

Chapitre 8

ANALYSE DES SÉRIES CHRONOLOGIQUES

Nous abordons dans ce chapitre l'analyse de données statistiques particulières en ce sens que les observations sont régulièrement échelonnées dans le temps. Ce genre de données est bien connu en économie : la quasi-totalité des indices de prix, de production etc. sont calculés régulièrement par l'INSEE ou d'autres établissements et constituent ce que l'on appelle des séries chronologiques. Elles sont fréquentes aussi en gestion : surveillance du niveau des stocks, suivi des ratios d'une entreprise etc.... Leur particularité vient de l'introduction du temps dans l'analyse de ces données : on étudie une suite de couples de la forme (t, x_t) , où x_t est l'observation de la variable à l'instant t .

1. DESCRIPTION D'UNE SÉRIE CHRONOLOGIQUE.

On distingue en général trois effets constitutifs d'une série chronologique :

- Un effet à long terme, appelée tendance (on ajoute parfois à long terme), composante tendancielle ou trend ;

- Un effet dit saisonnier, qui réapparaît à intervalles réguliers ; cet effet se traduit par une composante de la série appelée composante saisonnière.
- Un effet inexplicé : cet effet, que l'on suppose en général dû au hasard, se manifeste par des variations accidentelles.

Dans les séries économiques longues, on cite souvent un effet supplémentaire : c'est ce que l'on appelle le cycle de Kondratiev, qui résulte du fait que, suivant la théorie de Kondratiev, à une période de prospérité économique succède mécaniquement une période de dépression.

1.1 Description de la tendance.

La description initiale de la tendance repose sur l'interprétation de la représentation graphique de la série.

Définition : on appelle tendance (ou variation à long terme ou trend) de la série x_t la série c_t résultant de la totalité des effets permanents auxquels est soumise la série x_t .

Exemple : Nous donnons ci-dessous les cours (en €) du titre Alcatel de code sicovam 13000 du 4 janvier (n°1) au 5 mars 1999 (n°45). L'unité de temps est le jour boursier, et le cours est déterminé par l'offre et la demande elles-mêmes déterminées par l'évolution économique. Les données figurent sur le site (paramètres Alcatel.par) :

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

Tableau 1.8 : Cours du titre Alcatel du 4 janvier 1999 au 5 mars 1999

La tendance peut être décomposée en trois phases (figure 1.8) :

- Le cours baisse du début des observations (7 janvier) jusqu'à l'observation n°15 (c'est-à-dire le 22 janvier) ;

- De l'observation n°16 (25 janvier) à l'observation n°41 (1^{er} mars), le cours diminue légèrement ;

- Il augmente rapidement à partir du 2 mars.

On ne distingue pas sur la figure 1.8 de variations apparaissant à intervalles réguliers : la série n'est soumise à aucun effet saisonnier visible. La composante accidentelle est visualisée par les petites variations de cours d'un jour à l'autre. Par exemple le cours n°11 est supérieur au cours n°10 : cela ne remet pas en cause la baisse de la tendance compte tenu des cours du 7 au 22 janvier.

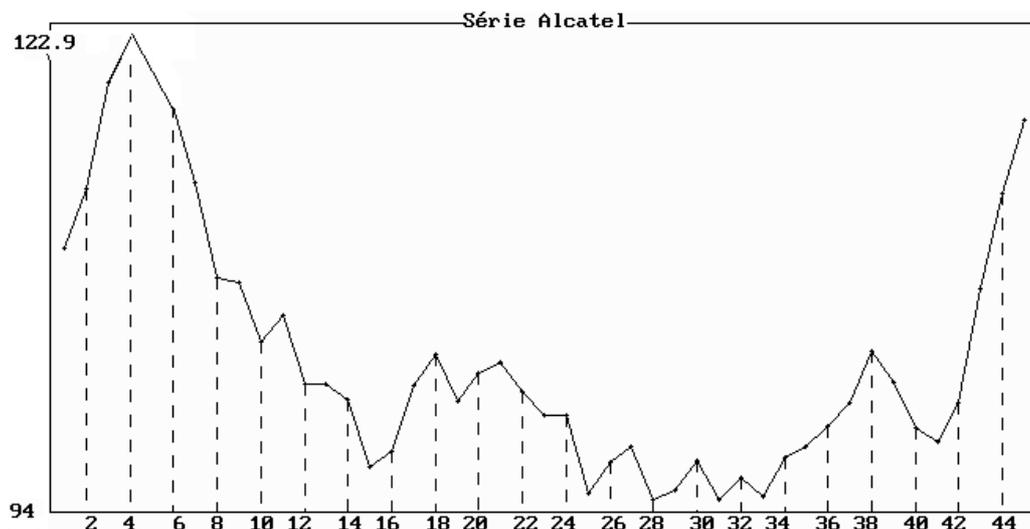


Figure 1.8 : cours journalier du titre Alcatel
(du 4 janvier 1999 au 5 mars 1999)

