilon_t, \varepsilon_s) = \mathbb{E}(\varepsilon_t \varepsilon_s) = \sigma^2 \delta_{ts}, \quad s, t = 1, \dots, T,$$

où  $\delta_{ts} = 1$  si  $s = t$ , 0 sinon, est le symbole de Kronecker. Cette hypothèse est en effet suffisante dans la mesure où les méthodes utilisées ne font intervenir que les deux premiers moments des variables, ce qui est souvent le cas en séries chronologiques. On ajoutera l'hypothèse gaussienne pour la construction d'intervalles de confiance ou de tests.

Ces hypothèses sont cependant encore trop générales pour espérer la moindre inférence statistique, sauf à estimer  $g(t)$  par  $Y_t$ , puisque la moyenne se situe dans le même espace  $\mathbb{R}^T$  que les observations. On introduit alors des contraintes sur la fonction  $g(t)$ . Celles-ci peuvent être de nature paramétrique et se traduisent par l'appartenance de  $g(t), t = 1, \dots, T$  à un sous espace

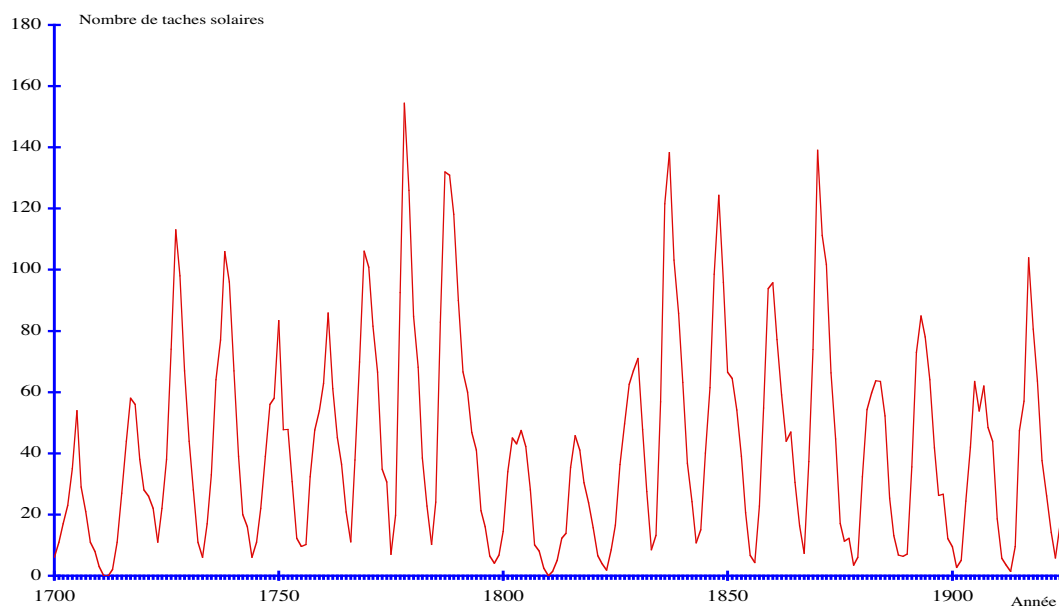


FIG. 1 – Nombre annuel de taches solaires selon Wolf de 1700 à 1924

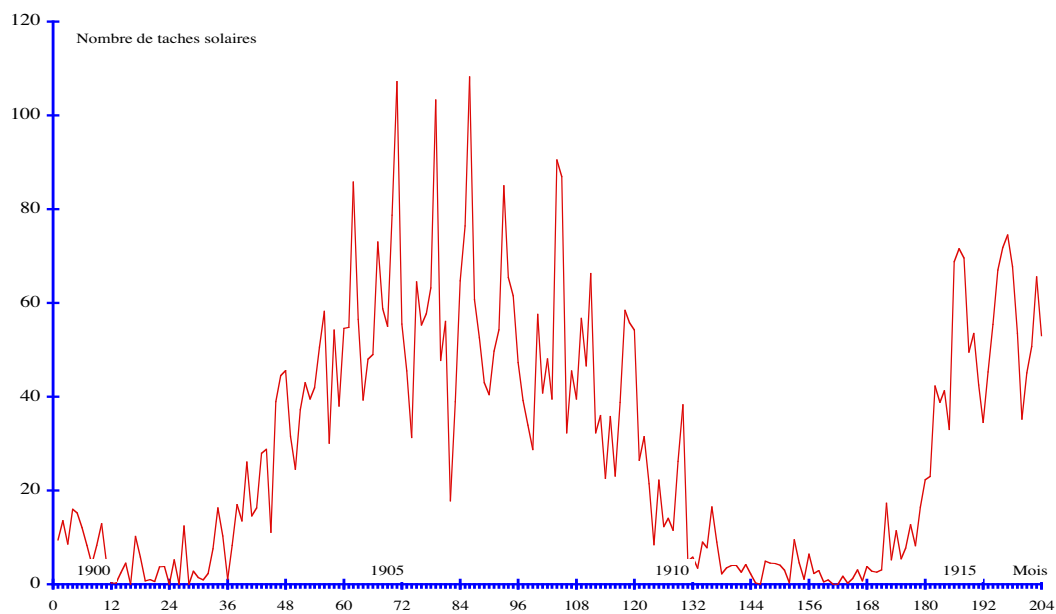


FIG. 2 – Nombre mensuel de taches solaires selon Wolf de 1900 à 1916

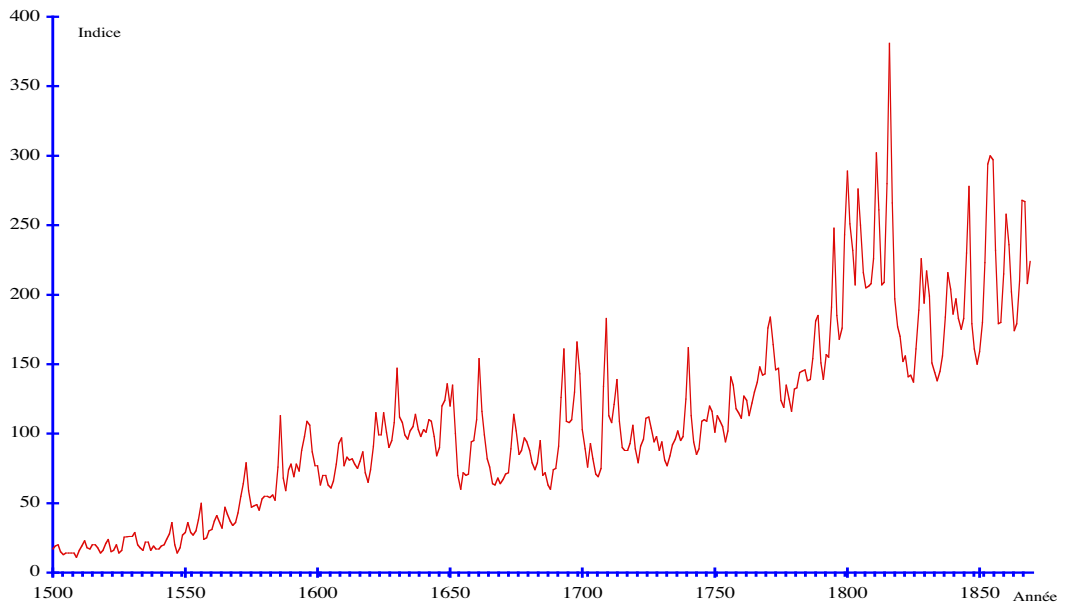


FIG. 3 – Indice annuel du prix du blé en Europe selon Beveridge de 1500 à 1869, base 100 : moyenne des années 1700 à 1745

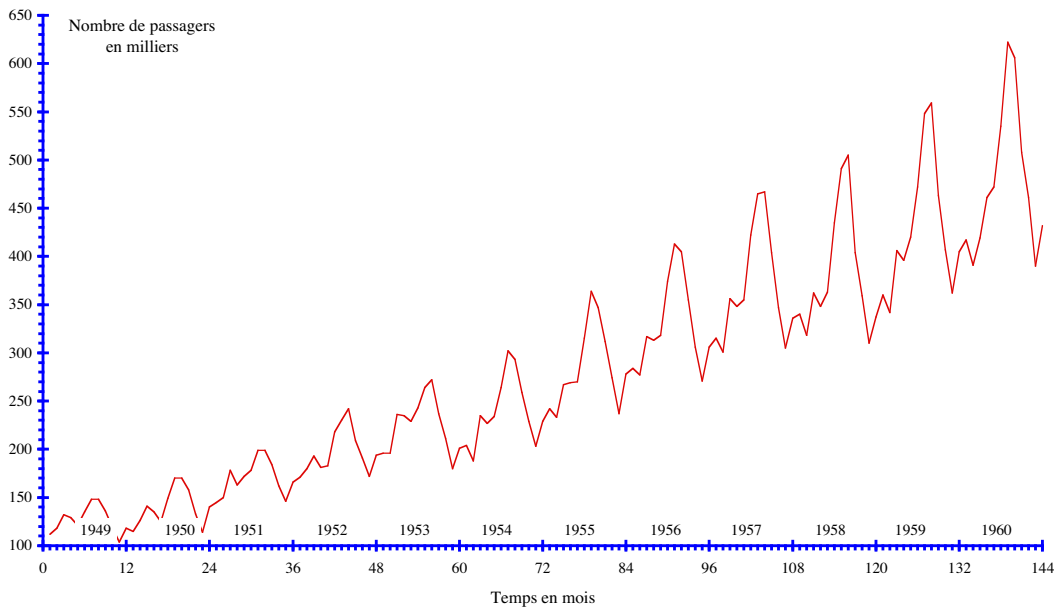


FIG. 4 – Nombre mensuel de passagers internationaux aux États Unis de 1949 à 1960

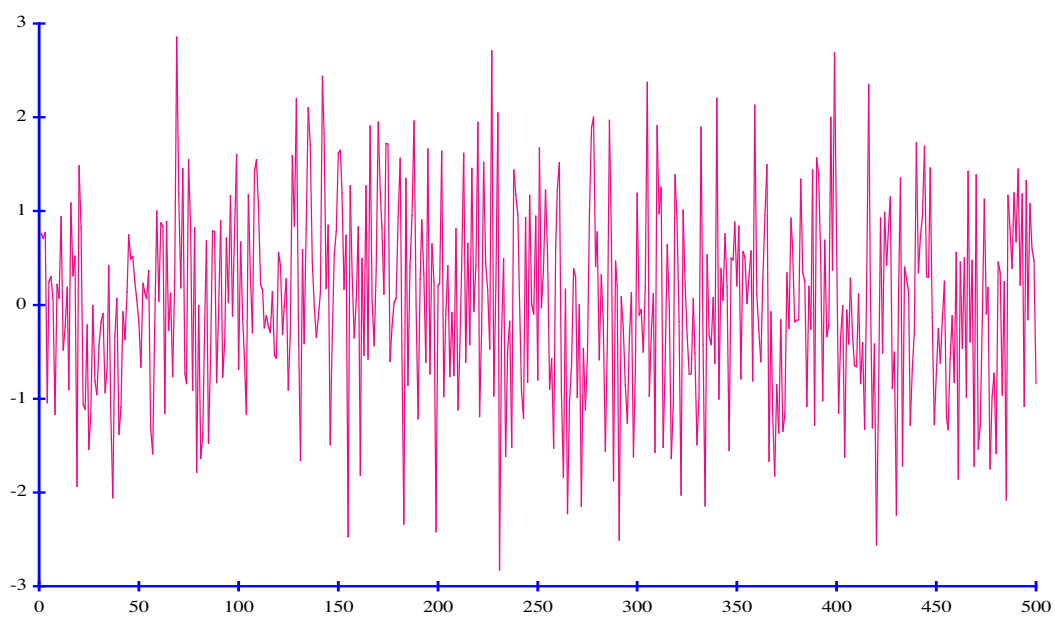


FIG. 5 – Bruit blanc gaussien de variance 1

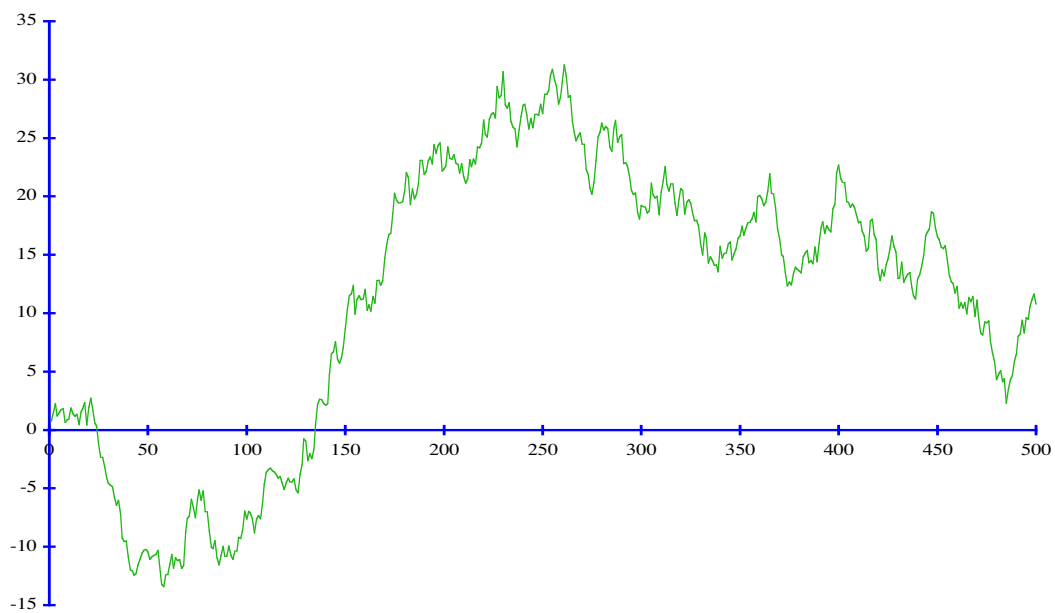


FIG. 6 – Marche aléatoire, valeurs cumulées du bruit blanc gaussien

vectorel de  $\mathbb{R}^T$ . C'est le cadre de la régression linéaire. La description de ce sous espace permet de séparer la tendance des aspects périodiques dans la moyenne  $g(t)$ . L'approche non paramétrique consiste à lisser les observations, le plus souvent par le biais de moyennes mobiles, de façon à éliminer la partie résiduelle  $\varepsilon_t$ . Elle permet également de dégager les deux composantes de  $g(t)$ . Les deux approches peuvent être utilisées de façon complémentaire sur une même série, on parle de méthodes semi-paramétriques.

Le premier chapitre précise les différentes notions évoquées jusqu'ici : temps, tendance, effet saisonnier, composante résiduelle. Il indique les premières démarches à effectuer sur une série brute avant de passer aux analyses statistiques développées dans les chapitres suivants. Le second chapitre est consacré aux méthodes basées sur la régression linéaire. Il permet d'aborder dans un cadre paramétrique l'estimation de la tendance sous forme polynomiale en présence ou non d'un effet saisonnier conduisant ainsi à des méthodes de prévision simples. Le dernier chapitre présente les méthodes de lissage. La tendance et l'effet saisonnier sont estimés de façon non paramétrique par le biais de moyennes mobiles. L'étude des solutions des équations de récurrence linéaires permet de faire le lien avec le chapitre précédent. La prévision est réalisée par le lissage exponentiel.

# Chapitre 1

## GÉNÉRALITÉS

Ce premier chapitre présente quelques notions élémentaires pour aborder l'étude d'une série chronologique. C'est une phase descriptive des données brutes, illustrée par quelques exemples, qui reste indispensable avant d'envisager d'appliquer les méthodes exposées par la suite.

Le premier paragraphe précise le rôle du temps dans la structure de la série. Les représentations graphiques mentionnées au second paragraphe constituent la première forme d'analyse destinée à éviter les erreurs grossières. Les composantes d'une série, tendance, effet saisonnier et erreurs ainsi que leur mode d'interaction font l'objet du dernier paragraphe.

### 1.1 LE TEMPS

#### 1.1.1 Définition d'une série chronologique

On appelle *série chronologique* (*série temporelle*, *chronique*) une suite d'observations numériques d'une grandeur effectuées à intervalles réguliers au cours du temps.

L'échelle de mesure et la variabilité de la grandeur sont telles que celle-ci sera toujours représentée par une variable continue à valeurs réelles. Le nombre mensuel de passagers aériens internationaux aux Etats Unis de 1949 à 1960, exprimé en milliers, varie entre 104 et 622 (*cf.* Figure 4). En général l'unité est choisie de sorte que la variable s'exprime avec 2 ou 3 chiffres significatifs.

La fréquence des observations peut être journalière, hebdomadaire, men-

suelle, trimestrielle, annuelle ou autre. Dans bien des situations économiques un effet saisonnier lié à une *période* connue est pressenti. Une chronique journalière sera observée pendant plusieurs semaines avec une périodicité de 5, 6 ou 7 jours selon le cas ; pour une chronique mensuelle (resp. trimestrielle) observée sur plusieurs années, la période est égale à 12 (resp. 4). Par la suite nous utiliserons systématiquement l'exemple mensuel dans nos commentaires.

La variable mesurée peut être l'état d'une grandeur à l'instant de mesure, on parle de *niveau* ou *stock*, ou le bilan d'une activité au cours de la dernière période écoulée à cet instant et on dit qu'il s'agit d'un *flux*. En météorologie la température est un niveau et la pluviométrie est un flux, de plus les relevés se font à heure fixe au cours du temps. En économie l'indice des prix à la consommation est un niveau et le taux d'inflation correspondant est un flux, dans ce cas l'indice fait référence au mois bien qu'il résulte de mesures pouvant être très étalées dans le temps. En finance un cours boursier est souvent considéré comme évoluant continûment bien qu'on utilise sa valeur en ouverture de séance durant cinq jours par semaine pour déterminer les rentabilités journalières puis hebdomadaires ou mensuels par sommation. L'analyse de ces trois séries, issues d'une même grandeur, est complémentaire et pas nécessairement redondante. Les techniques d'analyse de séries temporelles s'appliquent sur des séries pour lesquelles la chronologie est dictée par des événements pouvant avoir lieu à des intervalles de temps très irréguliers. On considère par exemple la série des intervalles de temps séparant les pannes successives dans un système informatique, celle donnant l'ampleur des tremblements de terre en un lieu donné. La variable espace sert parfois à indexer les observations.

### 1.1.2 Quelques précautions élémentaires

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

#### *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. En fait de légères entorses à la règle sont prises en compte dans la composante résiduelle ou dans la composante périodique lorsqu'elles sont systématiques

(février, jours fériés de mai,...). Par contre une période de grève nécessite plus de précautions.

#### *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 saisonnières (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).

#### *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).

#### *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  $j^e$  mois de la  $i^e$  année, et les données sont listées dans une table à double entrée.

### **1.1.3 Exemple**

L'indice mensuel des prix à la consommation (INSEE) de 1970 à 1978 est donné dans le Tableau 1.1, le Tableau 1.2 indique le taux d'inflation correspondant et les graphes associés sont reproduits dans les Figures 1.1 et 1.2 (*cf.* [GM90]).

Pour la période considérée l'indice est de base 100 en juillet 1970. La méthode de calcul de cet indice est modifiée périodiquement pour tenir compte des changements de mode de consommation. Il est cependant préférable de maintenir durablement la définition de la grandeur étudiée plutôt que de l'adapter systématiquement aux modifications de l'environnement car

<i>année</i>	<i>mois</i>	janv	fév	mars	avr	mai	juin	juil	août	sept	oct	nov	déc
	<i>j</i>	1	2	3	4	5	6	7	8	9	10	11	12
1970	1	97,9	98,2	98,5	99,0	99,4	99,8	100,0	100,4	100,8	101,2	101,6	101,9
1971	2	102,5	103,0	103,4	104,0	104,7	105,1	105,6	106,0	106,5	107,1	107,5	108,0
1972	3	108,3	108,9	109,4	109,8	110,4	111,0	111,9	112,5	113,2	114,2	114,9	115,5
1973	4	115,5	115,8	116,4	117,2	118,3	119,2	120,2	121,0	122,1	123,4	124,5	125,3
1974	5	127,4	129,1	130,6	132,7	134,3	135,8	137,5	138,6	140,1	141,8	143,1	144,3
1975	6	145,9	147,0	148,2	149,5	150,6	151,7	152,8	153,8	155,1	156,3	157,3	158,2
1976	7	159,9	161,0	162,4	163,8	164,9	165,6	167,2	168,4	170,2	171,8	173,2	173,8
1977	8	174,3	175,5	177,1	179,4	181,1	182,5	184,1	185,1	186,7	188,2	188,9	189,4
1978	9	190,3	191,7	193,4	195,5	197,4	198,9	201,5	202,5	203,8	205,7	206,8	207,8

TAB. 1.1 – Indice mensuel des prix à la consommation, base 100 en juillet 1970, de 1970 à 1978

<i>année</i>	<i>mois</i>	janv	fév	mars	avr	mai	juin	juil	août	sept	oct	nov	déc
	<i>j</i>	1	2	3	4	5	6	7	8	9	10	11	12
1970	1		0,31	0,31	0,51	0,40	0,40	0,20	0,40	0,40	0,40	0,40	0,30
1971	2	0,59	0,49	0,39	0,58	0,67	0,38	0,48	0,38	0,47	0,56	0,37	0,47
1972	3	0,28	0,55	0,46	0,37	0,55	0,54	0,81	0,54	0,62	0,88	0,61	0,52
1973	4	0,00	0,26	0,52	0,69	0,94	0,76	0,84	0,67	0,91	1,06	0,89	0,64
1974	5	1,68	1,33	1,16	1,61	1,21	1,12	1,25	0,80	1,08	1,21	0,92	0,84
1975	6	1,11	0,75	0,82	0,88	0,74	0,73	0,73	0,65	0,85	0,77	0,64	0,57
1976	7	1,07	0,69	0,87	0,86	0,67	0,42	0,97	0,72	1,07	0,94	0,81	0,35
1977	8	0,29	0,69	0,91	1,30	0,95	0,77	0,88	0,54	0,86	0,80	0,37	0,26
1978	9	0,48	0,74	0,89	1,09	0,97	0,76	1,31	0,50	0,64	0,93	0,53	0,48

TAB. 1.2 – Taux mensuel des prix à la consommation de 1970 à 1978

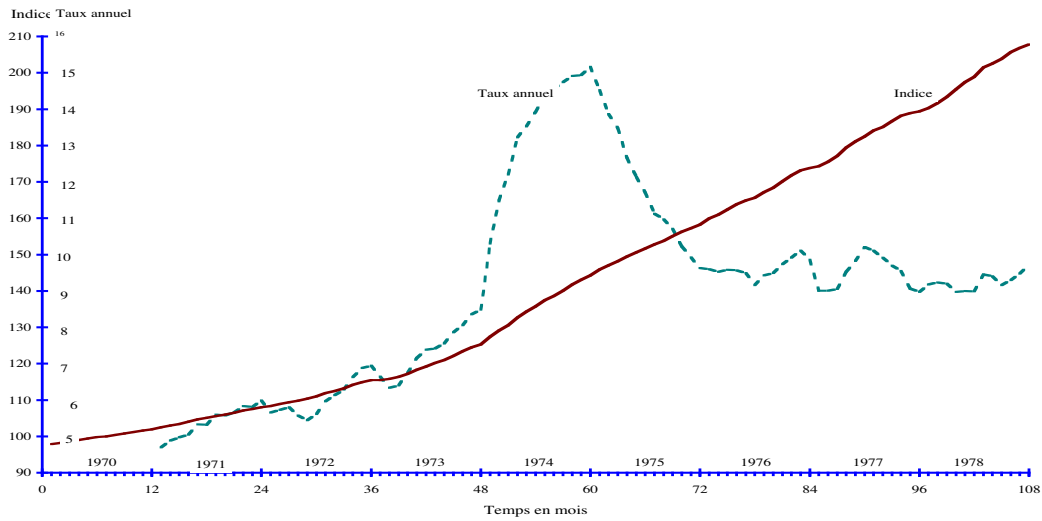


FIG. 1.1 – Indice mensuel des prix à la consommation, base 100 en juillet 1970, de 1970 à 1978

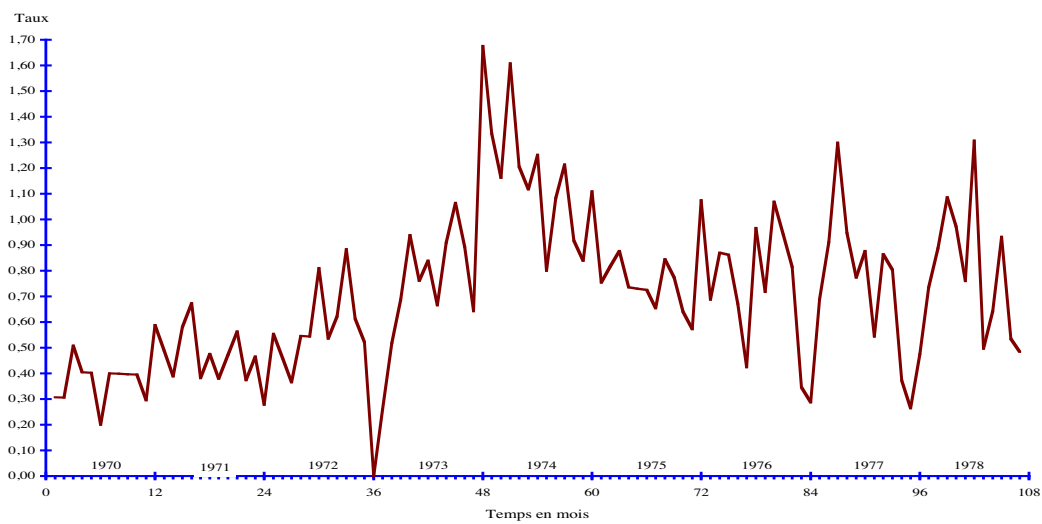


FIG. 1.2 – Taux d'inflation mensuel de 1970 à 1978

alors la série n'a plus aucun sens. L'indice actuel est de base 100 en 1980. On remarque que l'indice  $I_t$  et le taux,  $\tau_t = (I_t - I_{t-1})/I_{t-1}$ , sont deux grandeurs qui se comportent très différemment bien que liées fonctionnellement. On distingue deux grandes périodes pour la variation de l'indice, le changement intervient à la fin de l'année 1973 et correspond à la première augmentation brutale du prix du pétrole. L'indice traduit l'évolution à moyen terme des prix. Cependant il est courant de considérer le taux d'inflation annuel, calculé sur les 12 derniers mois, pour mesurer cette évolution. Celle-ci fait apparaître quatre périodes : une croissance moyenne jusqu'à la fin 1973, une forte croissance suivie d'une forte décroissance au cours des années 1974 et 1975 et une stabilité pour les années 1976 à 1978. La variation du taux d'inflation mensuel s'analyse différemment : on retrouve la croissance moyenne jusqu'à la fin 1973 où intervient une rupture, les années 1974 et 1975 se manifestent par une forte décroissance et la période de stabilité des années 1976 à 1978 est inchangée. L'analyse figurant dans [GM90] est différente : une période stable jusqu'au mois d'octobre 1973 avec un taux voisin de 0,4%, une croissance forte jusqu'au milieu de l'année 1974 suivie d'une décroissance pour atteindre une nouvelle période de stabilité à partir de 1976 avec un taux proche de 0,7%.

## 1.2 REPRÉSENTATIONS GRAPHIQUES

### 1.2.1 Représentation de la chronique

La représentation graphique des observations est une étape indispensable avant d'entreprendre une analyse plus technique de la chronique. Les points  $(t, y_t), t = 1, \dots, T$  sont représentés dans un système d'axes orthogonaux (échelles arithmétiques). Ils sont joints chronologiquement par des segments de droite pour faciliter la visualisation. Cette représentation permet d'apprécier l'évolution lente du phénomène (tendance), de dégager les périodes de stabilité. Elle suggère parfois d'opérer une transformation de la grandeur. C'est très nettement le cas pour le trafic aérien (*cf.* Figure 4) où la transformation logarithmique s'impose. Nous reviendrons sur ce point dans le prochain paragraphe.

L'interprétation de la tendance est délicate. Lorsqu'elle est naturellement liée au phénomène observé (trafic aérien, indice des prix,...), elle est reproductible dans un contexte similaire et sera considérée comme déterministe. Par contre une tendance apparente peut résulter du pur hasard comme le montre l'exemple de la marche aléatoire (*cf.* Figure 6). Dans ce cas elle est de na-

ture stochastique et n'a pas d'interprétation autre que descriptive. L'analyse d'une série chronologique ne doit pas se faire au vu de ses seules valeurs numériques mais doit prendre en compte le contexte des observations. Cette représentation graphique est également utile pour le choix d'un modèle. L'aspect graphique est un indicateur sommaire permettant d'opérer un premier tri.

### 1.2.2 Représentation du mouvement saisonnier

La représentation de la chronique peut faire apparaître, en plus de la tendance, un aspect périodique plus ou moins marqué de période connue, 12 pour une chronique mensuelle, 4 pour une chronique trimestrielle,... Il est très net dans le cas du trafic aérien (*cf.* Figure 4). L'évolution de la production industrielle française, considérée au troisième chapitre, présente un mouvement saisonnier exceptionnel directement observable à la lecture des données. En général l'effet saisonnier est moins spectaculaire. Une première appréciation de son importance est obtenue graphiquement en représentant, pour une chronique mensuelle, les courbes annuelles (sur 14 mois). La représentation polaire de la chronique a le même objectif, elle est plus originale mais moins lisible. L'étude quantitative du mouvement saisonnier est traitée dans les deux prochains chapitres.

### 1.2.3 Exemple

Nous avons repris le taux d'inflation mensuel introduit plus haut en ne considérant que les quatre dernières années. Les différentes représentations graphiques sont données dans les Figures 1.3, 1.4 et 1.5. Même en faisant abstraction de l'année 1975, qui n'appartient pas à la période de stabilité finale, l'effet saisonnier semble non négligeable mais est mal stabilisé dans le temps. L'approche numérique devrait permettre de préciser cette impression.

## 1.3 LES MODÈLES POUR LA MOYENNE

### 1.3.1 Les composantes du modèle

Nous avons déjà évoqué à plusieurs reprises les notions de tendance, effet saisonnier et composante résiduelle. On distingue en effet généralement trois *composantes* dans une chronique  $Y_t, t = 1, \dots, T$ .

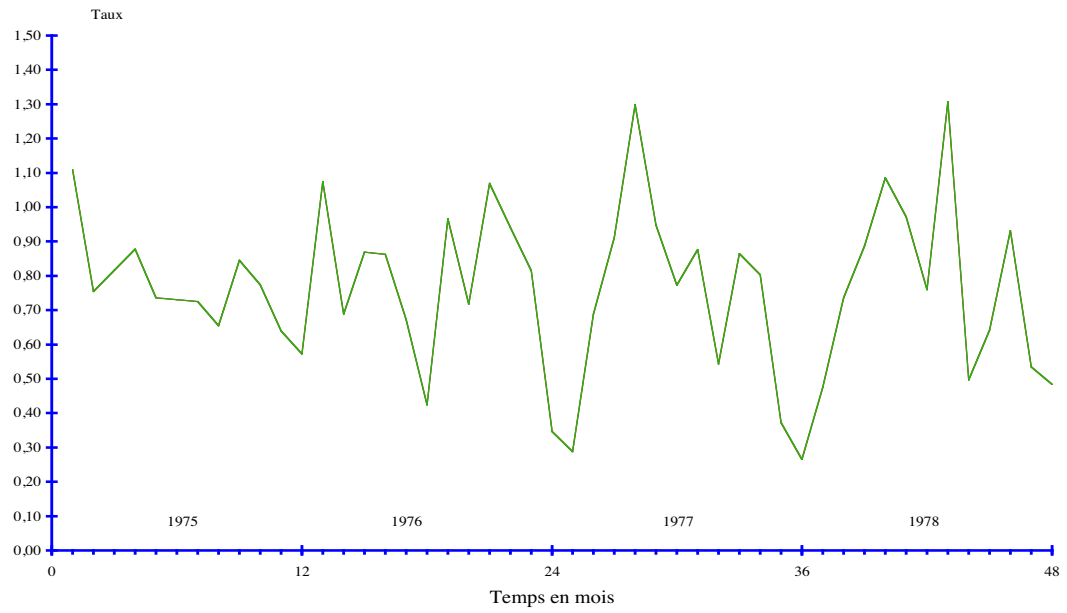


FIG. 1.3 – Taux d'inflation mensuel de 1975 à 1978

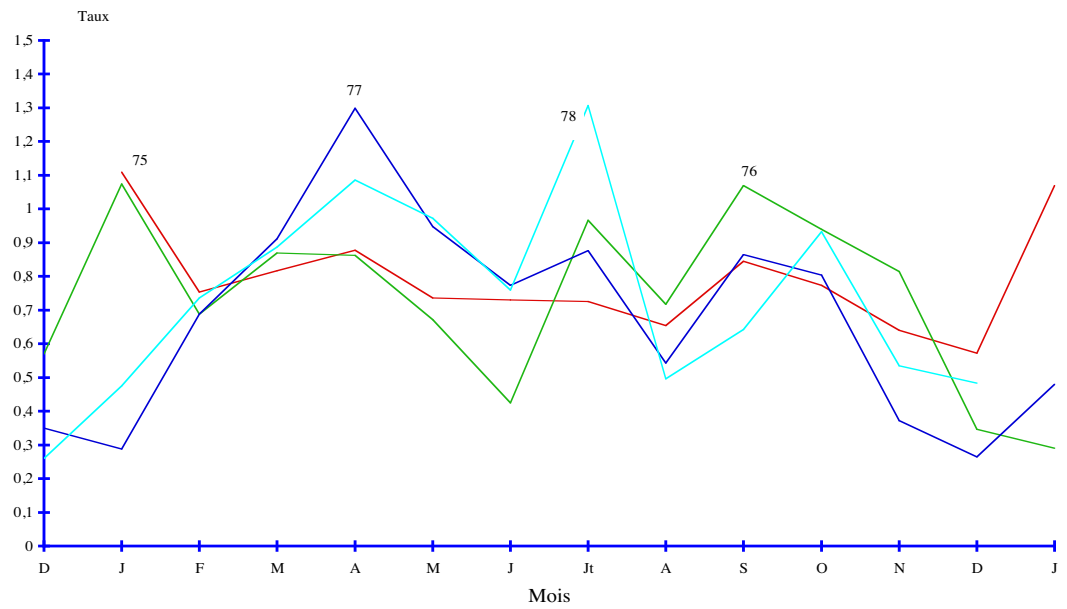


FIG. 1.4 – Mouvement saisonnier du taux d'inflation mensuel de 1975 à 1978

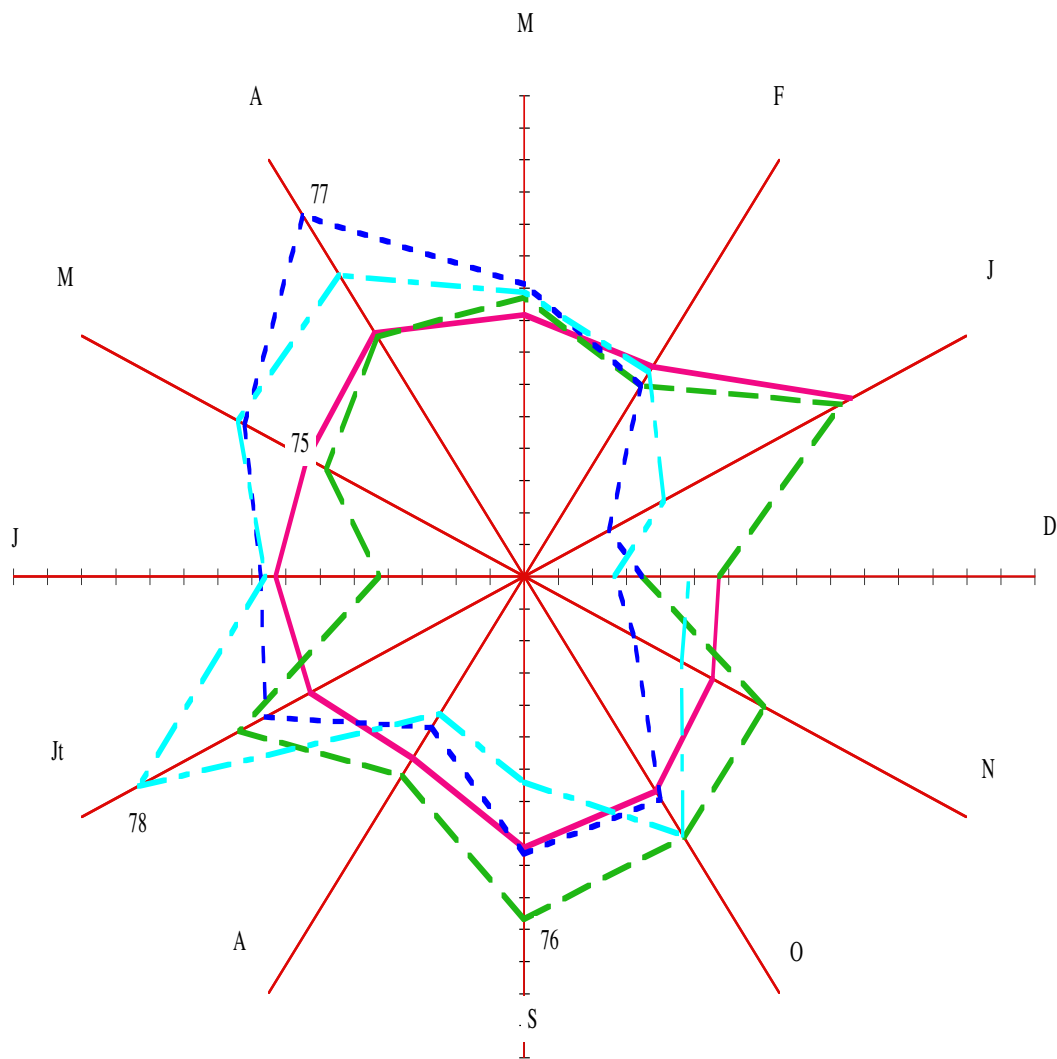


FIG. 1.5 – Représentation polaire du mouvement saisonnier du taux d'inflation mensuel de 1975 à 1978 (échelle polaire : pôle à 0%, graduation de 0,1%)

*Tendance*

La *composante fondamentale* ou *tendance* (trend) traduit l'évolution à moyen terme du phénomène. On parle aussi de mouvement conjoncturel ou mouvement extra-saisonnier. La chronique correspondante, notée  $f_t, t = 1, \dots, T$ , est une fonction à variation lente supposée déterministe dans cette approche. Elle sera estimée sous forme paramétrique (polynôme, exponentielle,...) ou comme le résultat d'une opération de lissage.

*La composante saisonnière*

La *composante saisonnière* ou *mouvement saisonnier* représente des effets périodiques de *période* connue  $p$  qui se reproduisent de façon plus ou moins identique d'une période sur l'autre. La chronique correspondante, également déterministe, est notée  $S_t, t = 1, \dots, T$ . Elle est généralement supposée rigoureusement périodique :  $S_{t+p} = S_t$  et les valeurs  $S_j = S_{ij}, j = 1, \dots, p$  d'une période sont appelées *coefficients saisonniers*. Le bilan de l'effet saisonnier sur une période doit être nul car il est pris en compte dans la tendance. La composante saisonnière permet simplement de distinguer à l'intérieur d'une même période une répartition stable dans le temps d'effets positifs ou négatifs qui se compensent sur l'ensemble de la période.

*La composante résiduelle*

La *composante résiduelle* ou *variations accidentelles* est la partie non structurée du phénomène. Elle est modélisée par une suite de variables aléatoires  $\varepsilon_t, t = 1, \dots, T$ , centrées, non corrélées et de même variance, on parle de *bruit blanc*.

Certains phénomènes économiques étudiés à très long terme présentent une *composante cyclique* (cycles d'activité) dont la période, de plusieurs années, est souvent mal définie. Cette composante est prise en compte dans la tendance sur les séries de taille moyenne et ne sera pas étudiée en tant que telle ici.

On comprend aisément l'intérêt de connaître la tendance. La composante saisonnière sert également à analyser le phénomène étudié et participe à la volonté de prévision. Cependant bien des séries économiques sont publiées en données corrigées des variations saisonnières (*série CVS*). De telles séries, dites désaisonnalisées, sont obtenues en éliminant la composante saisonnière de la série initiale. À cette fin la composante saisonnière est estimée en général de façon non paramétrique. La série CVS gomme l'effet saisonnier

en le répartissant de façon uniforme sur toute la période et conserve ainsi la grandeur observée dans son ensemble. En particulier elle contient la composante résiduelle et ne doit pas être confondue avec la tendance. Elle permet de comparer directement deux valeurs consécutives. Le chômage peut augmenter d'un mois sur l'autre en données brutes alors qu'il baisse en données corrigées des variations saisonnières.

On suppose généralement que la partie résiduelle a un ordre de grandeur très inférieur à celui des parties explicatives  $f_t$  et  $S_t$ . En fait il sera toujours possible de déceler la présence d'une partie déterministe, même de faible importance, à condition que la série observée soit suffisamment longue.

### 1.3.2 Les schémas de composition

Pour pouvoir séparer les trois composantes servant à décrire la série observée, il est nécessaire de préciser leur mode d'interaction. La plupart des séries chronologiques entrent dans l'un des schémas suivants.

- *Schéma additif* :  $Y_t = f_t + S_t + \varepsilon_t$ ,
- *Schéma multiplicatif* :  $Y_t = f_t \times S_t \times (1 + \varepsilon_t)$ ,
- *Schéma mixte* :  $Y_t = f_t \times S_t + \varepsilon_t$ .

En utilisant  $(1+\varepsilon_t)$  dans le cas multiplicatif, on conserve la même signification et les mêmes propriétés à chacune des trois composantes  $f_t$ ,  $S_t$  et  $\varepsilon_t$  dans les trois schémas de composition. Cependant il est nécessaire de supposer que  $(1 + \varepsilon_t)$  reste positif dans le modèle multiplicatif car la composante résiduelle ne peut être responsable du signe de la grandeur observée. Dans les schémas multiplicatif et mixte les oscillations dues à l'effet saisonnier ont une amplitude proportionnelle à la valeur de la tendance. C'est précisément l'argument utilisé pour faire le choix entre le schéma additif et les deux autres schémas au vu de la représentation graphique de la chronique. La distinction entre le schéma multiplicatif et le schéma mixte peut également s'apprécier graphiquement selon le même principe. Elle peut aussi relever de considérations sur l'origine des erreurs : une erreur structurelle (de modélisation) a des chances d'être proportionnelle à la grandeur étudiée alors qu'une erreur de mesure pourrait ne pas en dépendre. Le trafic aérien aux Etats Unis entre 1949 et 1960 (*cf.* Figure 4) fournit un exemple où le schéma multiplicatif s'impose. Nous avons déjà remarqué que dans ce cas il fallait considérer le logarithme de la variable qui est alors expliqué par un schéma additif. Par la suite nous ne considérerons plus que le cas additif. Il peut être nécessaire d'appliquer plusieurs fois la transformation logarithmique. Notons que la relation entre l'indice des prix et le taux d'inflation correspondant,  $I_t = I_{t-1}(1 + \tau_t)$ , invite à considérer la variable  $\ln(1 + \tau_t) = \ln I_t - \ln I_{t-1}$  plutôt que  $\tau_t$ . La différence

est négligeable pour des taux faibles. Les rentabilités d'un cours boursier sont définies de la même façon que le taux d'inflation. Le schéma mixte ne peut pas être transformé en schéma additif.