Pour faire apparaître plus clairement la tendance, il faut atténuer la composante accidentelle. On utilise pour cela les moyennes mobiles définies de la façon suivante :

Définitions :

- on appelle moyenne mobile centrée de longueur impaire $l_i = 2k + 1$ à l'instant t la valeur moyenne mm_t des observations $x_{t-k}, x_{t-k+1}, \dots, x_t, x_{t+1}, \dots, x_{t+k}$:

$$mm_t = (x_{t-k} + \dots + x_{t-1} + x_t + x_{t+1} + \dots + x_{t+k}) / l_i$$

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

$$mm_t = (0.5 x_{t-k} + x_{t-k+1} + \dots + x_{t-1} + x_t + x_{t+1} + \dots + x_{t+k-1} + 0.5 x_{t+k}) / l_p$$

Dans la première formule, le nombre de termes de la somme est égal à $2k + 1$: il s'agit bien d'une moyenne. Dans la seconde, la somme des coefficients est égale à $2k$, puisque le premier et le dernier sont égaux à 0.5 : il s'agit d'une moyenne pondérée. Dans les deux cas, le nombre d'observations prise en compte avant l'instant t est égal au nombre d'observations prises en compte après l'instant t : c'est pour cela que les moyennes sont dites centrées.

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

$$mm_3 = (0.5 x_1 + x_2 + x_3 + x_4 + 0.5 x_5) / 4 \quad (l = 4)$$

$$mm_3 = (x_1 + x_2 + x_3 + x_4 + x_5) / 5 \quad (l = 5)$$

De façon générale, ne peut calculer de moyenne mobile en $t = 1, t = 2, \dots, t = k$ puisque les formules ne peuvent être appliquées que si l'on connaît x_{t-k} . De même, si T est le nombre total d'observations, on ne peut calculer mm_T, \dots, mm_{T-k+1} puisqu'il faut connaître x_{t+k} .

L'avantage des moyennes mobiles est d'atténuer la composante accidentelle tout en conservant les tendances linéaires : la série est dite « lissée », et est d'autant plus lissée que la longueur de la moyenne mobile est élevée comme on peut le constater sur la figure 3.8 sur laquelle nous avons représenté les moyennes mobiles de longueur 14.

Exemple : nous donnons dans le tableau 2.8 un extrait des moyennes mobiles de longueur 5 et les représentations graphiques sur les figures 2.8 et 3.8 des cours du titre Alcatel et des moyennes mobiles de longueur 5 et 14.

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

Tableau 2.8 : cours du titre Alcatel du 4 janvier (n°1) au 5 mars 1999 (n°45) moyennes mobiles de longueur 5 (extrait)

On notera que, dans les journaux financiers, les moyennes mobiles ne sont pas centrées : on utilise les moyennes des 50 ou 100 dernières observations avant l'instant t pour définir la tendance à l'instant t .

L'inconvénient des moyennes mobiles de longueur 14 est qu'elles ne sont définies qu'à partir de la 8^e observation et jusqu'à la 38^e. On ne dispose d'aucune information sur la tendance ni au début ni à la fin de la période d'observation. Il faut donc choisir la longueur des

moyennes mobiles suivant le nombre d'observations et l'objectif de l'analyse. Nous verrons que la longueur de ces m.m. dépend aussi de la période des variations saisonnières pour faire apparaître la tendance.

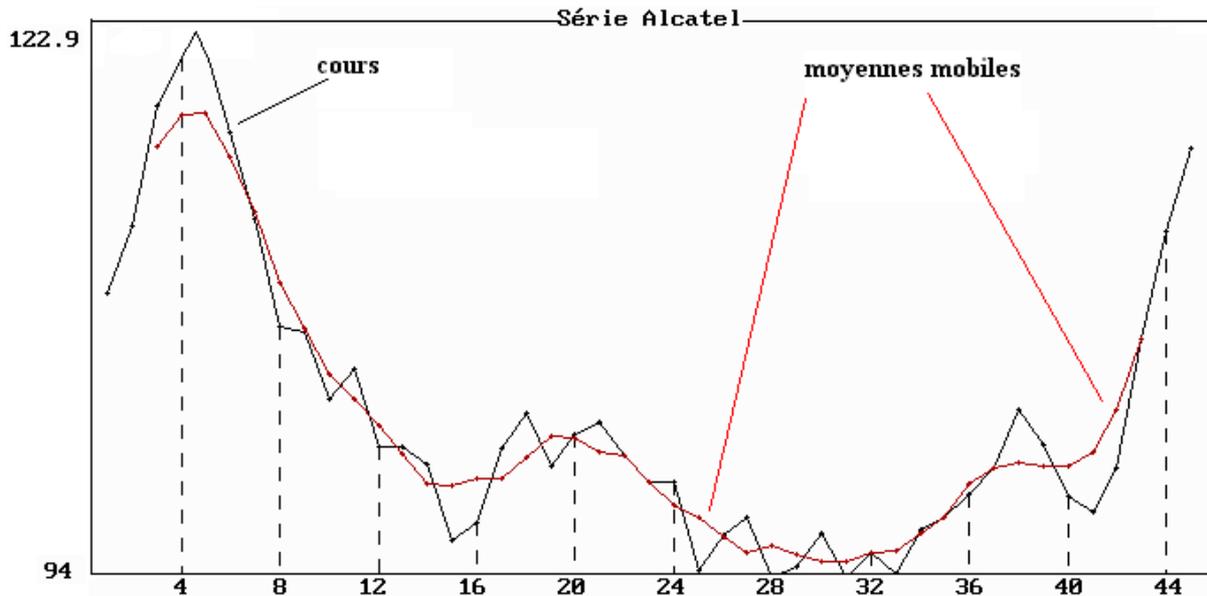


Figure 2.8 : cours du titre Alcatel du 4 janvier au 5 mars 1999
Moyennes mobiles de longueur 5.

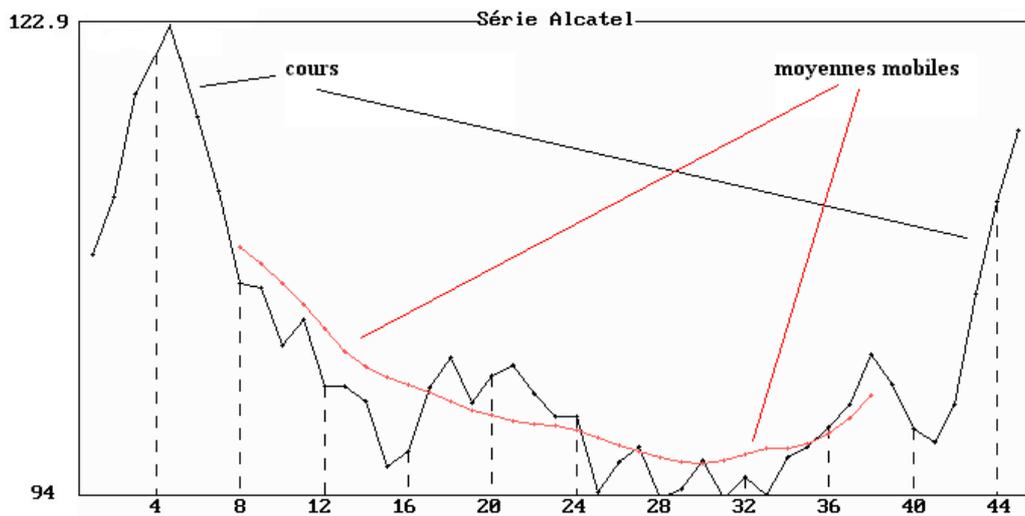


Figure 3.8 : cours du titre Alcatel du 4 janvier au 5 mars 1999
Moyennes mobiles de longueur 4.

Théorème : si la tendance d'une série chronologique x_t est linéaire et a pour équation $c_t = \beta t + \alpha$, les moyennes mobiles centrées ont pour tendance la même droite et en sont d'autant plus proches que la longueur des moyennes mobiles est élevée.

Ce théorème est démontré dans un complément pédagogique. Il peut être complété par l'étude de tendance de la forme $c_t = \beta_2 t^2 + \beta_1 t + \alpha$ proposée en application pédagogique.

1.2 Description simultanée des variations saisonnières et de la tendance.

Une variation saisonnière est caractérisée par le fait qu'elle se produit à intervalles de temps réguliers, d'où d'ailleurs le terme saisonnier.

Définition : on appelle variation saisonnière d'une série chronologique à l'instant t une variation due à un effet momentané se reproduisant régulièrement dans le temps.

Exemple : nous étudions la série chronologique suivante observée trimestriellement pendant 6 ans (tableau 3.8).