## Chapitre 2

# RÉGRESSION LINÉAIRE

La *régression linéaire* est certainement la méthode la plus simple pour analyser une série chronologique et effectuer des prévisions à court terme. Elle est cependant suffisante et assez riche pour aborder bien des problèmes économiques élémentaires.

Le modèle de régression général en séries chronologiques s'écrit :

$$Y_t = g(t; \theta) + \varepsilon_t, \quad t = 1, \dots, T,$$

où  $g(t; \theta)$  est une fonction déterministe connue du temps, à travers un paramètre vectoriel  $\theta = (\theta_1, \dots, \theta_r)$ , et  $\varepsilon_t, t = 1, \dots, T$ , une suite de variables aléatoires centrées. Ainsi la partie structurée de la grandeur étudiée est entièrement située dans la *moyenne*  $g(t; \theta) = \mathbb{E}(Y_t)$  et décrite de façon paramétrique. En effet l'*erreur*  $\varepsilon_t$ , que l'on suppose additive, est soumise à des hypothèses très restrictives. On peut distinguer trois situations principales.

- a) Les erreurs  $\varepsilon_t$  sont des variables aléatoires non corrélées, de moyenne nulle et de même variance  $\sigma^2$  ; on dit que  $\varepsilon_t, t = 1, \dots, T$  est un bruit blanc de variance  $\sigma^2$  et on note brièvement  $\varepsilon_t \sim b.b.(\sigma^2)$ .
- b) Les erreurs  $\varepsilon_t$  sont des variables aléatoires indépendantes, identiquement distribuées, de moyenne nulle et de variance  $\sigma^2$  ; on note  $\varepsilon_t \sim i.i.d.(0; \sigma^2)$ .
- c) Les erreurs  $\varepsilon_t$  sont des variables aléatoires gaussiennes indépendantes, de moyenne nulle et de variance  $\sigma^2$  ; on dit aussi qu'il s'agit d'un bruit blanc gaussien et on note  $\varepsilon_t \sim i.i.d.\mathcal{N}(0; \sigma^2)$  ou  $\varepsilon_t \sim b.b.\mathcal{N}(0; \sigma^2)$ .

L'hypothèse a) justifie la méthode des moindres carrés dans laquelle l'estimateur de  $\theta$  est recherché parmi les valeurs qui minimisent la somme des

carrés des erreurs :

$$\min_{\theta} \sum_{t=1}^T [Y_t - g(t; \theta)]^2.$$

Sous l'hypothèse gaussienne c), cette méthode équivaut à celle du maximum de vraisemblance. Cependant le problème de minimisation peut rester difficile selon la nature de la fonction  $g(t; \theta)$ . Nous nous placerons dans le cas où cette fonction est linéaire par rapport à  $\theta$ . Sous l'hypothèse a) les estimateurs obtenus sont sans biais et de variance minimum parmi les estimateurs linéaires et sans biais (estimateurs de Gauss-Markov). L'hypothèse c) est utilisée pour la construction d'intervalles de confiance ou de tests. L'hypothèse b) permet d'appliquer ces résultats sans l'hypothèse gaussienne mais dans un cadre asymptotique. Nous ne développerons pas ce dernier point.

La particularité en séries chronologiques est que le temps est la seule variable externe utilisée dans la description de la moyenne. Il existe cependant de nombreuses possibilités de modèles à travers le choix de la fonction  $g(t; \theta)$ . Citons trois exemples classiques où  $g(t; \theta) = g(t; \alpha, \beta, \gamma)$  représente une courbe de croissance avec seuil :

- Exponentielle modifiée :  $g(t; \alpha, \beta, \gamma) = \alpha\beta^t + \gamma$ .
- Fonction logistique :  $g(t; \alpha, \beta, \gamma) = 1/(\alpha\beta^t + \gamma)$ .
- Fonction de Gompertz :  $g(t; \alpha, \beta, \gamma) = \exp(\alpha\beta^t + \gamma)$ .

Dans ces situations, la fonction  $g(t; \theta)$  n'est pas linéaire par rapport à  $\theta$ , même après transformation. On entre dans le cadre de la régression non linéaire dont on trouvera une présentation très complète dans [ABC92]. Une méthode élémentaire, dite des trois points, est donnée dans [MEL90]. La forme non linéaire ne sera donc pas traitée ici puisque seules les fonctions envisagées sont de la forme,

$$g(t; \theta) = \sum_{k=1}^r \theta_k g_k(t),$$

où les fonctions  $g_k(t)$ ,  $k = 1, \dots, r$  sont fixées et connues.

L'autre point spécifique aux séries chronologiques est que la moyenne  $g(t; \theta)$  est souvent composée d'une tendance et d'une partie saisonnière qu'il est nécessaire de bien identifier.

Le premier paragraphe est consacré au modèle de Buys-Ballot dans lequel la moyenne est composée d'une tendance linéaire et d'un mouvement saisonnier rigoureusement périodique. L'étude complète de ce cas très particulier fournit une première approche des problèmes que l'on résout de façon simple.

Le cas général constitue le second paragraphe. Le modèle linéaire, dont nous rappelons les propriétés principales, permet de résoudre toutes les situations clairement définies.

Le lecteur désirant approfondir et compléter les points abordés dans ce chapitre peut se référer aux ouvrages [GM90] ou [MEL90], pour une approche pratique et illustrée des problèmes, [AND71], pour une analyse très complète des propriétés statistiques des méthodes utilisées ainsi que pour les résultats asymptotiques et [ABC92] pour les points liés spécifiquement à la régression.

## 2.1 MODÈLE DE BUYS-BALLOT

Le modèle de *Buys-Ballot* fournit un schéma additif simple que l'on peut traiter très complètement par des méthodes élémentaires. La tendance est représentée par une droite, l'effet saisonnier est rigoureusement périodique de période  $p$  connue et la partie résiduelle est une suite de variables aléatoires indépendantes identiquement distribuées de loi normale centrée et de variance  $\sigma^2$  :

$$Y_t = \alpha t + \beta + S_t + \varepsilon_t, t = 1, \dots, T; \quad S_t = S_{t+p}; \quad \varepsilon_t \sim i.i.d. \mathcal{N}(0; \sigma^2).$$

### 2.1.1 Table de Buys-Ballot

La simplicité des calculs est obtenue en supposant que la série est observée pendant  $n$  "années" de  $p$  "mois" :

$$Y_{ij} = \alpha[p(i-1)+j] + \beta_j + \varepsilon_{ij}, \quad \beta_j = \beta + S_j, j = 1, \dots, p, i = 1, \dots, n; \quad \sum_{j=1}^p S_j = 0.$$

Les coefficients saisonniers  $S_j, j = 1, \dots, p$ , caractérisent la composante périodique  $S_t$  dont l'effet annuel moyen est nul. Ainsi la partie déterministe du modèle est décrite par  $p+1$  paramètres linéairement indépendants :  $\alpha, \beta_1, \dots, \beta_p$ . La méthode des moindres carrés consiste à chercher, parmi les chroniques  $x_{ij}(a, b_1, \dots, b_p) = a[p(i-1)+j] + b_j$ , composées d'une tendance linéaire et d'un mouvement saisonnier périodique, celle qui est la plus proche de l'observation selon le critère :

$$\min_{a, b_1, \dots, b_p} \frac{1}{np} \sum_{i=1}^n \sum_{j=1}^p [y_{ij} - x_{ij}(a, b_1, \dots, b_p)]^2.$$

En d'autres termes elle retient les paramètres pour lesquels la moyenne des carrés des erreurs observées est minimum.

Notons plus simplement  $x_{ij} = x_{ij}(a, b_1, \dots, b_p)$  et introduisons les moyennes mensuelles :

$$\bar{y}_{.j} = \frac{1}{n} \sum_{i=1}^n y_{ij}, \quad \bar{x}_{.j} = \frac{1}{n} \sum_{i=1}^n x_{ij} = a[p(n-1)/2 + j] + b_j, \quad j = 1, \dots, p.$$

Le critère à minimiser se scinde en deux parties :

$$\begin{aligned} & \frac{1}{np} \sum_{i,j} [y_{ij} - x_{ij}]^2 = \frac{1}{p} \sum_j \{ \bar{y}_{.j} - a[p(n-1)/2 + j] - b_j \}^2 \\ & + \frac{1}{np} \sum_{i,j} [y_{ij} - \bar{y}_{.j}]^2 + \frac{1}{np} \sum_{i,j} [x_{ij} - \bar{x}_{.j}]^2 - \frac{2}{np} \sum_{i,j} [y_{ij} - \bar{y}_{.j}][x_{ij} - \bar{x}_{.j}]. \end{aligned}$$

Ainsi, pour une pente  $a$  fixée quelconque, le minimum par rapport à  $b_j$  est réalisé en annulant le premier terme et équivaut à écrire que les chroniques  $y_{ij}$  et  $x_{ij}$  ont mêmes moyennes mensuelles :

$$b_j = \bar{y}_{.j} - a[p(n-1)/2 + j] \iff \bar{x}_{.j} = \bar{y}_{.j}, \quad j = 1, \dots, p.$$

Soient  $t_{ij} = p(i-1) + j$  les dates d'observation et introduisons les moyennes annuelles ainsi que les moyennes globales :

$$\bar{y}_{.i} = \frac{1}{p} \sum_{j=1}^p y_{ij}, \quad \bar{t}_{.i} = \frac{1}{p} \sum_{j=1}^p t_{ij} = p(i-1) + (p+1)/2, \quad i = 1, \dots, n,$$

$$\bar{y}_{..} = \frac{1}{np} \sum_{i,j} y_{ij} = \frac{1}{n} \sum_i \bar{y}_{.i} = \frac{1}{p} \sum_j \bar{y}_{.j}, \quad \bar{t}_{..} = \frac{1}{np} \sum_{i,j} t_{ij} = (np+1)/2.$$

Tenant compte de la solution obtenue pour les variables  $b_j$  et de la relation  $x_{ij} - \bar{x}_{.j} = a(\bar{t}_{.i} - \bar{t}_{..})$ , il reste à minimiser par rapport à  $a$  la deuxième partie du critère :

$$\frac{1}{np} \sum_{i,j} [y_{ij} - \bar{y}_{.j}]^2 + a^2 \text{var}(\bar{t}_{.i}) - 2a \text{cov}(\bar{t}_{.i}, \bar{y}_{.i}),$$

où l'on a posé ( $\sum_{i=1}^n i^2 = n(n+1)(2n+1)/6$ ) :

$$\text{var}(\bar{t}_{.i}) = \frac{1}{n} \sum_i [\bar{t}_{.i} - \bar{t}_{..}]^2 = \frac{p^2(n^2-1)}{12}, \quad \text{cov}(\bar{t}_{.i}, \bar{y}_{.i}) = \frac{1}{n} \sum_i [\bar{t}_{.i} - \bar{t}_{..}][\bar{y}_{.i} - \bar{y}_{..}].$$

L'estimation de la pente  $\alpha$  de la tendance est alors :

$$\hat{\alpha} = \frac{\text{cov}(\bar{t}_i, \bar{y}_i)}{\text{var}(\bar{t}_i)} = \frac{12}{np(n^2 - 1)} \left[ \sum_{i=1}^n i \bar{y}_i - \frac{n(n+1)}{2} \bar{y}_{..} \right].$$

Son report dans la solution pour  $b_j$  donne les estimations des paramètres  $\beta_j$  :

$$\hat{\beta}_j = \bar{y}_{.j} - \hat{\alpha}[p(n-1)/2 + j], \quad j = 1, \dots, p,$$

qui, par centrage, fournissent les estimations de l'ordonnée à l'origine  $\beta$  de la tendance ainsi que celles des coefficients saisonniers  $S_j$  :

$$\hat{\beta} = \frac{1}{p} \sum_{j=1}^p \hat{\beta}_j = \bar{y}_{..} - \hat{\alpha} \bar{t}_{..} = \bar{y}_{..} - \hat{\alpha}(np+1)/2,$$

$$\hat{S}_j = \hat{\beta}_j - \hat{\beta} = \bar{y}_{.j} - \bar{y}_{..} - \hat{\alpha}[j - (p+1)/2], \quad j = 1, \dots, p.$$

En résumé on observe les résultats suivants :

- La tendance ne dépend que des moyennes annuelles, elle est la droite des moindres carrés construite sur les points  $(\bar{t}_i, \bar{y}_i), i = 1, \dots, n$ , c'est-à-dire que  $\hat{\alpha}$  et  $\hat{\beta}$  sont solution du problème de minimisation :

$$\min_{a,b} \frac{1}{n} \sum_{i=1}^n [\bar{y}_i - a\bar{t}_i - b]^2.$$

- La composante saisonnière est définie par les moyennes mensuelles de la chronique privée de la tendance estimée :

$$\hat{S}_j = \frac{1}{n} \sum_{i=1}^n [y_{ij} - \hat{\alpha}[p(i-1) + j] - \hat{\beta}], \quad j = 1, \dots, p.$$

### Illustration

Nous illustrons cette méthode sur la chronique mensuelle du chiffre d'affaires de la presse parisienne dans une petite ville de province ( $\simeq 8000$  h.) de 1981 à 1985. Les données, leurs différentes moyennes et les résultats sont présentés habituellement dans la *Table de Buys-Ballot*. Nous avons également fait figurer les valeurs de la *série ajustée* ainsi que celles de la *prévision* pour l'année 1986 (*cf.* Tableau 2.1). Celles-ci sont évidemment données par :

$$\hat{y}_{ij} = \hat{\alpha}[p(i-1) + j] + \hat{\beta} + \hat{S}_j, \quad j = 1, \dots, p, \quad i = 1, \dots, n+1.$$

La série ajustée correspond aux années  $i = 1, \dots, n$  et la prévision à l'année  $i = n+1$ . La représentation graphique de ces résultats permet de se faire

année	mois	janv	fév	mars	avr	mai	juin	juil	août	sept	oct	nov	déc	moy. an.
	$j$	1	2	3	4	5	6	7	8	9	10	11	12	
1981	1	84	92	90	83	85	100	96	104	107	120	102	105	97
	ajustée	99	93	102	90	86	94	105	101	112	122	94	106	
1982	2	112	112	119	109	109	103	135	111	140	133	123	125	119
	ajustée	116	110	119	107	103	111	122	118	129	139	111	123	
1983	3	139	129	142	123	124	124	140	151	149	147	130	139	136
	ajustée	133	127	136	124	120	128	139	134	146	156	127	139	
1984	4	158	150	171	137	138	145	155	149	155	178	139	156	153
	ajustée	149	143	152	140	136	144	155	151	162	172	144	156	
1985	5	171	150	157	167	142	167	167	157	177	200	143	171	164
	ajustée	166	160	169	157	153	161	172	168	179	189	161	173	
														moy. gén.
	moy. mens.	133	127	136	124	120	128	139	134	146	156	127	139	134
	coef. sais.	7	-1	7	-7	-12	-5	4	-2	8	17	-13	-2	
1986	prévues	183	177	186	174	170	178	189	184	196	206	177	189	

Tendance : pente = 1,39 ; ordonnée à l'origine = 92

TAB. 2.1 – Chiffre d'affaires mensuel de la presse parisienne dans une petite ville de province (unité : 1KF)

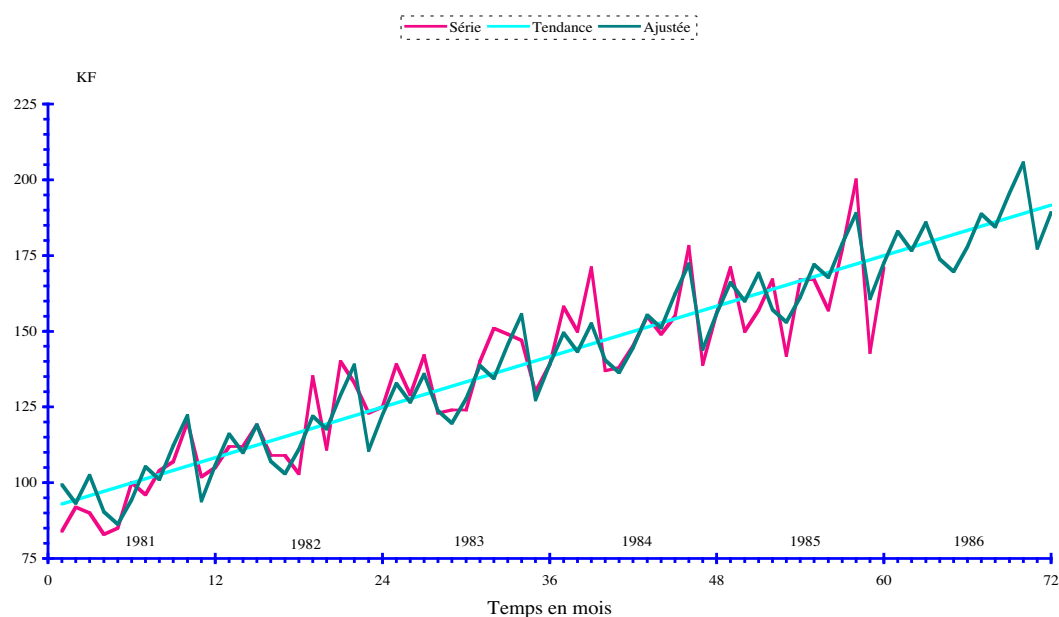


FIG. 2.1 – Chiffre d'affaires mensuel de la presse parisienne dans une petite ville de province

une première idée de la capacité de ce modèle à rendre compte des observations (*cf.* Figure 2.1). On constate que la série observée est au dessous de la série ajustée aux deux extrémités alors qu'elle est au dessus dans la partie centrale. L'étude des résidus (*cf.* Figure 2.2) permet de mieux visualiser ce phénomène et invite à ajuster une tendance parabolique (*cf.* Paragraphe 2.2). Notons aussi que l'ordre de grandeur du mouvement saisonnier est très faible par rapport aux données (environ 10%) bien que la concordance entre les pics et les creux des deux séries soit assez bien respectée. L'étude du mouvement saisonnier dans le cadre de ce modèle (*cf.* Paragraphe 2.1.3) montre, grâce à l'approche numérique, qu'il est effectivement présent avec un mois d'octobre fort et les mois de mai et novembre faibles alors que la représentation graphique n'est pas aussi nette sur ce point.

### 2.1.2 Moyenne et variance des estimateurs

Les propriétés des estimateurs obtenus à la section précédente résultent de celles du modèle linéaire général qui fera l'objet du Paragraphe 2.2. Ils sont sans biais et de variance minimum parmi les estimateurs sans biais qui sont linéaires en les observations. Cependant la particularité du modèle de Buys-Ballot permet de donner des expressions plus explicites de leurs variances.

Pour les paramètres de la tendance, on utilise le modèle linéaire :

$$\bar{Y}_i = \alpha \bar{t}_i + \beta + \bar{\varepsilon}_i, \quad i = 1, \dots, n; \quad \bar{\varepsilon}_i = \frac{1}{p} \sum_{j=1}^p \varepsilon_{ij} \sim i.i.d. \mathcal{N}(0; \sigma^2/p).$$

L'estimateur de la pente s'écrit :

$$\hat{\alpha} = \frac{1}{\text{var}(\bar{t}_i)} \frac{1}{n} \sum_{i=1}^n [\bar{t}_i - \bar{t}_..] \bar{Y}_i = \alpha + \frac{1}{\text{var}(\bar{t}_i)} \frac{1}{n} \sum_{i=1}^n [\bar{t}_i - \bar{t}_..] \bar{\varepsilon}_i.$$

Il est sans biais et sa variance est donnée par :

$$\text{Var}(\hat{\alpha}) = \frac{\sigma^2}{p} \frac{1/n}{\text{var}(\bar{t}_i)} = \frac{\sigma^2}{np} \frac{12}{p^2(n^2 - 1)}.$$

L'estimateur de l'ordonnée à l'origine,

$$\hat{\beta} = \bar{Y}_.. - \hat{\alpha}(np+1)/2 = \beta - (\hat{\alpha} - \alpha)\bar{t}_.. + \bar{\varepsilon}_.. = \beta + \frac{1}{n} \sum_{k=1}^n \left\{ 1 - \frac{[\bar{t}_k - \bar{t}_..]\bar{t}_..}{\text{var}(\bar{t}_i)} \right\} \bar{\varepsilon}_k.,$$

est également sans biais, il est corrélé avec  $\hat{\alpha}$  et, en utilisant les expressions en fonction des  $\bar{\varepsilon}_i$ , ou la non corrélation entre  $\hat{\alpha}$  et  $\bar{Y}_{..}$ , on obtient :

$$Var(\hat{\beta}) = \frac{\sigma^2}{np} \left[ 1 + \frac{3(np+1)^2}{p^2(n^2-1)} \right], \quad Cov(\hat{\alpha}, \hat{\beta}) = -\frac{\sigma^2}{np} \frac{6(np+1)}{p^2(n^2-1)}.$$

Pour les autres estimateurs, on introduit les erreurs mensuelles moyennes :

$$\bar{\varepsilon}_{.j} = \frac{1}{n} \sum_{i=1}^n \varepsilon_{ij} \sim i.i.d. \mathcal{N}(0, \sigma^2/n), \quad j = 1, \dots, p,$$

qui satisfont :

$$Cov(\bar{\varepsilon}_i, \bar{\varepsilon}_j) = Cov(\bar{\varepsilon}_{..}, \bar{\varepsilon}_j) = Cov(\bar{\varepsilon}_{..}, \bar{\varepsilon}_i) = \frac{\sigma^2}{np}, \quad Cov(\bar{\varepsilon}_j - \bar{\varepsilon}_{..}, \bar{\varepsilon}_i) = 0.$$

Les estimateurs des coefficients saisonniers s'écrivent :

$$\hat{S}_j = S_j + (\bar{\varepsilon}_{.j} - \bar{\varepsilon}_{..}) - [j - (p+1)/2](\hat{\alpha} - \alpha), \quad j = 1, \dots, p.$$