	<i>1^{er} trimestre</i>	<i>2^e trimestre</i>	<i>3^e trimestre</i>	<i>4^e trimestre</i>
<i>Année 1</i>	89.658	97.593	108.906	114.157
<i>Année 2</i>	96.205	99.399	112.763	119.185
<i>Année 3</i>	99.602	105.192	116.556	121.911
<i>Année 4</i>	103.272	109.644	121.208	126.508
<i>Année 5</i>	105.637	113.428	125.641	131.147
<i>Année 6</i>	111.118	117.215	129.776	133.000

Tableau 3.8 : série chronologique 1
(période $p = 4$)

L'observation de chaque trimestre est soumise à un effet particulier qui revient tous les ans ; il y a donc 4 variations saisonnières correspondant chacune à un trimestre.

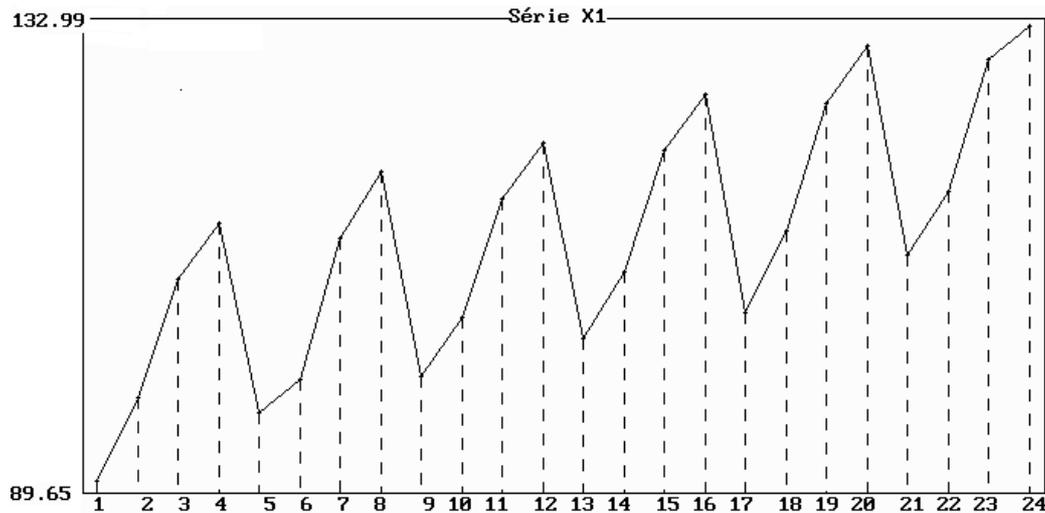


Figure 4.8 : représentation graphique de la série 1
(données observées trimestriellement pendant 6 ans)

Définition : la période notée p des variations saisonnières est la longueur exprimée en unités de temps séparant deux variations saisonnières dues à un même phénomène.

Remarque : nous supposons dans la suite que la série est soumise à des variations saisonnières de même période p . La période est alors le nombre de variations saisonnières. Cette hypothèse n'est pas toujours réalisée au départ : les ventes par tranches horaires d'un hypermarché sont soumises par exemple à une première variation saisonnière due à l'heure et à une seconde due à la journée. Ce cas est traité théoriquement en considérant une période égale au plus petit commun multiple des deux périodes : deux variations saisonnières de périodes 4 et 6 donnent une variation saisonnière de période 12 ($= 3 \times 4 = 2 \times 6$).

Il n'est pas toujours facile de distinguer la tendance lorsque la série chronologique est soumise à des variations saisonnières. La méthode mathématique consiste à calculer les moyennes mobiles en choisissant comme longueur la période des variations saisonnières, de façon à les faire disparaître. Si la moyenne mobile choisie est de longueur différente, les variations saisonnières ne sont pas toujours éliminées (cf. figures 5.8 et 6.8).

Ces moyennes mobiles ont en outre l'avantage d'atténuer les variations accidentelles comme nous l'avons vu précédemment, mais l'inconvénient de n'être définies ni au début ni à la fin de la période observée.

Exemple : on pourra comparer sur les figures 5.8 et 6.8 ci-dessous les moyennes mobiles de longueur 4 (dont les valeurs numériques sont données dans le tableau 4.8) et de longueur 5.

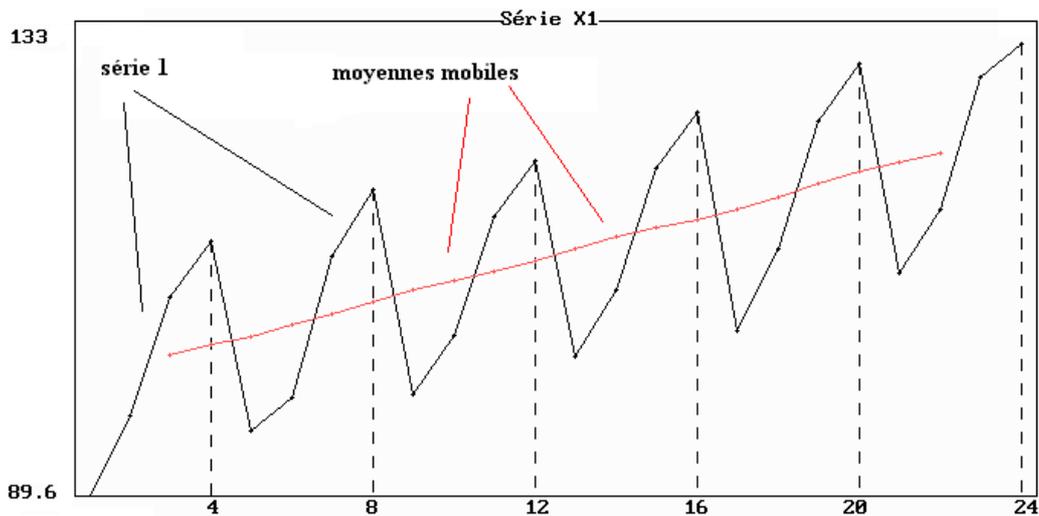


Figure 5.8 : représentation simultanée de la série 1 et des moyennes mobiles de longueur 4

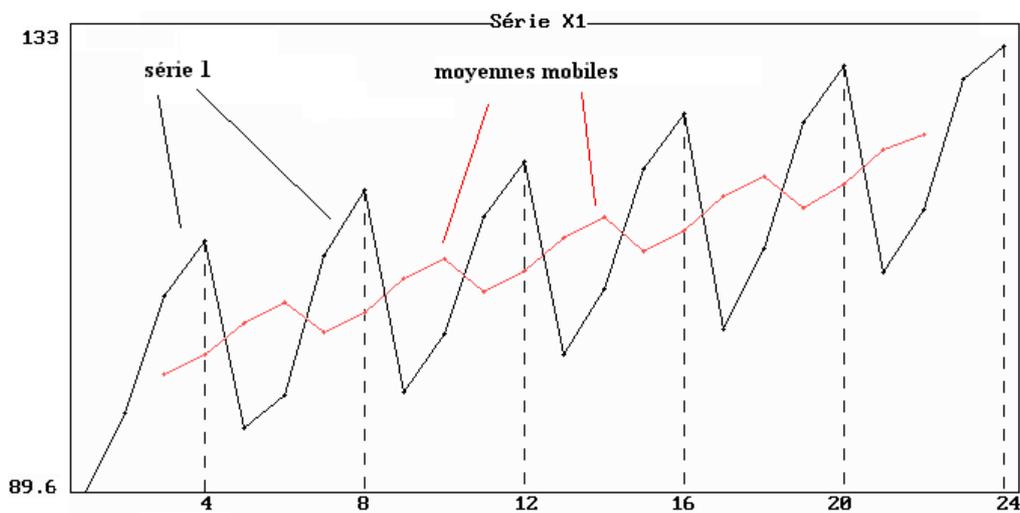


Figure 6.8 : représentation simultanée de la série 1 et des moyennes mobiles de longueur 4

Contrairement aux m.m. de longueur 4, les m.m. de longueur 5 n'éliminent pas les variations saisonnières.

Théorème : les moyennes mobiles d'une série soumise à des variations saisonnières de période p ne sont pas soumises à ces variations saisonnières si leur longueur l est égale à la période p , et plus généralement si leur longueur est un multiple de la période.

Ce théorème est démontré dans un complément pédagogique.

Conclusion : les moyennes mobiles d'une série chronologique dont la tendance est linéaire et les variations saisonnières sont de période p font apparaître la tendance et disparaître les variations saisonnières si leur longueur l est égale à la période p .

2. MODÉLISATION ET DÉSAISONNALISATION.

Un modèle de série chronologique est une équation précisant la façon dont les composantes s'articulent les unes par rapport aux autres pour constituer la série chronologique. Il existe de très nombreux modèles, et parmi eux deux modèles classiques simples : le modèle additif et le modèle multiplicatif, auxquels nous nous limiterons.

Dans les deux modèles présentés, la longueur des moyennes mobiles doit être impérativement égale à la période des variations saisonnières.

Nous avons présenté dans le tableau 3.8 les données sous une forme particulière : en lignes, ce sont les années, et en colonnes les trimestres : le terme x_t correspondant à la t^e observation est alors noté $x_{i,j}$, i donnant l'année (la ligne) et j le trimestre (la colonne).