Ils sont donc sans biais. D'autre part  $\bar{\varepsilon}_{.j} - \bar{\varepsilon}_{..}$  est non corrélé avec  $\hat{\alpha}$ , d'où :

$$Var(\hat{S}_j) = \frac{\sigma^2}{np} \left[ (p-1) + \frac{12[j - (p+1)/2]^2}{p^2(n^2-1)} \right], \quad j = 1, \dots, p.$$

Notons que  $Var(\hat{S}_j)$  est symétrique par rapport au milieu de l'année, où elle est minimum, et augmente lorsque l'on s'en écarte. Ces estimateurs sont évidemment corrélés entre eux puisque leur somme est nulle et on a :

$$Cov(\hat{S}_j, \hat{S}_k) = -\frac{\sigma^2}{np} \left[ 1 - \frac{12[j - (p+1)/2][k - (p+1)/2]}{p^2(n^2-1)} \right],$$

$$Cov(\hat{S}_j, \hat{\alpha}) = -\frac{\sigma^2}{np} \frac{12[j - (p+1)/2]}{p^2(n^2-1)},$$

$$Cov(\hat{S}_j, \hat{\beta}) = \frac{\sigma^2}{np} \frac{6[j - (p+1)/2](np+1)}{p^2(n^2-1)}.$$

La série ajustée et la prévision s'écrivent :

$$\hat{Y}_{ij} = \alpha t_{ij} + \beta + S_j + \bar{\varepsilon}_{.j} + p[i - (n+1)/2](\hat{\alpha} - \alpha), \quad j = 1, \dots, p, i = 1, \dots, n+1.$$

Ce sont des estimateurs sans biais de la valeur moyenne de la chronique pour chacune des dates considérées et la variance,

$$Var(\hat{Y}_{ij}) = \frac{\sigma^2}{np} \left[ p + 12 \frac{[i - (n+1)/2]^2}{(n^2-1)} \right],$$

ne dépend pas du mois, elle est symétrique par rapport à l'année centrale, où elle est minimum, et augmente lorsque l'on s'en éloigne. Plus généralement on a :

$$\text{Cov}(\hat{Y}_{ij}, \hat{Y}_{kl}) = \frac{\sigma^2}{np} \left[ p\delta_{jl} + 12 \frac{[i - (n+1)/2][k - (n+1)/2]}{(n^2 - 1)} \right],$$

où  $\delta_{jl} = 1$  si  $j = l$ , 0 sinon est le symbole de Kronecker. Enfin l'erreur résultant de ce modèle, appelée *résidu*,

$$\hat{\varepsilon}_{ij} = Y_{ij} - \hat{Y}_{ij} = \varepsilon_{ij} - \bar{\varepsilon}_{.j} - p[i - (n+1)/2](\hat{\alpha} - \alpha),$$

a pour variance

$$\text{Var}(\hat{\varepsilon}_{ij}) = \sigma^2 - \frac{\sigma^2}{np} \left[ p + 12 \frac{[i - (n+1)/2]^2}{(n^2 - 1)} \right] = \text{Var}(Y_{ij}) - \text{Var}(\hat{Y}_{ij}),$$

ce qui confirme la non corrélation entre  $\hat{\varepsilon}_{ij}$  et  $\hat{Y}_{ij}$  en chaque  $(i, j)$ . En fait les vecteurs aléatoires  $\hat{\varepsilon}$  et  $\hat{Y}$  formés par l'ensemble des  $np$  composantes sont non corrélés et vérifient  $\text{Var}(\hat{\varepsilon}) = \text{Var}(Y) - \text{Var}(\hat{Y})$  où  $Y$  est le vecteur représentant les observations.

Tous ces estimateurs ont une variance proportionnelle à  $\sigma^2$ . On peut vérifier avec les résultats ci-dessus que  $\mathbb{E}({}^t\hat{\varepsilon}\hat{\varepsilon}) = (np - p - 1)\sigma^2$ . On dispose ainsi d'un estimateur sans biais de  $\sigma^2$ ,

$$\hat{\sigma}^2 = \frac{1}{np - p - 1} \sum_{i,j} \hat{\varepsilon}_{ij}^2 = \frac{1}{np - p - 1} \sum_{i,j} [Y_{ij} - \hat{\alpha}[p(i-1) + j] - \hat{\beta} - \hat{S}_j]^2,$$

qui est non corrélé avec les estimateurs  $\hat{\alpha}, \hat{\beta}, \hat{S}_j$ , et  $\hat{Y}_{ij}$ .

### 2.1.3 Inférence statistique

Tous les résultats précédents restent vrais lorsque l'on suppose simplement que les erreurs sont centrées, non corrélées et de même variance  $\sigma^2$ . L'hypothèse de normalité des erreurs implique celle des variables  $\hat{\alpha}, \hat{\beta}, \hat{S}_j, \hat{Y}_{ij}$  et  $\hat{\varepsilon}_{ij}$  et la non corrélation équivaut à l'indépendance. La variable  $(np - p - 1)\hat{\sigma}^2/\sigma^2$  suit la loi du khi-deux à  $(np - p - 1)$  degrés de liberté. La loi de Student permet donc de construire des intervalles de confiance pour les paramètres  $\alpha, \beta$  et  $S_j, j = 1, \dots, p$ , ainsi que pour la prévision de  $Y_{n+1,j}$ . Dans ce dernier cas,  $\hat{Y}_{n+1,j}$  représente l'estimation de la moyenne et il faut tenir compte de

la variance de l'erreur pour construire l'intervalle. C'est ainsi que, pour un intervalle de confiance au niveau  $(1 - \alpha)$ , on obtient :

$$Y_{n+1,j} = \hat{Y}_{n+1,j} \pm St_{np-p-1}^{-1}(1 - \alpha/2) \frac{\hat{\sigma}}{\sqrt{np}} \sqrt{(n+1)\left[p + \frac{3}{n-1}\right]},$$

où  $St_{np-p-1}$  désigne la fonction de répartition de la loi de Student à  $np-p-1$  degrés de liberté. Il est clair que la bande de confiance ainsi obtenue en faisant varier  $j$  ne constitue pas une région de confiance de niveau  $(1 - \alpha)$ . Nous pourrions construire selon le même principe une bande de confiance pour la série elle-même afin d'apprécier la validité du modèle. Posant  $t_\alpha = St_{np-p-1}^{-1}(1 - \alpha/2)$ , l'observation de  $Y_{ij}$  sera dans la bande si et seulement si :

$$\left| \frac{\hat{\varepsilon}_{ij}}{\frac{\hat{\sigma}}{\sqrt{np}} \sqrt{\left[p(n+1) + 12 \frac{[i-(n+1)/2]^2}{(n^2-1)}\right]}} \right| \leq t_\alpha.$$

Cet intervalle n'est pas correct car  $Y_{ij}$  est utilisé dans sa construction. Une autre approche consiste à utiliser directement les résidus en négligeant le fait que  $\hat{\varepsilon}_{ij}$  et  $\hat{\sigma}^2$  ne sont pas indépendants :

$$\left| \frac{\hat{\varepsilon}_{ij}}{\frac{\hat{\sigma}}{\sqrt{np}} \sqrt{\left[p(n-1) - 12 \frac{[i-(n+1)/2]^2}{(n^2-1)}\right]}} \right| \leq t_\alpha.$$

Les résidus ainsi normalisés sont appelés *résidus standardisés* et seront notés  $\hat{\varepsilon}_{ij}^S$ . En toute rigueur il faut estimer l'erreur ainsi que l'ensemble des paramètres en supprimant la variable  $Y_{ij}$ . Le nouveau résidu standardisé ainsi obtenu est appelé résidu par *validation croisée* ou *résidu studentisé* et sera noté  $\hat{\varepsilon}_{ij}^V$ . On montre qu'il s'exprime simplement en fonction du précédent ([ABC92], page 33),

$$\hat{\varepsilon}_{ij}^V = \hat{\varepsilon}_{ij}^S \sqrt{\frac{np-p-2}{np-p-1 - \hat{\varepsilon}_{ij}^{S^2}}},$$

et qu'il suit la loi de Student à  $np-p-2$  degrés de liberté.

Le premier test portant sur  $Y_{ij}$  est moins sévère que celui utilisant  $\hat{\varepsilon}_{ij}^S$ . Par contre, il n'y a pas d'ordre systématique entre ceux utilisant  $\hat{\varepsilon}_{ij}^S$  ou  $\hat{\varepsilon}_{ij}^V$ , bien qu'il y ait une relation monotone entre ces deux variables, car le seuil  $t_\alpha$

est plus faible pour  $\hat{\varepsilon}_{ij}^V$  que pour  $\hat{\varepsilon}_{ij}^S$ . Bien souvent la série est suffisamment longue pour que tous ces tests soient équivalents et reviennent à considérer que la variable  $\hat{\varepsilon}_{ij}/\hat{\sigma}$  est  $\mathcal{N}(0; 1)$ . C'est le cas dans l'exemple du chiffre d'affaires de la presse parisienne. Avec  $\alpha = 5\%$  on obtient  $t_\alpha = 2,01$  pour les trois premiers tests et  $t_\alpha = 1,96$  pour l'approximation normale. Les différents résidus sont donnés dans le Tableau 2.2 et la représentation de  $\hat{\varepsilon}_{ij}^V$  fait l'objet de la Figure 2.2.

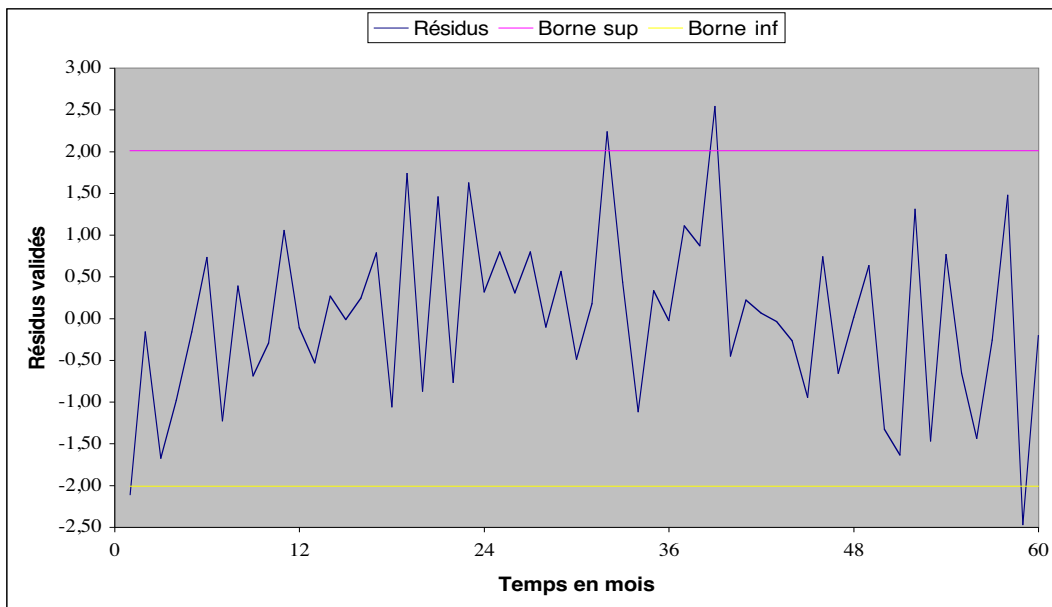


FIG. 2.2 – Chiffre d'affaires mensuel de la presse parisienne dans une petite ville de province : représentation des résidus obtenus par validation croisée

Dans le cadre du modèle de Buys-Ballot, il est naturel de vouloir apprécier l'importance de l'effet saisonnier. La Figure 2.3 donne la représentation graphique de cet effet pour notre illustration.

Pour comparer entre eux les coefficients saisonniers, il faut prendre en compte leurs variances. On définit ainsi les *coefficients studentisés* :

$$\hat{S}_j^S = \frac{\hat{S}_j}{\hat{\sigma}} \left[ \frac{1}{np} \left[ (p-1) + \frac{12[j - (p+1)/2]^2}{p^2(n^2-1)} \right] \right]^{-1/2}, \quad j = 1, \dots, p.$$

Le Tableau 2.3 donne ces résultats dans le cas de notre illustration.

Sous l'hypothèse  $S_j = 0$ , la variable suit la loi de Student à  $np - p - 1$  degrés de liberté. Cela permet de considérer leurs valeurs sur une échelle de

année	mois												
	janv	fév	mars	avr	mai	juin	juil	août	sept	oct	nov	déc	
1981	Résidus												
	Prévus	-1,61	-0,13	-1,29	-0,77	-0,13	0,58	-0,96	0,31	-0,54	-0,23	0,83	-0,09
	Standardisés	-2,04	-0,16	-1,64	-0,98	-0,16	0,74	-1,22	0,39	-0,69	-0,29	1,05	-0,11
	Validés	-2,11	-0,16	-1,67	-0,98	-0,16	0,73	-1,23	0,39	-0,69	-0,29	1,05	-0,11
1982	Normalisés	-1,78	-0,14	-1,44	-0,86	-0,14	0,64	-1,07	0,34	-0,61	-0,26	0,92	-0,10
	Prévus	-0,43	0,22	-0,01	0,20	0,64	-0,85	1,38	-0,71	1,17	-0,62	1,29	0,26
	Standardisés	-0,53	0,27	-0,02	0,24	0,79	-1,05	1,70	-0,87	1,44	-0,77	1,60	0,32
	Validés	-0,53	0,27	-0,01	0,24	0,79	-1,06	1,74	-0,87	1,46	-0,77	1,62	0,32
1983	Normalisés	-0,48	0,24	-0,01	0,22	0,70	-0,94	1,51	-0,78	1,28	-0,68	1,42	0,29
	Prévus	0,65	0,25	0,65	-0,08	0,46	-0,40	0,15	1,75	0,36	-0,91	0,27	-0,02
	Standardisés	0,80	0,31	0,80	-0,10	0,57	-0,49	0,18	2,15	0,44	-1,11	0,34	-0,03
	Validés	0,80	0,31	0,80	-0,10	0,56	-0,49	0,18	2,24	0,44	-1,11	0,33	-0,03
1984	Normalisés	0,72	0,28	0,72	-0,09	0,51	-0,44	0,16	1,92	0,39	-0,99	0,30	-0,02
	Prévus	0,90	0,71	1,95	-0,37	0,18	0,05	-0,03	-0,22	-0,77	0,60	-0,53	0,01
	Standardisés	1,11	0,87	2,41	-0,45	0,22	0,07	-0,04	-0,27	-0,95	0,74	-0,66	0,02
	Validés	1,11	0,87	2,54	-0,45	0,22	0,07	-0,04	-0,27	-0,95	0,74	-0,66	0,01
1985	Normalisés	0,98	0,78	2,14	-0,40	0,20	0,06	-0,03	-0,24	-0,84	0,66	-0,59	0,01
	Prévus	0,50	-1,04	-1,27	1,02	-1,14	0,61	-0,52	-1,12	-0,20	1,15	-1,85	-0,16
	Standardisés	0,64	-1,32	-1,61	1,30	-1,45	0,77	-0,66	-1,42	-0,26	1,46	-2,35	-0,21
	Validés	0,63	-1,33	-1,63	1,31	-1,47	0,77	-0,65	-1,44	-0,26	1,47	-2,47	-0,20
	Normalisés	0,56	-1,15	-1,41	1,14	-1,27	0,67	-0,57	-1,24	-0,23	1,28	-2,05	-0,18

TAB. 2.2 – Chiffre d'affaires mensuel de la presse parisienne dans une petite ville de province : les différentes formes de résidus

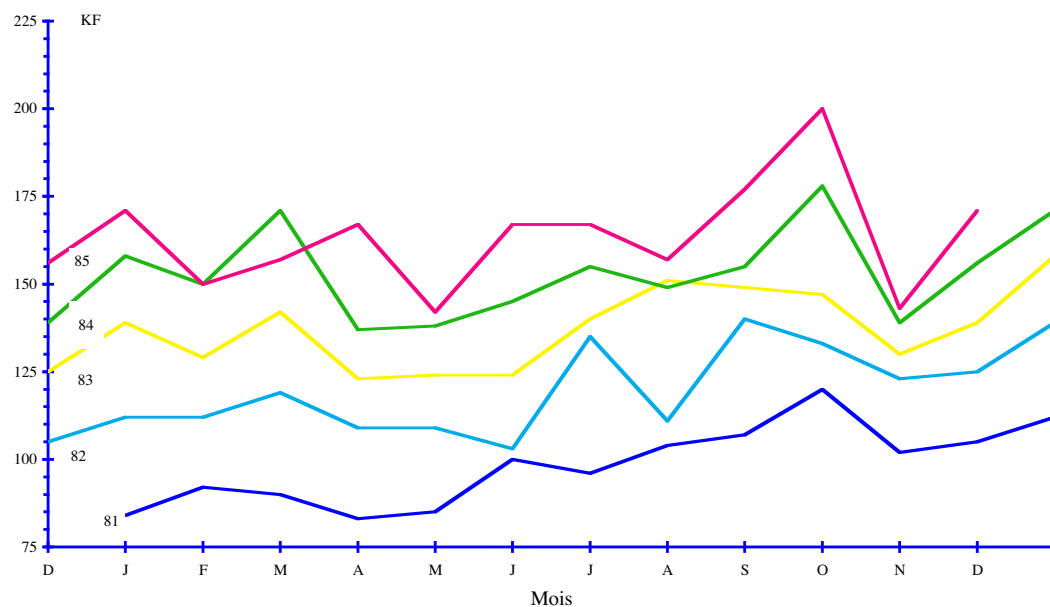


FIG. 2.3 – Mouvement saisonnier du chiffre d'affaires mensuel de la presse parisienne dans une petite ville de province

mesure standard. Rappelons que le niveau de signification du test de  $S_j = 0$  contre  $S_j \neq 0$  ne peut être contrôlé rigoureusement que si un tel test est envisagé pour un seul mois donné avant de consulter les observations. Notons que ce test de Student équivaut à comparer par un test de Fisher-Snedecor le modèle global au sous-modèle défini par l'hypothèse  $S_j = 0$ . Par contre il est possible de tester "systématiquement" la présence ou non de l'effet saisonnier de façon globale. Dans ce cas le sous-modèle est limité à la tendance linéaire. Celle-ci est estimée par la droite des moindres carrés ajustée sur l'ensemble des observations mensuelles :

$$\hat{\alpha}_0 = \frac{\text{cov}(t_{ij}, Y_{ij})}{\text{var}(t_{ij})}, \quad \hat{\beta}_0 = \bar{Y}_{..} - \hat{\alpha}_0 \bar{t}_{..},$$

où

$$\begin{aligned} \text{cov}(t_{ij}, Y_{ij}) &= \frac{1}{np} \sum_{t=1}^{np} t Y_t - \frac{np+1}{2} \bar{Y}_{..}, \\ \text{var}(t_{ij}) &= \frac{1}{np} \sum_{t=1}^{np} t^2 - \left[ \frac{np+1}{2} \right]^2 = \frac{n^2 p^2 - 1}{12}. \end{aligned}$$

L'estimateur sans biais de  $\sigma^2$ , sous cette hypothèse, est

$$\hat{\sigma}_0^2 = \frac{1}{np-2} \sum_{t=1}^{np} [Y_t - \hat{\alpha}_0 t - \hat{\beta}_0]^2 = \frac{\|\hat{\epsilon}_0\|^2}{np-2} = \frac{np}{np-2} \left[ \text{var}(Y_{ij}) - \frac{n^2 p^2 - 1}{12} \hat{\alpha}_0^2 \right],$$

où

$$\text{var}(Y_{ij}) = \frac{1}{np} \sum_{t=1}^{np} [Y_t - \bar{Y}_{..}]^2.$$

Sous l'hypothèse nulle, le rapport de deux estimateurs sans biais de  $\sigma^2$ ,

$$\frac{\tilde{\sigma}^2}{\hat{\sigma}^2} = \frac{(np-p-1)[\|\hat{\epsilon}_0\|^2 - \|\hat{\epsilon}\|^2]}{(p-1)\|\hat{\epsilon}\|^2}$$

suit la loi de Fisher-Snedecor à  $(p-1, np-p-1)$  degrés de liberté. Les résultats de ce test pour notre illustration sont reportés dans le Tableau 2.4.

## 2.2 MODÈLE LINÉAIRE GÉNÉRAL

### 2.2.1 Présentation du modèle

Le modèle de Buys-Ballot est un cas particulier du modèle suivant,

$$Y_t = f(t) + S(t) + \varepsilon_t = \sum_{j=0}^q \alpha_j \phi_j(t) + \sum_{j=1}^p \beta_j \psi_j(t) + \varepsilon_t, \quad t = 1, \dots, T,$$

dans lequel  $\phi_j(t)$  et  $\psi_j(t)$  sont des fonctions connues du temps, utilisées pour modéliser respectivement la tendance  $f(t)$  et la composante saisonnière  $S(t)$  à l'aide de paramètres naturels  $\alpha_j$  et  $\beta_j$ , l'erreur  $\varepsilon_t$  étant un bruit blanc gaussien. En effet la tendance linéaire est obtenue avec  $\phi_0(t) = 1$  et  $\phi_1(t) = t$ . Pour la partie périodique de période  $p$ , nous introduisons les fonctions

$$\psi_j(t) = \delta_j(t) = 1 \text{ si } t = j \text{ mod}(p), 0 \text{ sinon, } \quad t = 1, \dots, T, \quad j = 1, \dots, p.$$

Les paramètres naturels sont  $\alpha_0 = \beta$ ,  $\alpha_1 = \alpha$  et  $\beta_j = S_j$ ,  $j = 1, \dots, p$ . Cependant pour que le modèle soit identifiable, nous avons imposé la contrainte  $S_1 + \dots + S_p = 0$ . En général les fonctions  $\phi_j(t)$ ,  $j = 0, \dots, q$  constituent  $q + 1$  vecteurs linéairement indépendants de  $\mathbb{R}^T$  et il en est de même pour les  $p$  fonctions  $\psi_j$ ,  $j = 1, \dots, p$ . Par contre l'intersection des sous-espaces de dimensions  $q + 1$  et  $p$  ainsi définis est rarement réduite au vecteur nul. En d'autres termes, les deux systèmes de fonctions engendrent un sous-espace de  $\mathbb{R}^T$  de dimension  $r \leq p + q + 1$ . Il sera donc nécessaire d'introduire  $p + q + 1 - r$  contraintes sur les paramètres naturels. Cela conduit à restructurer le modèle sous la forme,

$$Y_t = f(t) + S(t) + \varepsilon_t = \sum_{j=1}^r \theta_j x_j(t) + \varepsilon_t, \quad t = 1, \dots, T,$$

où les fonctions  $x_j(t)$ ,  $j = 1, \dots, r$ , combinaisons linéaires des fonctions  $\phi_j(t)$  et  $\psi_j(t)$  initiales, définissent  $r$  vecteurs linéairement indépendants de  $\mathbb{R}^T$ .