La relation entre les indices i et j d'une part et l'instant t d'autre part est la suivante :

$$t = (i-1)p + j$$

Exemple :

	$j = 1$	$j = 2$	$j = 3$	$j = 4$
$i = 1$	$t = 1$	$t = 2$	$t = 3$	$t = 4$
$i = 2$	$t = 5$	$t = 6$	$t = 7$	$t = 8$
$i = 3$	$t = 9$	$t = 10$	$t = 11$	$t = 12$

Exemple pour $n = 3$ et $p = 4$

Nous utiliserons cette notation très souvent dans la suite du texte.

2.1 Modèle additif de série chronologique.

La série chronologique x_t se décompose en une tendance notée c_t , des variations saisonnières s_t de période p (égales à $s_1, s_2, s_3, \dots, s_p$) et d'une composante accidentelle e_t .

Le modèle le plus simple est le modèle additif, dans lequel la variation saisonnière s'ajoute simplement à la tendance :

$$\text{pour tout } t = 1, \dots, T \quad x_t = c_t + s_t + e_t$$

Le modèle additif s'exprime donc en général de la façon suivante :

$$\text{pour tout } i = 1, \dots, n \quad \text{pour tout } j = 1, \dots, p \quad x_{i,j} = c_{i,j} + s_j + e_{i,j}$$

Le terme s_j caractérise la variation saisonnière à l'instant j de chaque période i : du trimestre j dans le cas particulier des séries 1 et 2 ($p = 4$), du mois j dans des données mensuelles ($p = 12$) etc.... Les moyennes mobiles seront aussi notées $mm_{i,j}$.

Définition : les termes s_j du modèle additif exprimé sous la forme précédente sont appelés coefficients saisonniers du modèle additif.

On peut calculer la différence entre l'observation et la tendance :

$$\text{pour tout } i = 1, \dots, n \quad \text{pour tout } j = 1, \dots, p \quad x_{i,j} - c_{i,j} = s_j + e_{i,j}$$

Pour un même trimestre, la différence entre l'observation et la tendance est donc à peu près constant et égale à s_j (on suppose que la composante accidentelle est relativement faible).

Nous avons vu précédemment que les moyennes mobiles de longueur l égale à la période des variations saisonnières sont des approximations de la tendance. On peut donc considérer que la différence entre une observation $x_{i,j}$ et la moyenne mobile $mm_{i,j}$ correspondante est à peu près constante pour j fixé :

$$\text{pour tout } i = 1, \dots, n \quad \text{pour tout } j = 1, \dots, p \quad x_{i,j} - mm_{i,j} \approx s_j$$

Cette propriété est recherchée sur la représentation graphique de la série x_t pour déterminer si cette série suit un modèle additif ou non. Elle peut être observée sur la figure 5.8 dans laquelle la tendance est caractérisée par les moyennes mobiles de longueur 4 : les différences entre x_3 et mm_3 , entre x_7 et mm_7 , entre x_{11} et c_{11} sont à peu près constantes, de même les différences entre x_4 et mm_4 , x_8 et mm_8 , x_{12} et mm_{12} etc.

On peut en déduire les différences $x_{i,j} - mm_{i,j}$. Les moyennes mobiles donnant une première approximation de la tendance $c_{i,j}$, les colonnes du tableau des différences contiennent des approximations des coefficients s_j .

Exemple : Les moyennes mobiles et par suite différences $x_{i,j} - mm_{i,j}$ ne sont pas définies aux premier et deuxième trimestres de la première année, ni aux troisième et quatrième trimestres de la dernière (tableaux 4.8 et 5.8).

	1 ^{er} trimestre	2 ^e trimestre	3 ^e trimestre	4 ^e trimestre
Année 1			103.39678	104.44080
Année 2	105.14860	106.25917	107.31233	108.46116
Année 3	109.65950	110.47448	111.27404	112.28937
Année 4	113.42748	114.58360	115.45379	116.22236
Année 5	117.24943	118.38337	119.64839	120.80691
Année 6	121.79719	122.54573		

Tableau 4.8 : moyennes mobiles de longueur 4 de la série 1

	1 ^{er} trimestre	2 ^e trimestre	3 ^e trimestre	4 ^e trimestre
Année 1			5.50932	9.71580
Année 2	-8.94393	-6.86050	5.45047	10.72334
Année 3	-10.05746	-5.28258	5.28226	9.62153
Année 4	-10.15538	-4.93910	5.75471	10.28534
Année 5	-11.61263	-4.95497	5.99271	10.33979
Année 6	-10.67929	-5.33023		

Tableau 5.8 : différences entre les observations et les moyennes mobiles de la série 1

Les différences apparaissant dans une même colonne sont proches les uns des autres et caractérisent le modèle additif.

Les différences $x_{i,j} - mm_{i,j}$ sont donc des approximations des coefficients s_j . Leur moyenne (ou leur médiane), pour chaque colonne j , donne une première estimation s_j' :

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

On obtiendra enfin les estimations définitives s_j en centrant ces termes s_j' :

- on calcule la moyenne des s_j' :

$$m_{s'} = \frac{1}{p} (s_1' + s_2' + \dots + s_p')$$

- on centre en posant :

$$\text{pour tout } j = 1, \dots, p \quad s_j = s_j' - m_{s'}$$

Exemple : le tableau 5.8 donne les différences entre les observations et les moyennes mobiles. On en déduit les moyennes suivantes :

$s_1' = -10.2897$	$s_2' = -5.4735$	$s_3' = 5.5979$	$s_4' = 10.1371$
-------------------	------------------	-----------------	------------------

- On calcule la moyenne des s_j' : $m_{s'} = -0.007039938$
- Les valeurs définitives sont obtenues en posant $s_j = s_j' - m_{s'}$:

$s_1 = -10.2827$	$s_2 = -5.4664$	$s_3 = 5.6049$	$s_4 = 10.1442$
------------------	-----------------	----------------	-----------------

règle de calcul des estimations des coefficients saisonniers du modèle additif

- on calcule les différences entre les observations et les moyennes mobiles ;
- on calcule la moyenne ou la médiane s_j' des différences de chaque colonne du tableau ;
- on calcule la moyenne $m_{s'}$ de ces valeurs s_j' ;
- on obtient les estimations s_j en centrant les valeurs s_j' : $s_j = s_j' - m_{s'}$.

2.2 Modèle multiplicatif de série chronologique.

Le second modèle que nous étudions ici est le modèle multiplicatif suivant :

$$\text{pour tout } t = 1, \dots, T \quad x_t = c_t (1 + s_t) + e_t$$

En présentant les données comme dans le paragraphe précédent, le modèle multiplicatif s'exprime de la façon suivante :

$$\text{pour tout } i = 1, \dots, n \quad \text{pour tout } j = 1, \dots, p \quad x_{i,j} = c_{i,j} (1 + s_j) + e_{i,j}$$

Le terme s_j caractérise la variation saisonnière du trimestre j dans le cas particulier des séries 1 et 2, du mois j dans des données mensuelles etc.

On peut calculer la différence entre l'observation et la tendance :

$$\text{pour tout } i = 1, \dots, n \quad \text{pour tout } j = 1, \dots, p \quad x_{i,j} - c_{i,j} = c_{i,j} s_j + e_{i,j}$$

Considérons le cas particulier $j = 1$ (1^{er} trimestre de l'année i).

$$\text{pour tout } i = 1, \dots, n \quad x_{i,1} - c_{i,1} = c_{i,1} s_1 + e_{i,1}$$

La différence $x_{i,1} - c_{i,1}$ entre l'observation et la tendance est proportionnelle à la tendance $c_{i,1}$: lorsque cette tendance est croissante, la différence augmente, lorsqu'elle est décroissante, il diminue.

Le même raisonnement peut évidemment être tenu pour j fixé quelconque. Les différences permettent ainsi de déterminer si la série chronologique étudiée suit un modèle multiplicatif.

Exemple : on considère la série chronologique ci-dessous :

	1 ^{er} trimestre	2 ^e trimestre	3 ^e trimestre	4 ^e trimestre
Année 1	224.3705	253.2811	201.2421	248.9411
Année 2	274.3802	300.1641	248.9038	298.4386
Année 3	331.9657	371.4032	303.4313	365.9029
Année 4	406.6326	437.9967	361.5774	444.8447
Année 5	488.4166	536.5268	435.5698	549.3614
Année 6	598.0016	659.2896	533.2156	669.2675

Tableau 6.8 : série chronologique 2
(modèle multiplicatif, période $p = 4$)