Par exemple une généralisation immédiate du modèle de Buys-Ballot consiste à poser :

$$f(t) = \sum_{k=0}^q \alpha_k t^k, \quad S(t) = \sum_{j=1}^p \left[ \sum_{k=0}^s \beta_{jk} t^k \right] \delta_j(t) = \sum_{j=1}^p S_j(t) \delta_j(t), \quad t = 1, \dots, T.$$

Dans ce cas la tendance est un polynôme de degré  $q$  et les "coefficients saisonniers" évoluent au cours du temps comme un polynôme de degré  $s$ . La

relation

$$\sum_{j=1}^p \delta_j(t) = 1, \quad t = 1, \dots, T,$$

conduit aux contraintes

$$\sum_{j=1}^p \beta_{jk} = 0, \quad k = 0, \dots, \min(s, q),$$

qui seront suffisantes si  $T$  est assez grand. Notons qu'il n'est plus nécessaire de supposer que  $T$  soit un multiple de la période.

De façon générale le modèle initial, en fonction des paramètres naturels, et sa restructuration se décrivent sous forme vectorielle par :

$$Y = \tilde{X}\tilde{\theta} + \varepsilon = X\theta + \varepsilon, \quad \tilde{X} = XG, \quad \tilde{\theta} = G^{-1}\theta, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I_T).$$

Dans cette écriture  $Y = {}^t[Y_1, \dots, Y_T]$  est le vecteur des observations,  $\varepsilon = {}^t[\varepsilon_1, \dots, \varepsilon_T]$  est celui des erreurs,  $I_T$  étant la matrice identité d'ordre  $T$ , et  $\tilde{\theta} = {}^t[\alpha_0, \dots, \alpha_q, \beta_1, \dots, \beta_p]$  est le vecteur des paramètres naturels. La matrice  $\tilde{X}$ , de dimension  $T \times (p + q + 1)$ , représente le système des fonctions initiales  $\phi_j(t)$  et  $\psi_j(t)$  et  $X$ , de dimensions  $T \times r$ , celui des nouvelles fonctions  $x_j(t)$ . La matrice  $G$ , de dimension  $r \times (p + q + 1)$ , est alors parfaitement définie. Le choix de l'inverse généralisé  $G^-$  traduit les contraintes sur  $\tilde{\theta}$  assurant son identification à partir du paramètre  $\theta = {}^t[\theta_1, \dots, \theta_r]$  du modèle. Le choix des fonctions  $x_j(t)$  peut être guidé par l'interprétation du paramètre  $\theta$  qui en résulte. Par exemple, on pourra faire en sorte que l'absence d'une composante (saisonniers ou tendance) se traduise par la nullité des dernières composantes de  $\theta$ .

### 2.2.2 Rappels sur le modèle linéaire

On considère le modèle

$$Y = X\theta + \varepsilon, \quad \mathbb{E}(\varepsilon) = 0, \quad \text{Var}(\varepsilon) = \sigma^2 I_T,$$

dans lequel  $Y$  est le vecteur aléatoire de dimension  $T$  représentant les observations,  $X$  est une matrice connue de dimension  $T \times r$  et de rang  $r < T$ , appelée matrice des régresseurs (ou *plan d'expérience*),  $\theta$  est un paramètre vectoriel de dimension  $r$  et  $\varepsilon$  est le vecteur aléatoire de dimension  $T$  représentant les erreurs. On a  $\mathbb{E}(Y) = X\theta$  et  $\text{Var}(Y) = \sigma^2 I_T$ . Pour ne pas alourdir les notations, nous utiliserons souvent le même symbole pour désigner un élément

aléatoire et l'observation de cet élément, sinon la lettre minuscule  $y$  représente l'observation de son homologue majuscule  $Y$ .

L'estimation des *moindres carrés* de  $\theta$  est obtenue en minimisant, par rapport à  $a$ , la somme des carrés des écarts entre les observations  $y_t$  et leurs moyennes de la forme  $(Xa)_t$  :

$$\min_a \|y - Xa\|^2 = \min_a \sum_{t=1}^T \left[ y_t - \sum_{j=1}^r a_j x_{tj} \right]^2.$$

L'estimateur  $\hat{\theta}$  qui en résulte est solution du système des équations dites normales,

$${}^t X X \hat{\theta} - {}^t X Y = 0, \quad \Rightarrow \quad \hat{\theta} = ({}^t X X)^{-1} {}^t X Y,$$

et vérifie :

$$\mathbb{E}(\hat{\theta}) = \theta, \quad \text{Var}(\hat{\theta}) = \sigma^2 ({}^t X X)^{-1}.$$

Il est donc sans biais. On montre qu'il est de variance minimum parmi les estimateurs linéaires en les observations et sans biais, on dit alors que c'est un estimateur de *Gauss-Markov*. Lorsque l'erreur  $\varepsilon$  est de loi gaussienne, la propriété reste vraie sans la contrainte de linéarité,  $\hat{\theta}$  est aussi l'estimateur de maximum de vraisemblance et sa loi est gaussienne. Si  $\phi = A\theta$  est une transformation linéaire connue de  $\theta$ , l'estimateur  $\hat{\phi} = A\hat{\theta}$  a les mêmes propriétés et  $\text{Var}(\hat{\phi}) = \sigma^2 A ({}^t X X)^{-1} {}^t A$  quelque soit la matrice  $A$ . C'est en particulier le cas de la moyenne  $X\theta$  de  $Y$  dont l'estimateur est noté improprement  $\hat{Y} = X\hat{\theta}$ . L'estimateur de l'erreur non observable  $\varepsilon$ , appelé *vecteur des résidus*, est  $\hat{\varepsilon} = Y - \hat{Y}$ , il est centré et non corrélé avec  $\hat{Y}$  ou  $\hat{\theta}$ . Soit  $H = X ({}^t X X)^{-1} {}^t X$  la matrice, définissant, dans l'espace des observations  $\mathbb{R}^T$ , la projection orthogonale sur l'espace  $\mathcal{M}(X)$  engendré par les colonnes de  $X$ , dit *espace des moyennes*. Elle est appelée *matrice chapeau* ou *matrice de prédiction* et on a :

$$\hat{Y} = X\hat{\theta} = HY, \quad \mathbb{E}(\hat{Y}) = X\theta, \quad \text{Var}(\hat{Y}) = \sigma^2 H,$$

$$\hat{\varepsilon} = Y - \hat{Y} = (I - H)Y, \quad \mathbb{E}(\hat{\varepsilon}) = 0, \quad \text{Var}(\hat{\varepsilon}) = \sigma^2 (I - H), \quad \text{Cov}(\hat{\varepsilon}, \hat{Y}) = 0.$$

L'estimateur  $\hat{Y}$  de la moyenne, et par suite le vecteur des résidus  $\hat{\varepsilon}$ , ne dépendent pas du choix du paramètre  $\theta$  mais uniquement du sous-espace des moyennes  $\mathcal{M}(X)$ . Tous ces estimateurs ont une variance proportionnelle à  $\sigma^2$  qui en général est inconnu. Son estimateur sans biais est :

$$\hat{\sigma}^2 = \frac{\|\hat{\varepsilon}\|^2}{T - r} = \frac{1}{T - r} \sum_{t=1}^T \hat{\varepsilon}_t^2,$$

et  $\frac{(T-r)\hat{\sigma}^2}{\sigma^2}$  suit la loi du khi-deux à  $T - r$  degrés de liberté lorsque  $\varepsilon$  est gaussien. Notons que l'estimateur de maximum de vraisemblance  $\|\hat{\varepsilon}\|^2/T$  est biaisé et que  $\|\hat{\varepsilon}\|^2$  est la valeur minimum du critère des moindres carrés.

Le vecteur des résidus  $\hat{\varepsilon}$  est centré mais ses composantes sont en général corrélées. De plus elles n'ont pas la même variance puisque  $Var(\hat{\varepsilon}_t) = \sigma^2(1 - h_{tt})$  où  $h_{tt}, t = 1, \dots, T$  sont les termes diagonaux de  $H$ . On appelle *résidus standardisés* les valeurs

$$\hat{\varepsilon}_t^S = \frac{\hat{\varepsilon}_t}{\hat{\sigma}\sqrt{1 - h_{tt}}}, \quad t = 1, \dots, T,$$

qui suivent approximativement la loi de Student à  $T - r$  degrés de liberté. Le résultat n'est pas exact car  $\hat{\sigma}$  et  $\hat{\varepsilon}_t$  ne sont pas indépendants. Une approche plus rigoureuse consiste à estimer les paramètres du modèle et l'erreur qui en découle en se privant de l'observation de  $Y_t$ . En notant avec un indice ( $t$ ) entre parenthèses les quantités qui résultent de cette suppression et  ${}^t x_t$  désignant la  $t^e$  ligne de la matrice  $X$ , on montre que (cf. [ABC92], page 30) :

$$\hat{\theta}_{(t)} = \hat{\theta} - ({}^t X X)^{-1} x_t \frac{\hat{\varepsilon}_t}{1 - h_{tt}}, \quad \hat{\varepsilon}_{(t)} = Y_t - \hat{Y}_{(t)} = \frac{\hat{\varepsilon}_t}{1 - h_{tt}},$$

$$(T - r - 1)\hat{\sigma}_{(t)}^2 = (T - r)\hat{\sigma}^2 - \frac{\hat{\varepsilon}_t^2}{1 - h_{tt}}.$$

Le résidu proprement normalisé est appelé résidu par *validation croisée* ou *résidu studentisé*. Noté  $\hat{\varepsilon}_t^V$ , il vérifie :

$$\hat{\varepsilon}_t^V = \frac{\hat{\varepsilon}_{(t)}}{\hat{\sigma}_{(t)}\sqrt{1 + {}^t x_t ({}^t X X)^{-1} x_t}} = \frac{\hat{\varepsilon}_t}{\hat{\sigma}_{(t)}\sqrt{1 - h_{tt}}} = \hat{\varepsilon}_t^S \sqrt{\frac{T - r - 1}{T - r - \hat{\varepsilon}_t^{S^2}}},$$

et suit la loi de Student à  $T - r - 1$  degrés de liberté. Ces résidus sont utilisés pour détecter les valeurs aberrantes. Ils sont évidemment corrélés mais leur représentation en fonction du temps permet de déceler une erreur de modèle importante.

Pour mieux interpréter  $\hat{\theta}_j$  on peut mesurer la diminution de l'erreur due à la variable  $X_j$ . Si  $\hat{\varepsilon}_{[j]}$  désigne l'erreur obtenue sans cette variable et  $x_{[j]}$  l'erreur de la régression de  $X_j$  sur les autres colonnes de  $X$ , on a :

$$\hat{\varepsilon} = \hat{\varepsilon}_{[j]} - \hat{\theta}_j x_{[j]}, \quad \|\hat{\varepsilon}\|^2 = \|\hat{\varepsilon}_{[j]}\|^2 - \hat{\theta}_j^2 \|x_{[j]}\|^2,$$

$$\hat{\theta}_j = \frac{\langle \hat{\varepsilon}_{[j]}, x_{[j]} \rangle}{\|x_{[j]}\|^2} = \frac{\langle \hat{\varepsilon}_{[j]}, x_{[j]} \rangle}{\|\hat{\varepsilon}_{[j]}\| \times \|x_{[j]}\|} \frac{\|\hat{\varepsilon}_{[j]}\|}{\|x_{[j]}\|}.$$

Le cosinus d'angle qui apparaît dans la dernière expression de  $\hat{\theta}_j$  ci-dessus est la “corrélation partielle” entre  $Y$  et  $X_j$  dans l'ensemble  $\{Y, X_1, \dots, X_p\}$ , c'est-à-dire la “corrélation” qui subsiste entre  $Y$  et  $X_j$  lorsque l'on a éliminé celle due aux variables intermédiaires  $X_k, k \neq j$ .

Sous l'hypothèse gaussienne, il est possible de construire des intervalles de confiance pour les grandeurs scalaires. La considération simultanée de plusieurs d'entre eux est délicate, car ils ne sont en général pas indépendants, et la région de confiance parallélépipédique obtenue n'est pas optimale même en cas d'indépendance. Par exemple la région de confiance au niveau  $(1 - \alpha)$  pour le paramètre vectoriel  $\theta$  est l'ellipsoïde

$${}^t(\theta - \hat{\theta}) {}^t X X (\theta - \hat{\theta}) \leq \hat{\sigma}^2 r F_{r, T-r}^{-1}(1 - \alpha),$$

où  $F_{r, T-r}$  est la fonction de répartition de la loi de Fisher-Snedecor à  $r$  et  $T - r$  degrés de liberté.

Enfin pour tester qu'une partie des composantes de  $\theta$  est nulle, ou plus généralement que la moyenne de  $Y$  appartient à un sous-espace  $\mathcal{W}$  de  $\mathcal{M}(X)$  de dimension  $s < r$ , on utilise la statistique

$$\frac{T - r}{r - s} \frac{\|\hat{Y} - \hat{Y}_{\mathcal{W}}\|^2}{\|Y - \hat{Y}\|^2} = \frac{T - r}{r - s} \frac{\|\hat{\varepsilon}_{\mathcal{W}}\|^2 - \|\hat{\varepsilon}\|^2}{\|\hat{\varepsilon}\|^2},$$

où  $\hat{Y}_{\mathcal{W}}$  est l'estimation de la moyenne de  $Y$  dans  $\mathcal{W}$  et  $\hat{\varepsilon}_{\mathcal{W}}$  est le résidu correspondant. Cette statistique suit, sous l'hypothèse nulle, la loi de Fisher-Snedecor à  $(r - s, T - r)$  degrés de liberté.

Lorsque les erreurs sont corrélées entre elles d'une manière connue, c'est-à-dire que l'on a  $Var(\varepsilon) = \sigma^2 \Sigma$ , où  $\Sigma$  est une matrice connue et  $\sigma^2$  un facteur scalaire inconnu, la transformation  $Z = \Sigma^{-1/2} Y$  permet de retrouver le modèle précédent. Cela équivaut à utiliser la géométrie dans  $\mathbb{R}^T$  définie par  $\Sigma^{-1}$ . On parle de *moindres carrés généralisés* ou de *moindres carrés pondérés* lorsque  $\Sigma^{-1}$  est diagonale. Sans tenir compte de cette corrélation, les estimateurs des moindres carrés ordinaires restent sans biais mais n'ont plus, en général, la propriété de Gauss-Markov.

### Illustration

Les résultats concernant l'ajustement d'un modèle avec une tendance quadratique (noté modèle quadratique) sur le chiffre d'affaires de la presse parisienne sont indiqués dans les Tableaux 2.5 et 2.6 ainsi que sur la Figure 2.4.

La représentation graphique des résidus normalisés par l'approximation gaussienne est donnée dans la Figure 2.5 et les tests concernant la présence de l'effet saisonnier ou la nécessité d'introduire le terme quadratique dans la tendance sont regroupés dans le Tableau 2.7.

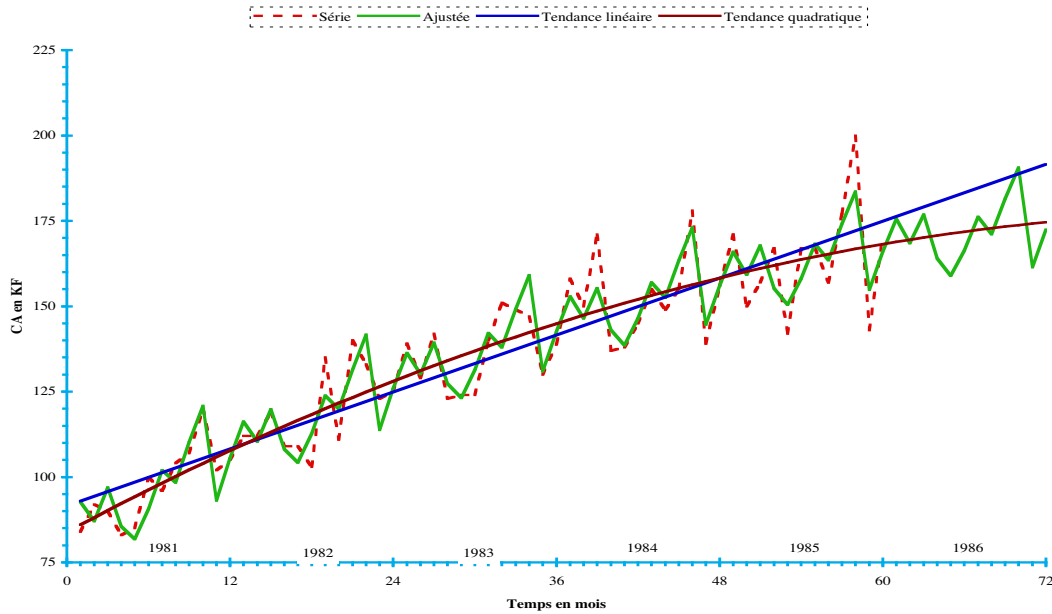


FIG. 2.4 – Chiffre d'affaires de la presse parisienne dans une petite ville de province, modèle quadratique

### Remarque

Pour être complet il faudrait s'assurer que l'hypothèse de non corrélation des erreurs est satisfaite. On sait que les résidus sont corrélés. Le test de Durbin-Watson (*cf.* [GM90], page 45) concerne l'alternative d'une autocorrélation d'ordre 1. Pour une étude plus complète du problème, voir [AND71], section 10.4.

<i>mois</i>	janv	fév	mars	avr	mai	juin	juil	août	sept	oct	nov	déc
<i>Coef. réels</i>	6,5	-1,1	6,7	-6,7	-12,2	-5,4	4,0	-1,6	8,2	16,8	-12,8	-2,4
<i>Coef. studentisés</i>	1,8	-0,3	1,8	-1,8	-3,3	-1,5	1,1	-0,4	2,2	4,5	-3,4	-0,6

Seuil du test de Student de niveau 5% à 47 d.l. : 2,01

TAB. 2.3 – Coefficients saisonniers du chiffre d'affaires mensuel de la presse parisienne dans une petite ville de province

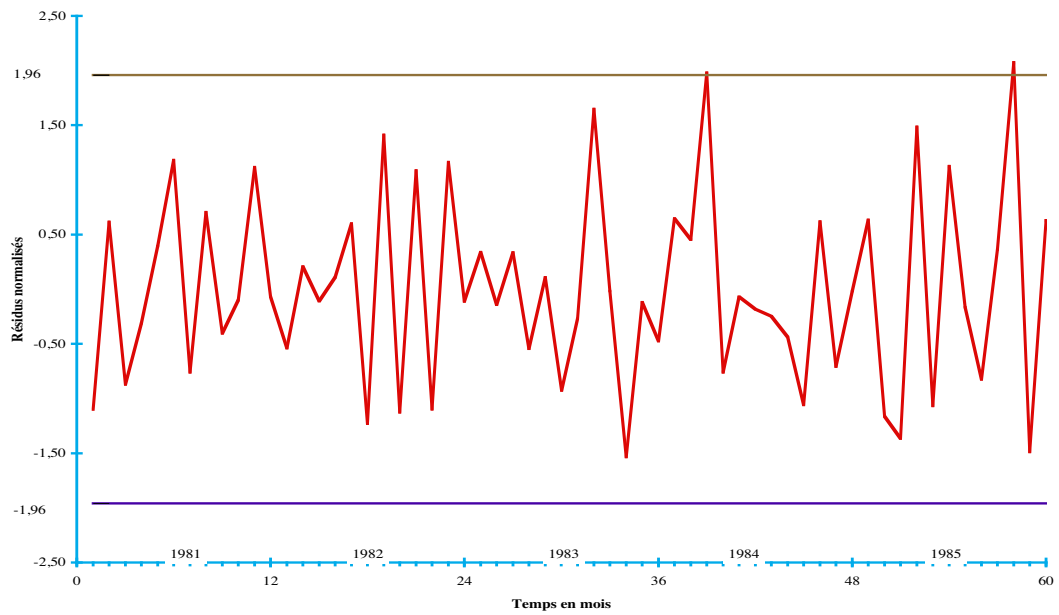


FIG. 2.5 – Résidus normalisés du chiffre d'affaires de la presse parisienne dans une petite ville de province, modèle quadratique

Modèle	pente	ordonnée à l'origine	écart-type de l'erreur	variance de l'erreur
Avec effet saisonnier	1,390	91,5	8,6	74,82
Sans effet saisonnier	1,388	91,6	11,6	133,88

$var(Y) = 707,13$ ;  $cov(t, Y) = 416,25$ ; statistique du test observée = 5,16

Seuil du test de Fisher-Snedecor de niveau 5% à 11 et 47 d.l. : 1,98

TAB. 2.4 – Test de l'effet saisonnier du chiffre d'affaires mensuel de la presse parisienne dans une petite ville de province

année	mois	janv	fév	mars	avr	mai	juin	juil	août	sept	oct	nov	déc
	<i>j</i>	1	2	3	4	5	6	7	8	9	10	11	12
1981	1	84	92	90	83	85	100	96	104	107	120	102	105
	ajustée	93	87	97	85	82	91	102	98	110	121	93	106
1982	2	112	112	119	109	109	103	135	111	140	133	123	125
	ajustée	116	110	120	108	104	113	124	120	131	142	114	126
1983	3	139	129	142	123	124	124	140	151	149	147	130	139
	ajustée	136	130	139	127	123	131	142	138	149	159	131	143
1984	4	158	150	171	137	138	145	155	149	155	178	139	156
	ajustée	153	146	155	143	139	146	157	152	163	173	145	156
1985	5	171	150	157	167	142	167	167	157	177	200	143	171
	ajustée	166	159	168	155	150	158	168	164	174	184	155	166
	coef. sais.	7	-1	7	-7	-12	-6	4	-2	8	17	-13	-2
1986	prévues	176	169	177	164	159	166	176	171	181	191	161	172

Tendance : "pente" = 2,13; "ordonnée à l'origine" = 84; terme quadratique = -0,0121

TAB. 2.5 – Chiffre d'affaires mensuel de la presse parisienne dans une petite ville de province, modèle quadratique (unité : 1KF)

mois	janv	fév	mars	avr	mai	juin	juil	août	sept	oct	nov	déc
Coef. réels	6,7	-1,0	6,7	-6,7	-12,4	-5,6	3,8	-1,7	8,1	16,8	-12,7	-2,2
Coef. studentisés	2,0	-0,3	2,0	-2,0	-3,6	-1,6	1,1	-0,5	2,4	4,9	-3,7	-0,6

Seuil du test de Student de niveau 5% à 46 d.l. : 2,01

TAB. 2.6 – Coefficients saisonniers du chiffre d'affaires mensuel de la presse parisienne dans une petite ville de province, modèle quadratique

Paramètres du modèle	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\sigma}$	$\hat{\sigma}^2$
Avec effet saisonnier	83,9	2,131	-0,0121	7,9	62,65
Sans effet saisonnier	84,0	2,120	-0,0120	11,2	125,35

Seuil du test de Fisher-Snedecor de niveau 5% à 11 et 46 d.l.	1,97
Statistique du test observée pour la présence du mouvement saisonnier	6,19
Seuil du test de Fisher-Snedecor de niveau 5% à 1 et 46 d.l.	4,06
Statistique du test observée pour la présence du terme quadratique	10,14

TAB. 2.7 – Tests de l'effet saisonnier et de la présence du terme quadratique pour le chiffre d'affaires mensuel de la presse parisienne dans une petite ville de province, modèle quadratique



## Chapitre 3

# LES MÉTHODES DE LISSAGE

Ce chapitre aborde les mêmes points que le précédent, recherche de la tendance, de la composante saisonnière et prévision, dans un cadre non paramétrique. Nous conservons le principe du schéma additif :

$$Y_t = f_t + S_t + \varepsilon_t, \quad t = 1, \dots, T, \quad \varepsilon_t \sim b.b.(\sigma^2).$$

La *composante saisonnière*  $S_t$ , lorsqu'elle est présente, est souvent supposée rigoureusement périodique de période  $p$  connue et définie par les *coefficients saisonniers*  $S_j, j = 1, \dots, p$ , satisfaisant la contrainte  $\sum_{j=1}^p S_j = 0$ . Cependant on cherchera parfois à prendre en compte une évolution lente de cet effet saisonnier. La différence fondamentale avec le chapitre précédent est de ne pas structurer la *tendance*  $f_t$ . Celle-ci est une fonction non paramétrique à variation lente. Il n'est donc pas nécessaire d'imposer de fortes contraintes à la partie aléatoire  $\varepsilon_t$  car aucune étude statistique fine ne peut être envisagée dans ce cadre.

En fait, l'approche est ici beaucoup plus descriptive et l'un des buts principaux est de désaisonnaliser la série observée. Ceci consiste à éliminer l'effet saisonnier, ou plus exactement à le répartir de façon uniforme à l'intérieur de chaque période. Pour cela il suffit de retrancher à la série brute  $Y_t$  la composante saisonnière  $S_t$ , dont l'effet est en moyenne nul sur une période. La nouvelle série ainsi obtenue, dite *désaisonnalisée* ou *Corrigée des Variations Saisonnières* (*série CVS*) et notée  $Y_t^c$ , constitue une série artificielle sans effet saisonnier conservant tout le potentiel de la série initiale. Elle conserve en particulier la partie résiduelle  $\varepsilon_t$  et ne doit pas être confondue avec la tendance. L'intérêt est de rendre comparables deux valeurs consécutives. Le principe de la méthode consiste à retrancher la tendance, estimée par lissage

de la série brute, puis à estimer la composante saisonnière sur la nouvelle série ainsi obtenue. Cette estimation est réalisée de façon paramétrique ou par une seconde opération de lissage. On comprend qu'il n'est pas nécessaire, pour un tel objectif, de contraindre la tendance à une forme paramétrique. Par ailleurs son estimation sous forme de fonction lisse permet de localiser les différentes phases du *mouvement conjoncturel* (croissance, stagnation,...).

Les opérations de lissage évoquées jusqu'ici sont réalisées par le biais de moyennes mobiles. Une chronique  $Y_t, t = 1, \dots, T$  est lissée en remplaçant chaque valeur  $Y_t$  par une moyenne pondérée des valeurs qui l'entourent :

$$\tilde{Y}_t = \sum_{j=-l}^k \gamma_j Y_{t-j}, \quad t = k+1, \dots, T-l.$$

Cette écriture regroupe une très grande diversité dans la mesure où aucune restriction n'est faite sur les coefficients  $\gamma_j$ . De simples *moyennes arithmétiques* permettent, par composition, de décrire une classe importante de moyennes mobiles qui s'avère suffisante pour la construction de séries *CVS* dans la plupart des situations. Dans ce cas les coefficients sont simples, positifs et symétriques ( $\gamma_{-j} = \gamma_j$ ). Plus généralement les coefficients  $\gamma_j$  sont déterminés afin de réaliser de façon optimale, selon différents critères, l'élimination d'une ou deux composantes sans détériorer les autres. Ils ne sont plus nécessairement positifs. On parle de *filtrage* de la chronique et le filtre est dit *causal* lorsque  $\gamma_j = 0$  pour  $j < 0$ .

L'inconvénient de l'approche non paramétrique est de ne pas pouvoir effectuer de prévisions. Un compromis est obtenu par les méthodes de *lissage exponentiel*. Celles-ci consistent à ajuster un modèle paramétrique de façon locale, c'est-à-dire dont les paramètres évoluent au cours du temps. Une prévision à très court terme est alors possible. Par exemple, dans le lissage exponentiel simple, la prévision à un pas est donnée par :

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

où la *constante de lissage*  $\gamma$  satisfait  $0 < \gamma < 1$  et  $Y_t = 0$  pour  $t \leq 0$  par convention. Le filtre est causal et infini. L'intérêt de la méthode réside dans la facilité de *mise à jour* de cette prévision lors de l'acquisition d'une nouvelle donnée  $Y_{T+1}$  :

$$\hat{Y}_{T+1}(1) = \hat{Y}_T(1) + (1 - \gamma) \left[ Y_{T+1} - \hat{Y}_T(1) \right].$$

La nouvelle prévision est égale à la précédente modifiée par la dernière erreur de prévision effectuée. Ceci est un exemple simple du célèbre *filtre de Kalman*. Le lissage exponentiel illustre deux des principaux objectifs des méthodes introduites dans ce chapitre : simplicité des calculs et facilité de mise à jour des résultats.

Le premier paragraphe présente la notion de *série CVS* dans le cadre simple des moyennes arithmétiques. L'étude générale des moyennes mobiles fait l'objet du paragraphe suivant. Le dernier paragraphe traite les méthodes de lissage exponentiel.

Les ouvrages de base pour ce chapitre sont ceux déjà présentés au chapitre précédent : [AND71], [GM90] et [MEL90].

## 3.1 MOYENNES ARITHMÉTIQUES

Les moyennes arithmétiques conduisent à une classe particulière de moyennes mobiles. Celles-ci sont cependant très utilisées car elles sont à la fois de conception simple, faciles à mettre en œuvre et suffisantes dans bien des situations.

### 3.1.1 Définitions et propriétés immédiates

Le lissage d'une chronique  $Y_t, t = 1, \dots, T$  par une *moyenne mobile d'ordre impair*  $m = 2k + 1$  est défini pour  $t = k + 1, \dots, T - k$ , par :

$$\tilde{Y}_t = M_m(Y_t) = \frac{1}{m} \{Y_{t-k} + \dots + Y_t + \dots + Y_{t+k}\} = \frac{1}{2k+1} \sum_{j=-k}^k Y_{t+j}.$$

Chaque valeur  $Y_t$  est donc remplacée par une simple moyenne arithmétique des valeurs qui l'entourent.

Afin de préserver la symétrie, la *moyenne mobile d'ordre pair*  $m = 2k$  est définie aux mêmes instants  $t = k + 1, \dots, T - k$  et porte sur les mêmes valeurs mais avec un poids 0,5 aux deux extrémités :

$$\tilde{Y}_t = M_m(Y_t) = \frac{1}{m} \left[ \frac{Y_{t-k}}{2} + \sum_{j=1-k}^{k-1} Y_{t+j} + \frac{Y_{t+k}}{2} \right] = \frac{1}{2k} \left[ \sum_{j=-k}^k Y_{t+j} - \frac{Y_{t-k} + Y_{t+k}}{2} \right].$$

Elle est égale à la moyenne d'ordre 2 de deux moyennes arithmétiques d'ordre  $m$  consécutives :

$$\tilde{Y}_t = \frac{\tilde{Y}_{t-0,5} + \tilde{Y}_{t+0,5}}{2} = \frac{1}{2} \left[ \frac{1}{m} \sum_{j=-k}^{k-1} Y_{t+j} + \frac{1}{m} \sum_{j=1-k}^k Y_{t+j} \right].$$

Il existe bien d'autres moyennes mobiles d'ordre  $m$  (*cf.* Paragraphe 3.2) mais cette terminologie est souvent employée pour désigner les moyennes définies ci-dessus.

La série lissée est plus courte que l'originale puisque  $[m/2]$  valeurs sont manquantes à chaque extrémité de la période d'observation. Un calcul récursif en temps de  $M_m(Y_t)$  est facile à mettre en œuvre. Le point  $(t, \tilde{Y}_t)$  est le centre de gravité des points  $(s, Y_s)$  qui ont servi à sa définition (avec un poids 0,5 aux extrémités dans le cas pair). En conséquence, à concavité constante, la série lissée reste d'un même côté par rapport à la série brute (sous-estimation ou surestimation de la tendance) et les points de retournement sont décalés dans les cas asymétriques (mauvaise localisation des changements de tendance) (*cf.* Figure 3.1). En effet l'opération de moyenne mobile est clairement linéaire et dans le cas du modèle additif avec composante saisonnière,  $Y_t = f_t + S_t + \varepsilon_t$ , on a :

$$\tilde{Y}_t = M_m(Y_t) = M_m(f_t) + M_m(S_t) + M_m(\varepsilon_t).$$

En particulier une tendance linéaire n'est pas modifiée par moyenne mobile.

### Illustration

On considère la chronique trimestrielle donnant l'évolution en pourcentage de la production industrielle française au cours des années 1981 à 1986 (source INSEE). Nous avons représenté (*cf.* Figure 3.2) la série brute ainsi que les séries lissées par les moyennes mobiles d'ordre 3, 4 et 5. Les données et les valeurs lissées sont dans le Tableau 3.1. Il est clair que la moyenne mobile lisse la série puisqu'elle atténue les oscillations. Dans cet exemple, où l'effet saisonnier (période 4) est très prononcé, les moyennes mobiles d'ordre 3 et 5 ne sont pas utilisables sur la série brute car elles perturbent complètement l'effet saisonnier (les pics et les creux sont déplacés). Par contre la moyenne mobile d'ordre 4 est particulièrement bien adaptée. La section qui suit justifie ce constat.

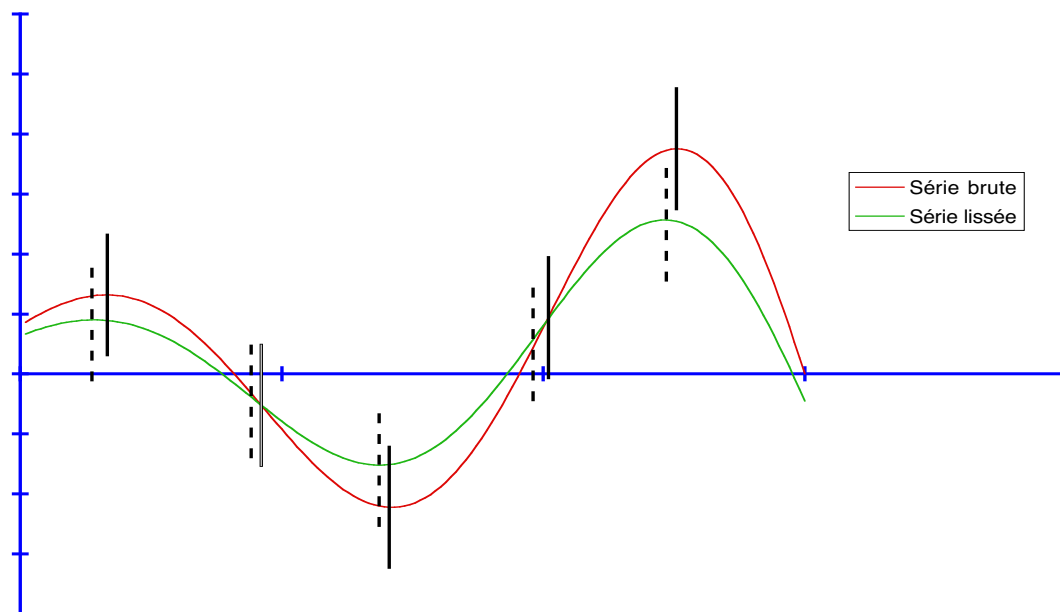


FIG. 3.1 – Le lissage par moyennes mobiles décale les changements de tendance (moyenne mobile d'ordre 51, séries représentées de taille 150)

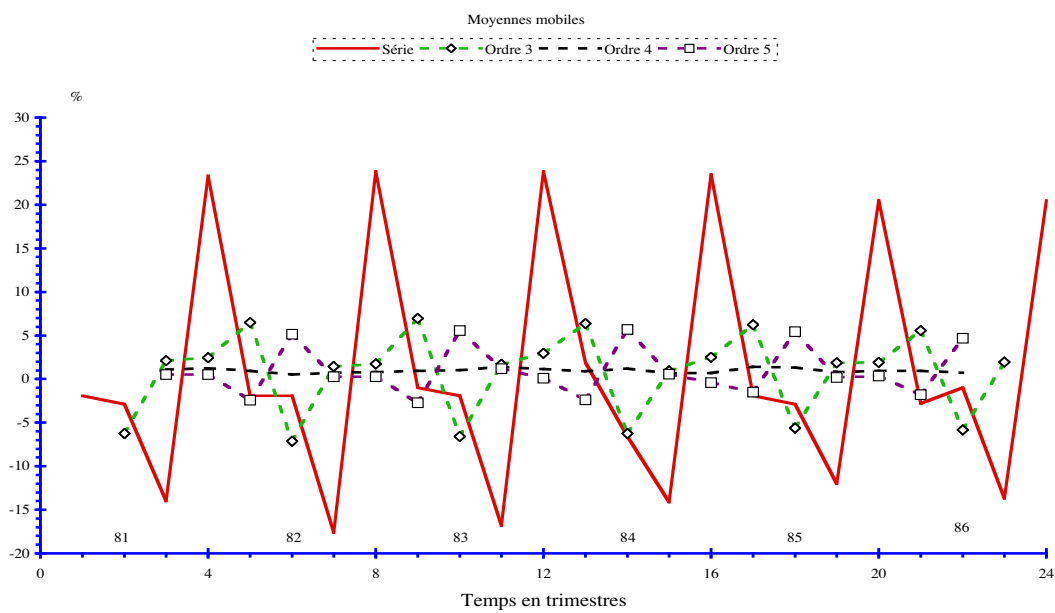


FIG. 3.2 – Évolution trimestrielle de la production industrielle française de 1981 à 1986 : moyennes mobiles d'ordre variable

année	trimestre		JFM	AMJ	JAS	OND
	<i>i</i>	<i>j</i>	1	2	3	4
1981	1	$Y_t$	-1,9	-2,9	-14,0	23,3
		$M3(Y_t)$		-6,3	2,1	2,5
		$M4(Y_t)$			1,1	1,3
		$M5(Y_t)$			0,5	0,5
		$Y_t - M4(Y_t)$				-15,1
		CVS	1,4	0,5	1,6	1,1
1982	2	$Y_t$	-1,9	-1,9	-17,6	23,8
		$M3(Y_t)$	6,5	-7,1	1,4	1,7
		$M4(Y_t)$	0,9	0,5	0,7	0,8
		$M5(Y_t)$	-2,4	5,1	0,3	0,3
		$Y_t - M4(Y_t)$	-2,8	-2,4	-18,3	23,0
		CVS	1,4	1,5	-2,0	1,6
1983	3	$Y_t$	-1,0	-1,9	-16,8	23,8
		$M3(Y_t)$	7,0	-6,6	1,7	3,0
		$M4(Y_t)$	0,9	1,0	1,4	1,2
		$M5(Y_t)$	-2,7	5,6	1,2	0,1
		$Y_t - M4(Y_t)$	-1,9	-2,9	-18,2	22,6
		CVS	2,3	1,5	-1,2	1,6
1984	4	$Y_t$	1,9	-6,6	-14,1	23,5
		$M3(Y_t)$	6,4	-6,3	0,9	2,5
		$M4(Y_t)$	0,9	1,2	0,7	0,7
		$M5(Y_t)$	-2,4	5,7	0,6	-0,4
		$Y_t - M4(Y_t)$	1,0	-7,8	-14,8	22,8
		CVS	5,2	-3,2	1,5	1,3
1985	5	$Y_t$	-1,9	-2,9	-12,0	20,5
		$M3(Y_t)$	6,2	-5,6	1,9	1,9
		$M4(Y_t)$	1,4	1,3	0,8	0,9
		$M5(Y_t)$	-1,5	5,4	0,2	0,4
		$Y_t - M4(Y_t)$	-3,3	-4,2	-12,8	19,6
		CVS	1,4	0,5	3,6	-1,7
1986	6	$Y_t$	-2,8	-1,0	-13,7	20,5
		$M3(Y_t)$	5,6	-5,8	1,9	
		$M4(Y_t)$	1,0	0,8		
		$M5(Y_t)$	-1,8	4,7		
		$Y_t - M4(Y_t)$	-3,8	-1,8		
		CVS	0,5	2,4	1,9	-1,7
		$S'_j$	-2,8	-2,9	-15,1	22,6
		$\hat{S}_j$	-3,3	-3,4	-15,6	22,2

TAB. 3.1 – Évolution en pourcentage de la production industrielle française de 1981 à 1986

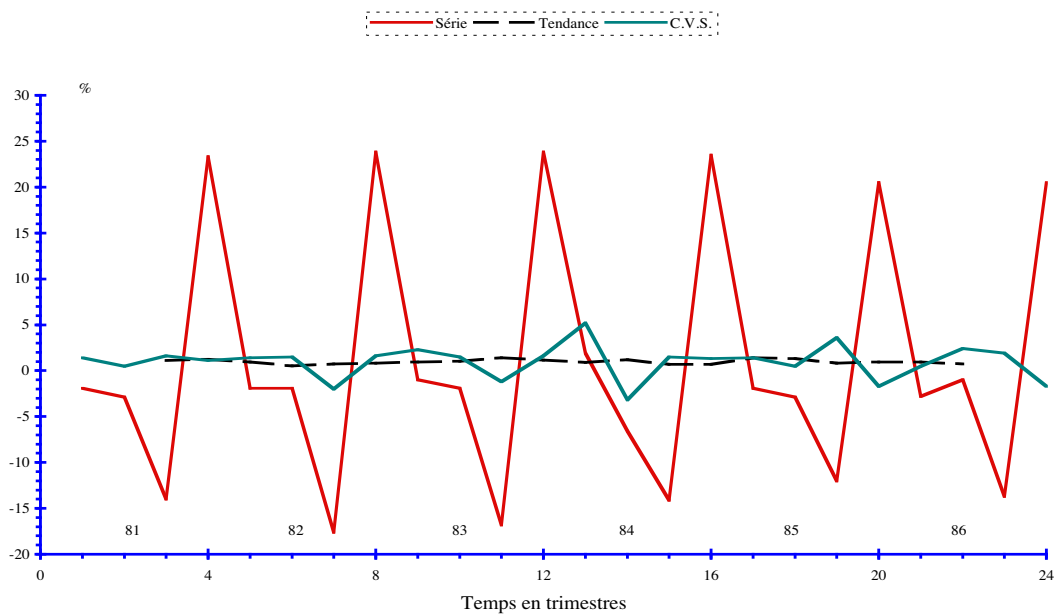


FIG. 3.3 – Évolution trimestrielle de la production industrielle française de 1981 à 1986 : série CVS

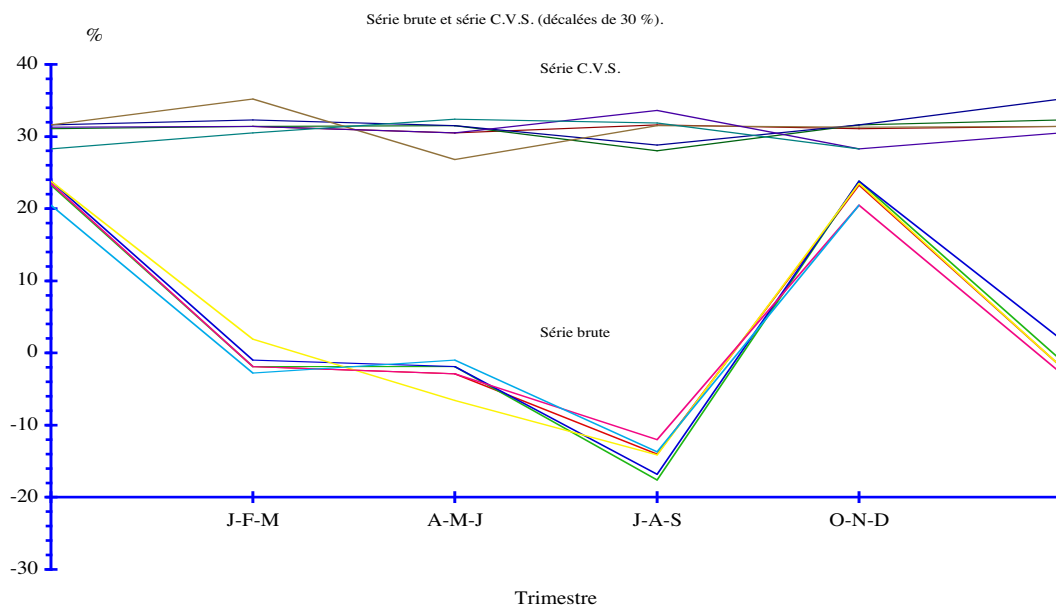


FIG. 3.4 – Évolution trimestrielle de la production industrielle française de 1981 à 1986 : mouvements saisonniers

### 3.1.2 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 :

$$Y_t = f_t + S_t + \varepsilon_t, \quad t = 1, \dots, T, \quad \varepsilon_t \sim b.b.(\sigma^2),$$

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  $p$  d'une série périodique de période  $p$  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, 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,  $\Delta_t = Y_t - M_p(Y_t), t = [p/2] + 1, \dots, T - [p/2]$ . Pour chaque "mois"  $j = 1, \dots, p$  fixé, les différences  $\Delta_{ij}$  observées sur les "années"  $i = 1$  ou  $2, \dots, n - 1$  ou  $n$  constituent  $n$  ou  $n - 1$  estimations du coefficient saisonnier  $S_j$ . Le nombre d'années étant en général faible, on retient les médianes  $S'_j$  de ces valeurs comme première estimation car la moyenne est trop sensible aux valeurs extrêmes. Lorsque la médiane porte sur un nombre pair de valeurs, on prend la demi-somme des valeurs centrales. L'estimation définitive est obtenue en centrant ces médianes en accord avec la contrainte  $\sum_{j=1}^p S_j = 0$  :

$$\hat{S}_j = S'(j) - \frac{1}{p} \sum_{k=1}^p S'_k, \quad j = 1, \dots, p.$$

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 :

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

**Illustration**

Les calculs concernant l'évolution de la production industrielle figurent dans le Tableau 3.1. La tendance est stable au voisinage de 1% et la série CVS conserve les irrégularités dues à la composante résiduelle (*cf.* Figure 3.3). La représentation des mouvements saisonniers (*cf.* Figure 3.4) montrent clairement l'élimination de l'effet saisonnier.

**Itération du procédé**

Il se peut que la série CVS obtenue ci-dessus présente encore un effet saisonnier. Dans ce cas on peut estimer à nouveau la tendance en lissant la série par une moyenne mobile d'ordre plus faible que  $p$  (2 ou 3 pour une série trimestrielle, 5 ou 7 pour une série annuelle). On reprend alors l'estimation des coefficients saisonniers comme précédemment à l'aide de la série brute  $Y_t$  et de la nouvelle tendance.

**3.2 MOYENNES MOBILES**

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

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

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

$$\hat{Y}_T(1) = (1 - \gamma) \sum_{k=0}^{\infty} \gamma^k Y_{T-k}.$$

**3.2.1 Définitions et propriétés immédiates**

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), qui à la série  $Y$  associe 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,

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

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

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 usuel 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 :

$$M(\alpha Y_t) = \alpha M(Y_t), \quad M(X_t + Y_t) = M(X_t) + M(Y_t),$$

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'ordre 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 symétriques 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*  $\{(B - 2I + F)^j; j \in \mathbb{N}\}$ . 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  $\Delta$ .

### 3.2.2 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 :

$$\alpha_0 x_t + \alpha_1 x_{t-1} + \dots + \alpha_k x_{t-k} = 0, \quad t \in \mathbb{Z}, \quad \alpha_0 \alpha_k \neq 0,$$

à laquelle on associe le *polynôme caractéristique*,

$$\alpha(z) = \alpha_0 z^k + \alpha_1 z^{k-1} + \dots + \alpha_{k-1} z + \alpha_k.$$

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

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

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

$$x_t = [(a_0 + a_1 t + \dots + a_{s-1} t^{s-1}) \cos \lambda t + (b_0 + b_1 t + \dots + b_{s-1} t^{s-1}) \sin \lambda t] \rho^t, \quad t \in \mathbb{Z},$$

est solution de l'équation de récurrence pour tout choix des 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,  $\alpha_j = \alpha_{k-j}, j = 0, \dots, k$ , et les solutions  $x_t$  ne présentent plus de formes explosives dues aux facteurs  $x^t$  ou  $\rho^t$ .

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 :

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

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

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

construites à partir des racines du polynôme caractéristique associé,

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

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 :

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

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

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

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

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

$$\gamma_{-l}Y_t + \gamma_{-l+1}Y_{t-1} + \dots + (\gamma_0 - 1)Y_{t-l} + \dots + \gamma_k Y_{t-l-k} = 0, \quad t \in \mathbb{Z}.$$

Les solutions forment un sous-espace vectoriel  $\mathcal{I}(M)$  de dimension  $m - 1$  obtenues à partir des solutions de l'équation  $g(z) - z^k = 0$ . Par composition on a l'inclusion  $\mathcal{I}(M) \cap \mathcal{I}(N) \subseteq \mathcal{I}(MN)$ . 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  $g(z) - z^k = 0$ .

### Moyennes de Spencer d'ordre 15 et 21

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

$$M = \frac{1}{320}(-3, -6, -5, 3, 21, 46, 67, \underline{74}) = \frac{1}{320}[4]^2[5](-3, 3, \underline{4}).$$

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,

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

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,  $\frac{1}{16}[4]^2$ , celles de période 5,  $\frac{1}{5}[5]$  et conserve les polynômes de degré 3 (effet conjugué des moyennes précédentes et de la moyenne  $\frac{1}{4}(-3, 3, \underline{4})$ ). 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 :

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

La *moyenne de Spencer d'ordre 21* est construite selon le même principe :

$$M = \frac{1}{350}(-1, -3, -5, -5, -2, 6, 18, 33, 47, 57, \underline{60}) = \frac{1}{350}[5]^2[7](-1, 0, 1, \underline{2}).$$

Elle élimine les saisonnalités de période 5 dont l'amplitude varie linéairement, celles de période 7 et conserve les polynômes de degré 3.

### Illustration

Nous reprenons l'exemple du chiffre d'affaires de la presse parisienne considéré au chapitre précédent. Nous déterminons la série CVS à l'aide de la moyenne arithmétique usuelle d'ordre 12,  $M_{12} = \frac{1}{24}(1, 2, 2, 2, 2, 2, \underline{2})$ , puis avec la moyenne mobile d'ordre 15,  $M_{15} = \frac{1}{24}(\frac{-73}{12}, 1, \frac{97}{12}, 2, 2, 2, 2, \underline{2})$ . Cette dernière est obtenue par composition de  $M_{12}$  avec  $M_3 = \frac{1}{12}(-73, \underline{158})$  de sorte à éliminer les composantes centrées périodiques de période 12 (effet de  $M_{12}$ ) et à conserver les polynômes de degré au plus 3 (effet combiné de  $M_{12}$  avec  $M_3$ ). Les résultats sont reportés dans le Tableau 3.2 et représentés sur les Figures 3.5 et 3.6. Les coefficients saisonniers obtenus par les deux méthodes sont assez proches et sont cohérents avec ceux fournis par le modèle quadratique. Ceux de  $M_{15}$  sont légèrement plus proches du cas quadratique que ceux de  $M_{12}$  au sens de l'écart quadratique moyen : 5,21 au lieu de 6,26. Par contre, on constate que la tendance estimée par  $M_{15}$  n'est pas du tout lisse. Il est donc préférable de retenir les résultats de  $M_{12}$ .

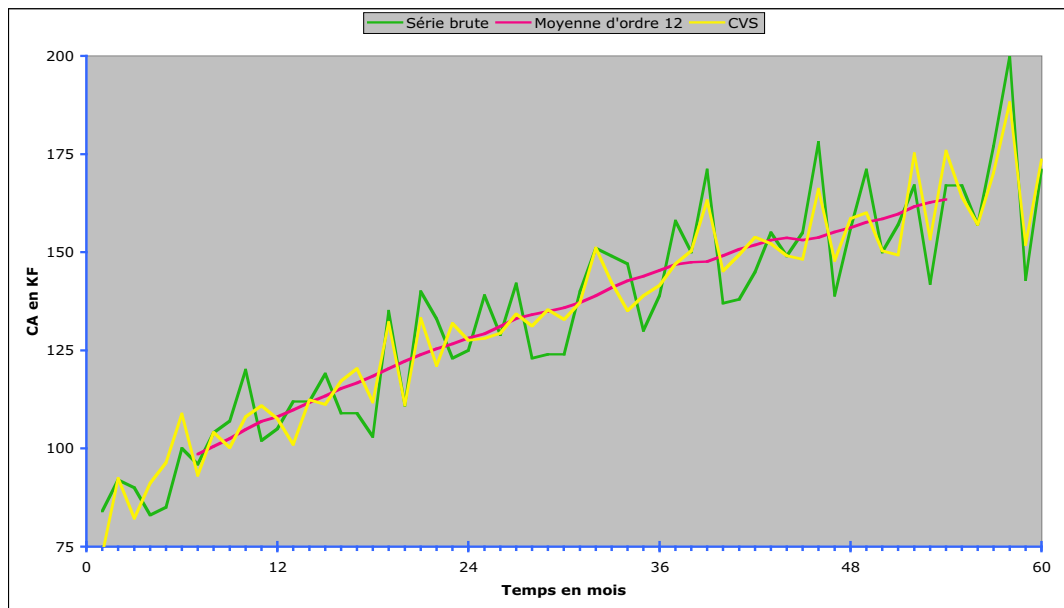


FIG. 3.5 – Chiffre d'affaires de la presse parisienne dans une petite ville de province, série CVS par moyenne arithmétique d'ordre 12

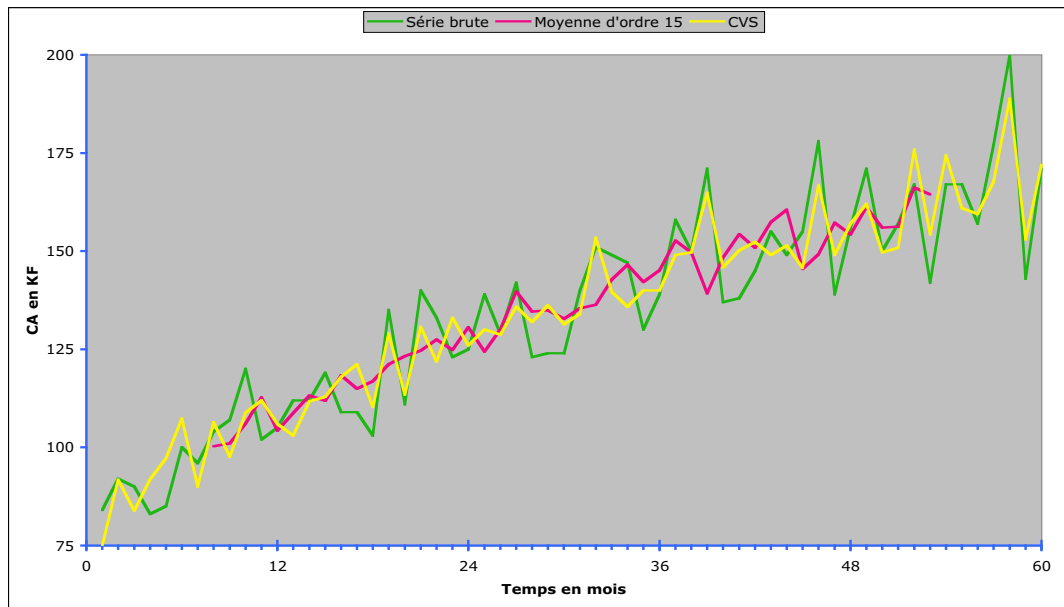


FIG. 3.6 – Chiffre d'affaires de la presse parisienne dans une petite ville de province, série CVS par moyenne mobile adaptée d'ordre 15

année	mois	janv	fév	mars	avr	mai	juin	juil	août	sept	oct	nov	déc
1981	chronique	84	92	90	83	85	100	96	104	107	120	102	105
	lissage $M_{12}$							99	101	103	105	107	108
	lissage $M_{15}$								100	101	106	113	104
	CVS $M_{12}$	73	92	82	91	96	109	93	104	100	108	111	108
	CVS $M_{15}$	75	92	84	92	97	107	90	106	98	109	112	106
1982	chronique	112	112	119	109	109	103	135	111	140	133	123	125
	lissage $M_{12}$	110	112	113	115	117	118	120	122	124	125	127	128
	lissage $M_{15}$	109	113	112	118	115	117	121	123	125	127	125	131
	CVS $M_{12}$	101	112	111	117	120	112	132	111	133	121	132	128
	CVS $M_{15}$	103	112	113	118	121	110	129	113	131	122	133	126
1983	chronique	139	129	142	123	124	124	140	151	149	147	130	139
	lissage $M_{12}$	129	131	133	134	135	136	137	139	141	143	144	145
	lissage $M_{15}$	124	130	140	135	135	133	135	136	143	147	142	145
	CVS $M_{12}$	128	129	134	131	135	133	137	151	142	135	139	142
	CVS $M_{15}$	130	129	136	132	136	131	134	153	140	136	140	140
1984	chronique	158	150	171	137	138	145	155	149	155	178	139	156
	lissage $M_{12}$	147	147	148	149	151	152	153	154	153	154	155	156
	lissage $M_{15}$	153	150	139	148	154	151	157	161	145	149	157	154
	CVS $M_{12}$	147	150	163	145	149	154	152	149	148	166	148	159
	CVS $M_{15}$	149	150	165	146	150	152	149	151	146	167	149	157
1985	chronique	171	150	157	167	142	167	167	157	177	200	143	171
	lissage $M_{12}$	158	159	160	162	163	163						
	lissage $M_{15}$	161	156	156	166	164							
	CVS $M_{12}$	160	150	149	175	153	176	164	157	170	188	152	174
	CVS $M_{15}$	162	150	151	176	154	174	161	159	168	189	153	172
coefficients saisonniers	$M_{12}$	11,0	-0,4	7,8	-8,2	-11,4	-8,8	2,9	-0,1	6,8	11,9	-8,9	-2,6
	$M_{15}$	9,0	0,3	6,2	-8,9	-12,2	-7,4	6,0	-2,4	9,3	11,2	-10,0	-1,0

TAB. 3.2 – Chiffre d'affaires mensuel de la presse parisienne dans une petite ville de province, lissages et séries CVS

### 3.2.3 Critères de sélection

Il est difficile de demander à une moyenne mobile de conserver certaines composantes tout en en supprimant d'autres. On retrouve la même difficulté qu'en modélisation paramétrique où la composante fondamentale n'est pas clairement dissociée de l'effet saisonnier. Il est donc nécessaire de privilégier l'élimination ou la conservation de l'une des composantes au détriment de l'opération contraire sur l'autre. Nous ne considérons désormais que des moyennes mobiles centrées et symétriques. Cela se justifie lorsque l'aspect descriptif et non évolutif est mis en avant, l'inversion du temps ne pouvant justifier une analyse différente. Cette précision est nécessaire dans la mesure où elle sera prise en compte dans la formulation du critère.

### Effet de Slutsky-Yule

Considérons l'effet d'une moyenne mobile sur la composante résiduelle  $\varepsilon_t$  constituée d'un bruit blanc de variance  $\sigma^2$  :

$$\tilde{\varepsilon}_t = \sum_{j=-k}^k \gamma_j \varepsilon_{t-j}, \quad \mathbb{E}(\tilde{\varepsilon}_t) = 0, \quad \tilde{\sigma}^2 = \text{Var}(\tilde{\varepsilon}_t) = \sigma^2 \sum_{j=-k}^k \gamma_j^2.$$

Sous la contrainte  $\sum_{j=-k}^k \gamma_j = 1$ , la variance  $\tilde{\sigma}^2$  est minimum dans le cas de la moyenne arithmétique  $\gamma_j = \frac{1}{2k+1}$ ,  $j = -k, \dots, k$ . Plus généralement la réduction du bruit est donnée par le rapport

$$\frac{\tilde{\sigma}^2}{\sigma^2} = \frac{\text{Var}(\tilde{\varepsilon}_t)}{\text{Var}(\varepsilon_t)} = \sum_{j=-k}^k \gamma_j^2.$$

Notons que le bruit transformé  $\tilde{\varepsilon}_t$  reste centré, sa variance est réduite mais ses composantes sont corrélées :

$$\text{Cov}(\tilde{\varepsilon}_t, \tilde{\varepsilon}_{t+h}) = \sigma^2 \sum_{j=h-k}^k \gamma_j \gamma_{j-h}, \quad \text{si } h \leq 2k, \quad 0 \text{ sinon.}$$

Cette corrélation entre les composantes de  $\tilde{\varepsilon}_t$  peut faire apparaître des oscillations artificielles sur la série lissée  $\tilde{Y}_t$  connues sous le nom de l'*effet de Slutsky-Yule*. On trouvera des compléments sur ce sujet dans [GM90].

### Méthode des moindres carrés

Considérons l'ajustement, par les moindres carrés, d'un polynôme de degré  $q$  aux observations dans une fenêtre  $[t-k, t+k]$ . On écrit le polynôme sous la forme  $a_0(t) + a_1(t)(s-t) + \dots + a_q(t)(s-t)^q$ . La solution  $\hat{a}_j(t)$ ,  $j = 0, \dots, q$  de

$$\min_{a_0(t), \dots, a_q(t)} \sum_{u=-k}^k \left[ Y_{t+u} - \sum_{j=0}^q a_j(t) u^j \right]^2$$

satisfait le système d'équations

$$\sum_{i=0}^q \sum_{u=-k}^k u^{i+j} \hat{a}_i(t) = \sum_{u=-k}^k u^j Y_{t+u}, \quad j = 0, \dots, q.$$

La valeur ajustée à l'instant  $t$  est celle du polynôme pour  $s = t$ , soit  $\hat{a}_0(t)$ . Le système ci-dessus montre qu'il s'agit d'une moyenne mobile,

$$\hat{a}_0(t) = \sum_{j=-k}^k \gamma_j Y_{t-j}.$$

Pour  $q = 1$ , on obtient la moyenne arithmétique  $\gamma_j = \frac{1}{2k+1}, j = -k, \dots, k$ . Pour  $q = 3$  et  $k = 2$ , on trouve la moyenne mobile  $M = \frac{1}{35}(-3, 12, \underline{17})$ . Il est clair que le degré de lissage augmente avec  $k$  et diminue lorsque  $q$  augmente. En général on utilise  $q = 3$  ou  $5$ . La moyenne mobile associée à une valeur paire  $q = 2r$  est identique à celle obtenue pour la valeur impaire supérieure  $2r + 1$  car  $\sum_{u=-k}^k u^j = 0$  lorsque  $j$  est impair. Les moyennes mobiles obtenues par cette méthode sont symétriques et conservent les polynômes de degré au plus  $q$ . En tenant compte de la symétrie, cette dernière condition s'écrit :

$$\gamma_0 + 2 \sum_{j=1}^k \gamma_j = 1, \quad \sum_{j=1}^k \gamma_j j^{2l} = 0, \quad l = 1, \dots, [q/2].$$

Pour  $q = 2k$  ou  $2k + 1$ , ce système conduit à la solution triviale  $\gamma_0 = 1$  et  $\gamma_j = 0, j = 1, \dots, k$ . En général  $k$  est plus grand et le critère des moindres carrés intervient. On pourrait le remplacer par d'autres contraintes (conservation de fonctions périodiques, minimisation du pouvoir de réduction du bruit,...).

### 3.3 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.

#### 3.3.1 Le lissage exponentiel simple

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

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

montrent que  $(1 - \gamma)$ , 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  $\beta$  ajustée à la série observée par la méthode des moindres carrés pondérés :

$$\min_b \sum_{k=0}^{\infty} \gamma^k (Y_{T-k} - b)^2.$$

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]. Des critères de choix optimal sont présentés et illustrés par des exemples dans [MEL90]. Les valeurs 0,7 ou 0,8 sont généralement recommandées.

### 3.3.2 Le 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  $h$ ,  $\hat{Y}_T(h) = \hat{\alpha}_T h + \hat{\beta}_T$ , 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 :

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

On note  $S_1(T) = \hat{Y}_T$  la *série lissée* introduite à la section 3.3.1 et  $S_2(T)$  la *série doublement lissée* définie par :

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

La solution du système est alors donnée par :

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

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

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

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

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

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

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

est égale à celle déjà obtenue à l'instant précédent  $T$  corrigée par un facteur proportionnel à la dernière erreur de prévision observée. Les valeurs initiales  $\hat{\beta}_2 = Y_2$  et  $\hat{\alpha}_2 = Y_2 - Y_1$  consistent à prévoir  $Y_3$  à l'aide de la droite définie par les points  $(1, Y_1)$  et  $(2, Y_2)$ .

### Illustration

On considère le taux d'inflation mensuel en France au cours des années 1970 à 1978 présenté dans le premier chapitre. Le Tableau 3.3 rappelle les données ainsi que les résultats des lissages exponentiels simple et double obtenus avec la constante de lissage égale à 0,8 (les calculs sont effectués avec les valeurs initiales non arrondies). La Figure 3.7 représente les chroniques correspondantes. On constate un effet de lissage, mais avec un retard sur l'évaluation des changements de tendance analogue à celui observé sur le taux annuel de la Figure 1.1. La méthode n'est pas vraiment adaptée à cette situation.

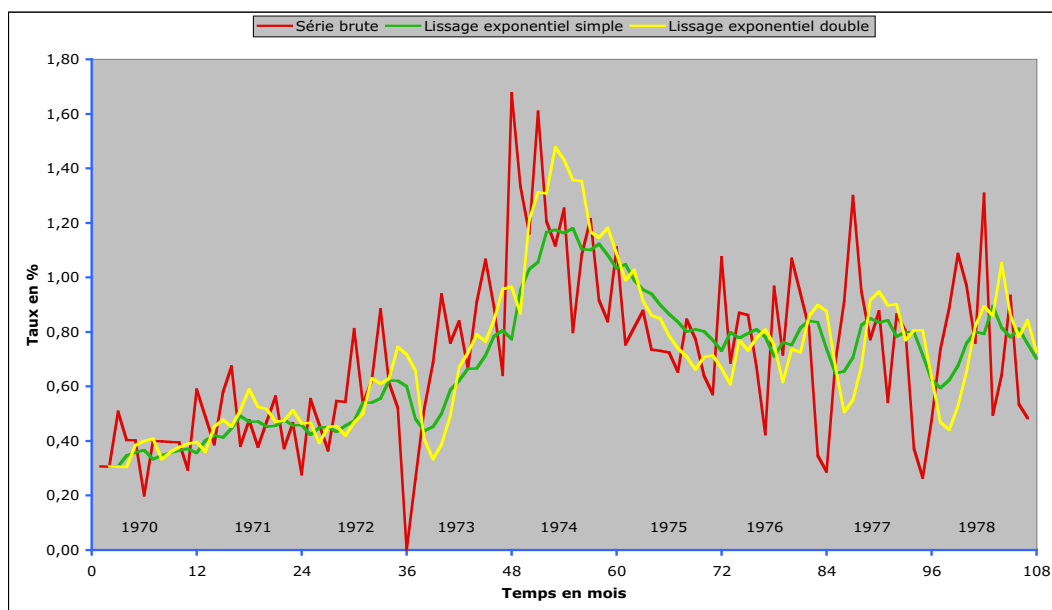


FIG. 3.7 – Taux d'inflation mensuel de 1970 à 1978 : lissage exponentiel,  $\gamma = 0,8$

année	mois		janv	fév	mars	avr	mai	juin	juil	août	sept	oct	nov	déc
	$i$	$j$	1	2	3	4	5	6	7	8	9	10	11	12
1970 simple double	1			0,31	0,31	0,51	0,40	0,40	0,20	0,40	0,40	0,40	0,40	0,30
				0,31	0,31	0,35	0,36	0,37	0,33	0,35	0,36	0,37	0,37	0,36
				0,31	0,31	0,30	0,38	0,40	0,41	0,33	0,36	0,38	0,39	0,40
1971 simple double	2		0,59	0,49	0,39	0,58	0,67	0,38	0,48	0,38	0,47	0,56	0,37	0,47
			0,40	0,42	0,41	0,45	0,49	0,47	0,47	0,45	0,46	0,48	0,46	0,46
			0,36	0,45	0,48	0,45	0,51	0,59	0,53	0,52	0,47	0,47	0,51	0,46
1972 simple double	3		0,28	0,55	0,46	0,37	0,55	0,54	0,81	0,54	0,62	0,88	0,61	0,52
			0,42	0,45	0,45	0,43	0,46	0,47	0,54	0,54	0,56	0,62	0,62	0,60
			0,47	0,39	0,45	0,45	0,42	0,47	0,50	0,63	0,61	0,63	0,75	0,72
1973 simple double	4		0,00	0,26	0,52	0,69	0,94	0,76	0,84	0,67	0,91	1,06	0,89	0,64
			0,48	0,44	0,45	0,50	0,59	0,62	0,67	0,67	0,71	0,78	0,81	0,77
			0,66	0,41	0,33	0,39	0,49	0,67	0,72	0,79	0,76	0,84	0,96	0,97
1974 simple double	5		1,68	1,33	1,16	1,61	1,21	1,12	1,25	0,80	1,08	1,21	0,92	0,84
			0,95	1,03	1,06	1,17	1,17	1,16	1,18	1,10	1,10	1,12	1,08	1,03
			0,87	1,21	1,31	1,31	1,48	1,43	1,36	1,35	1,17	1,15	1,18	1,09
1975 simple double	6		1,11	0,75	0,82	0,88	0,74	0,73	0,73	0,65	0,85	0,77	0,64	0,57
			1,05	0,99	0,95	0,94	0,90	0,86	0,84	0,80	0,81	0,80	0,77	0,73
			0,99	1,03	0,91	0,86	0,85	0,79	0,74	0,71	0,66	0,71	0,71	0,67
1976 simple double	7		1,07	0,69	0,87	0,86	0,67	0,42	0,97	0,72	1,07	0,94	0,81	0,35
			0,80	0,78	0,80	0,81	0,78	0,71	0,76	0,75	0,82	0,84	0,84	0,74
			0,61	0,77	0,73	0,78	0,81	0,75	0,62	0,74	0,72	0,86	0,90	0,88
1977 simple double	8		0,29	0,69	0,91	1,30	0,95	0,77	0,88	0,54	0,86	0,80	0,37	0,26
			0,65	0,66	0,71	0,83	0,85	0,83	0,84	0,78	0,80	0,80	0,71	0,62
			0,67	0,51	0,55	0,67	0,92	0,95	0,90	0,90	0,77	0,81	0,81	0,63
1978 simple double	9		0,48	0,74	0,89	1,09	0,97	0,76	1,31	0,50	0,64	0,93	0,53	0,48
			0,59	0,62	0,68	0,76	0,80	0,79	0,90	0,82	0,78	0,81	0,76	0,70
			0,47	0,44	0,53	0,65	0,82	0,89	0,86	1,05	0,86	0,78	0,84	0,73

TAB. 3.3 – Taux mensuel des prix à la consommation de 1970 à 1978 : lissage exponentiel,  $\gamma = 0,8$

### 3.3.3 Le 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, changent 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*,

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

où  $Y(T) = {}^t[Y_T, \dots, Y_1]$  est le vecteur des observations ordonnées dans le sens inverse du sens habituel,  $\theta(T) = {}^t[\theta_1(T), \dots, \theta_r(T)]$  est le vecteur des

paramètres et  $\Gamma(T)$  est une matrice diagonale avec

$$\text{Diag}\{\Gamma(T)\} = (1, \gamma^{-1}, \dots, \gamma^{1-T}).$$

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

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

puisque ses lignes sont définies par une fonction vectorielle  $x(t), t \in \mathbb{Z}$ , satisfaisant la récurrence  $x(t) = Ax(t-1)$  où  $A$  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  $x(t-T)$  :

$$Y_t = {}^t\theta(T)x(t-T) + \varepsilon_t(T), \quad t = 1, \dots, T.$$

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

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

Ainsi l'espace des moyennes  $\mathcal{M}\{X(T) {}^t A\}$  du sous modèle de dimension  $T$  obtenu à l'étape  $T+1$  est identique au précédent  $\mathcal{M}\{X(T)\}$ . Par contre la matrice de variance des erreurs associées est divisée par  $\gamma$  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  $\hat{\theta}(T+1) = {}^t A \hat{\theta}(T)$ . La solution est en fait :

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

où

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

Pour cela considérons le problème des *moindres carrés pondérés*,

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

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

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

A 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  $\hat{Y}_T(h) = {}^t \hat{\theta}(T) x(h)$ . L'initialisation qui conduit aux solutions des moindres carrés ci-dessus est

$$\hat{\theta}(r) = M^{-1} {}^t X(\infty) \Gamma(\infty)^{-1} Y(r).$$

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  $t \leq 0$ ,

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

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

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

Pour un polynôme de degré  $q$  on utilise

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

et la matrice  $A$  est triangulaire inférieure avec  $A_{ij} = 1$  si  $i = j$  ou  $i = j + 1$ , 0 sinon. Les fonctions sinusoidales  $\theta_1 \sin \lambda t + \theta_2 \cos \lambda t$  sont associées à la rotation d'angle  $-\lambda$  et la fonction exponentielle  $e^{\alpha t}$  à  $e^\alpha$  avec  $\alpha > 0$  pour l'existence de  $M$ .

### 3.3.4 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.

**Modèle de Holt**

Posant  $\alpha = 1 - \gamma^2$  et  $\beta = \frac{1-\gamma}{1+\gamma}$ , la mise à jour des coefficients du lissage exponentiel double s'écrit :

$$\begin{aligned}\hat{\beta}_{T+1} &= \alpha Y_{T+1} + (1 - \alpha)(\hat{\beta}_T + \hat{\alpha}_T) = \alpha Y_{T+1} + (1 - \alpha)\hat{Y}_T(1), \\ \hat{\alpha}_{T+1} &= (1 - \beta)\hat{\alpha}_T + \beta(\hat{\beta}_{T+1} - \hat{\beta}_T).\end{aligned}$$

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 :

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

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

**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 :

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

où  $\alpha, \beta$  et  $\delta$  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), 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  $(t, Y_t), t = 1, \dots, p$  :

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

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

**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 :

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

avec les formules de mise à jour suivantes :

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

Pour les valeurs initiales des paramètres, on utilise  $\hat{S}_j = \frac{Y_j}{\hat{\alpha}_p(j-p) - \hat{\beta}_p}$ ,  $j = 1, \dots, p$  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.



# Liste des tableaux

1.1	Indice mensuel des prix à la consommation, base 100 en juillet 1970, de 1970 à 1978 . . . . .	14
1.2	Taux mensuel des prix à la consommation de 1970 à 1978 . . . . .	14
2.1	Chiffre d'affaires mensuel de la presse parisienne dans une petite ville de province (unité : 1KF) . . . . .	28
2.2	Chiffre d'affaires mensuel de la presse parisienne dans une petite ville de province : les différentes formes de résidus . . . . .	34
2.3	Coefficients saisonniers du chiffre d'affaires mensuel de la presse parisienne dans une petite ville de province . . . . .	42
2.4	Test de l'effet saisonnier du chiffre d'affaires mensuel de la presse parisienne dans une petite ville de province . . . . .	43
2.5	Chiffre d'affaires mensuel de la presse parisienne dans une petite ville de province, modèle quadratique (unité : 1KF) . . . . .	43
2.6	Coefficients saisonniers du chiffre d'affaires mensuel de la presse parisienne dans une petite ville de province, modèle quadratique . . . . .	43
2.7	Tests de l'effet saisonnier et de la présence du terme quadratique pour le chiffre d'affaires mensuel de la presse parisienne dans une petite ville de province, modèle quadratique . . . . .	43
3.1	Évolution en pourcentage de la production industrielle française de 1981 à 1986 . . . . .	50
3.2	Chiffre d'affaires mensuel de la presse parisienne dans une petite ville de province, lissages et séries CVS . . . . .	59
3.3	Taux mensuel des prix à la consommation de 1970 à 1978 : lissage exponentiel, $\gamma = 0,8$ . . . . .	65



# Table des figures

1	Nombre annuel de taches solaires selon Wolf de 1700 à 1924 . . . . .	7
2	Nombre mensuel de taches solaires selon Wolf de 1900 à 1916 . . . . .	7
3	Indice annuel du prix du blé en Europe selon Beveridge de 1500 à 1869, base 100 : moyenne des années 1700 à 1745 . . . . .	8
4	Nombre mensuel de passagers internationaux aux États Unis de 1949 à 1960 . . . . .	8
5	Bruit blanc gaussien de variance 1 . . . . .	9
6	Marche aléatoire, valeurs cumulées du bruit blanc gaussien . . . . .	9
1.1	Indice mensuel des prix à la consommation, base 100 en juillet 1970, de 1970 à 1978 . . . . .	15
1.2	Taux d'inflation mensuel de 1970 à 1978 . . . . .	15
1.3	Taux d'inflation mensuel de 1975 à 1978 . . . . .	18
1.4	Mouvement saisonnier du taux d'inflation mensuel de 1975 à 1978 . . . . .	18
1.5	Représentation polaire du mouvement saisonnier du taux d'inflation mensuel de 1975 à 1978 (échelle polaire : pôle à 0%, graduation de 0,1%) . . . . .	19
2.1	Chiffre d'affaires mensuel de la presse parisienne dans une petite ville de province . . . . .	28
2.2	Chiffre d'affaires mensuel de la presse parisienne dans une petite ville de province : représentation des résidus obtenus par validation croisée . . . . .	33
2.3	Mouvement saisonnier du chiffre d'affaires mensuel de la presse parisienne dans une petite ville de province . . . . .	34
2.4	Chiffre d'affaires de la presse parisienne dans une petite ville de province, modèle quadratique . . . . .	41
2.5	Résidus normalisés du chiffre d'affaires de la presse parisienne dans une petite ville de province, modèle quadratique . . . . .	42

3.1	Le lissage par moyennes mobiles décale les changements de tendance (moyenne mobile d'ordre 51, séries représentées de taille 150) . . . . .	49
3.2	Évolution trimestrielle de la production industrielle française de 1981 à 1986 : moyennes mobiles d'ordre variable . . . . .	49
3.3	Évolution trimestrielle de la production industrielle française de 1981 à 1986 : série CVS . . . . .	51
3.4	Évolution trimestrielle de la production industrielle française de 1981 à 1986 : mouvements saisonniers . . . . .	51
3.5	Chiffre d'affaires de la presse parisienne dans une petite ville de province, série CVS par moyenne arithmétique d'ordre 12 .	58
3.6	Chiffre d'affaires de la presse parisienne dans une petite ville de province, série CVS par moyenne mobile adaptée d'ordre 15	58
3.7	Taux d'inflation mensuel de 1970 à 1978 : lissage exponentiel, $\gamma = 0,8$ . . . . .	64

# Bibliographie

- [ABC92] A. ANTONIADIS, J. BERRUYER, and R. CARMONA. *Régression non linéaire et applications*. Economica, Paris, 1992.
- [AND71] T.W. ANDERSON. *The statistical Analysis of Time series*. Wiley, New York, 1971.
- [BD91] P.J. BROCKWELL and R.A. DAVIS. *Time Series : Theory and Methods*. Springer-Verlag, New-York, 1991.
- [BJR94] G.E.P. BOX, G.M. JENKINS, and G.C. REINSEL. *Time Series Analysis, Forecasting and Control*. Prentice Hall, Englewood Cliffs, New Jersey, third edition, 1994.
- [BRI81] D.R. BRILINGER. *Time Series, Data Analysis and Theory*. Rinehart & Winston, New-York, 1981.
- [CAL65] G. CALOT. *Cours de statistique descriptive*. Dunod, Paris, 3<sup>e</sup> édition, 1965. Collection Statistique et Programmes Économiques.
- [DFT89] J.-J. DROESBECKE, B. FICHET, and P. TASSI. *Séries chronologiques : théorie et pratique des modèles ARIMA*. Economica, Paris, 1989.
- [DFT94] J.-J. DROESBECKE, B. FICHET, and P. TASSI. *Modélisation ARCH. Théorie statistique et applications dans le domaine de la finance*. Editions de l'Université de Bruxelles, Bruxelles et Editions Ellipses, Paris, 1994.
- [GM89] C. GOURIEROUX and A. MONFORT. *Statistique et modèles économétriques*, volume 1 & 2. Economica, Paris, 1989.
- [GM90] C. GOURIEROUX and A. MONFORT. *Séries temporelles et modèles dynamiques*. Economica, Paris, 1990.
- [GUE94] D. GUEGAN. *Séries chronologiques non linéaires à temps discret*. Economica, Paris, 1994.
- [HAN70] E.J. HANNAN. *Multiple Time Series*. Wiley, New-York, 1970.

- [IZE83] A.J. IZENMAN. J.r. wolf and h.a. wolfer : an historical note on zurich sunspot relative numbers. *J. R. Stat. Soc.*, A146 :311–318, 1983.
- [KAY87] S.M. KAY. *Modern Spectral Estimation. Theory & Application*. Prentice-Hall, Englewood Cliffs, 1987.
- [MAR87] Jr. MARPLE, S.L. *Digital Spectral Analysis with Applications*. Prentice-Hall, Englewood Cliffs, 1987.
- [MEL90] G. MELARD. *Méthodes de prévision à court terme*. Editions de l'Université de Bruxelles, Bruxelles et Editions Ellipses, Paris, 1990.
- [OWY83] A.V. OPPENHEIM, A.S. WILLSKY, and I.T. YOUNG. *Signals and Systems*. Prentice-Hall International, Englewood Cliffs, 1983.
- [PRI81] M.B. PRIESTLEY. *Spectral analysis and time series*, volume 1 & 2. Academic-Press, New-York, 1981.
- [TAS89] P. TASSI. *Méthodes statistiques*. Economica, Paris, 2<sup>e</sup> edition, 1989.

# Index

- équation caractéristique, 55
- équation de récurrence linéaire, 55
  
- bruit blanc, 20
- Buy-Ballot, 25
  
- causal, 46
- centrée, 54
- chronique, 11
- coefficients saisonniers, 20, 45, 52
- coefficients studentisés, 33
- composante cyclique, 20
- composante fondamentale, 20
- composante résiduelle, 20
- composante saisonnière, 20, 45
- composantes, 17
- constante de lissage, 46, 61
- Corrigée des Variations Saisonnières, 45
  
- désaisonnalisée, 45
  
- effet de Slutsky-Yule, 60
- erreur, 23
- espace des moyennes, 38
- espace des observations, 38
  
- facteur d'oubli, 61, 62
- filtrage, 46, 62
- filtre, 53
- filtre de Kalman, 47
- flux, 12
  
- Gauss-Markov, 38
  
- horizon, 61
  
- invariante, 56
- invariants, 55
  
- lissage exponentiel, 46
- lissage exponentiel double, 63
- lissage exponentiel généralisé, 65
- lissage exponentiel simple, 61
  
- matrice chapeau, 38
- matrice de prédiction, 38
- mise à jour, 46, 61
- modèle de Holt, 68
- modèle de Holt-Winter additif avec saisonnalité, 68
- modèle de Holt-Winter multiplicatif avec saisonnalité, 69
- modèle linéaire généralisé, 65
- moindres carrés, 38
- moindres carrés généralisés, 40
- moindres carrés pondérés, 40, 66
- mouvement conjoncturel, 46
- mouvement saisonnier, 20
- moyenne, 23
- moyenne de Spencer, 56
- moyenne de Spencer d'ordre 21, 57
- moyenne mobile d'ordre  $m$ , 54
- moyenne mobile d'ordre impair, 47
- moyenne mobile d'ordre pair, 47
- moyennes arithmétiques, 46
  
- niveau, 12
- noyau, 55
  
- opérateur de différence arrière, 54

- opérateur de différence symétrique, 54
- opérateur de retard, 53
- période, 12, 20
- plan d'expérience, 37
- polynôme caractéristique, 55
- prévision, 27
  
- régression linéaire, 23
- résidu, 31
- résidu studentisé, 32, 39
- résidus standardisés, 32, 39
  
- série ajustée, 27
- série chronologique, 11
- série Corrigée des Variations Saisonnières, 52
- série CVS, 20, 45, 47
- série doublement lissée, 63
- série lissée, 63
- série temporelle, 11
- séries désaisonnalisées, 52
- stock, 12
- symétrique, 54
  
- Table de Buys-Ballot, 27
- tendance, 20, 45
  
- valeur lissée, 62
- validation croisée, 32, 39
- variations accidentelles, 20
- vecteur d'état à transition fixe, 66
- vecteur des résidus, 38

# Table des matières

INTRODUCTION	3
<b>1 GÉNÉRALITÉS</b>	<b>11</b>
1.1 LE TEMPS . . . . .	11
1.1.1 Définition d'une série chronologique . . . . .	11
1.1.2 Quelques précautions élémentaires . . . . .	12
1.1.3 Exemple . . . . .	13
1.2 REPRÉSENTATIONS GRAPHIQUES . . . . .	16
1.2.1 Représentation de la chronique . . . . .	16
1.2.2 Représentation du mouvement saisonnier . . . . .	17
1.2.3 Exemple . . . . .	17
1.3 LES MODÈLES POUR LA MOYENNE . . . . .	17
1.3.1 Les composantes du modèle . . . . .	17
1.3.2 Les schémas de composition . . . . .	21
<b>2 RÉGRESSION LINÉAIRE</b>	<b>23</b>
2.1 MODÈLE DE BUYS-BALLOT . . . . .	25
2.1.1 Table de Buys-Ballot . . . . .	25
2.1.2 Moyenne et variance des estimateurs . . . . .	29
2.1.3 Inférence statistique . . . . .	31
2.2 MODÈLE LINÉAIRE GÉNÉRAL . . . . .	36
2.2.1 Présentation du modèle . . . . .	36
2.2.2 Rappels sur le modèle linéaire . . . . .	37
<b>3 LES MÉTHODES DE LISSAGE</b>	<b>45</b>
3.1 MOYENNES ARITHMÉTIQUES . . . . .	47
3.1.1 Définitions et propriétés immédiates . . . . .	47
3.1.2 Série corrigée des variations saisonnières (Série CVS) . . . . .	52
3.2 MOYENNES MOBILES . . . . .	53
3.2.1 Définitions et propriétés immédiates . . . . .	53
3.2.2 Noyau et invariants . . . . .	55

3.2.3	Critères de sélection . . . . .	59
3.3	LISSAGE EXPONENTIEL . . . . .	61
3.3.1	Le lissage exponentiel simple . . . . .	61
3.3.2	Le lissage exponentiel double . . . . .	63
3.3.3	Le lissage exponentiel généralisé . . . . .	65
3.3.4	L'approche de Holt et Winters . . . . .	67
<b>LISTE DES TABLEAUX</b>		<b>69</b>
<b>LISTE DES FIGURES</b>		<b>71</b>
<b>BIBLIOGRAPHIE</b>		<b>74</b>
<b>INDEX</b>		<b>76</b>
<b>TABLE DES MATIÈRES</b>		<b>79</b>