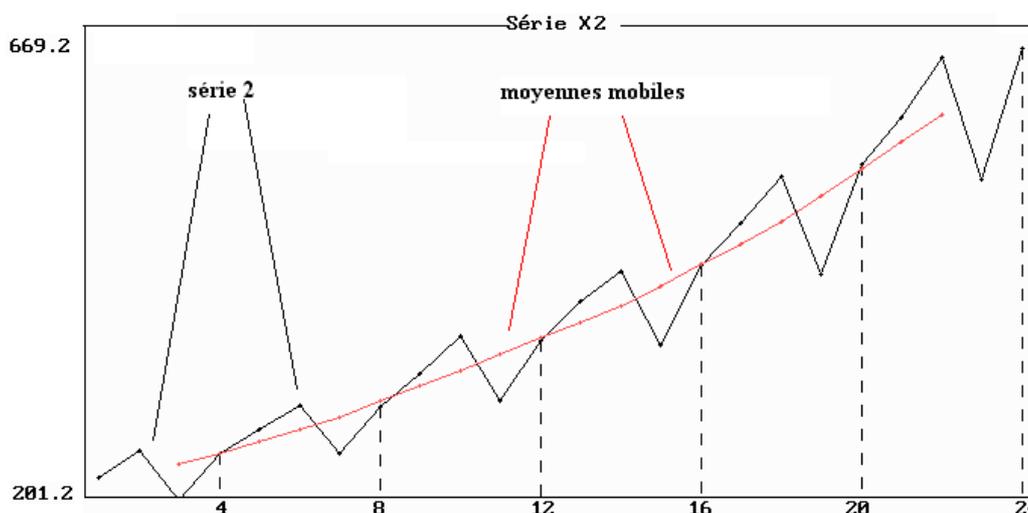


Figure 7.8 : série 2 et moyennes mobiles de longueur 4

Cette série est soumise à des variations saisonnières de période 4 ; la tendance, caractérisée par les moyennes mobiles de longueur 4, est croissante, et la différence entre une observation x_t et la moyenne mobile mm_t a tendance à augmenter pour une même variation saisonnière : les différences entre x_3 et mm_3 , entre x_7 et mm_7 , entre x_{11} et c_{11} augmentent visiblement, de même que les différences entre x_4 et mm_4 , x_8 et mm_8 , x_{12} et mm_{12} etc. (figure 7.8).

Pour quantifier les variations saisonnières, on considère les rapports $x_{i,j} / c_{i,j}$:

$$\text{pour tout } i = 1, \dots, n \text{ pour tout } j = 1, \dots, p \quad x_{i,j} / c_{i,j} = 1 + s_j + e_{i,j} / c_{i,j}$$

En considérant que les variations accidentelles $e_{i,j}$ sont faibles par rapport à la tendance $c_{i,j}$ et en utilisant l'approximation de la tendance par les moyennes mobiles, on constate donc que les rapports x_3 / mm_3 , x_7 / mm_7 , x_{11} / mm_{11} , ... sont à peu près constants et donnent une approximation de $1 + s_3$, de même les rapports x_4 / mm_4 , x_8 / mm_8 , x_{12} / mm_{12} etc. donnent une approximation de $1 + s_4$:

$$\text{pour tout } j = 1, \dots, p \quad \frac{x_{i,j}}{mm_{i,j}} = 1 + s_j = S_j$$

Exemple : les tableaux 7.8 et 8.8 ci-dessous contiennent les moyennes mobiles de la série et les rapports $x_{i,j} / mm_{i,j}$.

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

Tableau 7.8 : moyennes mobiles de longueur 4 de la série 2

	1 ^{er} trimestre	2 ^e trimestre	3 ^e trimestre	4 ^e trimestre
Année 1			0.84481	0.99449
Année 2	1.04670	1.09435	0.86524	0.98244
Année 3	1.03904	1.10952	0.86078	0.98848
Année 4	1.05411	1.08712	0.85482	0.99847
Année 5	1.04566	1.09629	0.84385	1.00761
Année 6	1.04407	1.09890		

Tableau 8.8 : rapports des observations aux moyennes mobiles de la série 2

Les rapports dans chaque colonne du tableau 8.8 sont à peu près constants.

Les rapports $x_{i,j} / mm_{i,j}$ sont donc des approximations des termes $1 + s_j$ que l'on appelle coefficients saisonniers dans le cas du modèle multiplicatif.

Définition : les termes $S_j = 1 + s_j$ du modèle multiplicatif exprimé sous la forme précédente sont appelés coefficients saisonniers du modèle multiplicatif.

On obtient des premières estimations S_j' des coefficients saisonniers en calculant la moyenne (ou la médiane) des rapports figurant dans chaque colonne. Par analogie avec les coefficients saisonniers s_j du modèle additif, dont la moyenne est égale à 0, on cherche des estimations définitives S_j de moyenne 1 :

- on calcule la moyenne :

$$m_{S'} = \frac{1}{p} (S_1' + S_2' + \dots + S_p')$$

- on pose :

$$\text{pour tout } j=1, \dots, p \quad S_j = S_j' / m_{S'}$$

Les coefficients saisonniers estimés S_j sont ainsi de somme p :

$$\begin{aligned} S_1 + S_2 + \dots + S_p &= \frac{1}{m_{S'}} (S_1' + S_2' + \dots + S_p') \\ &= p \end{aligned}$$

ce qui équivaut à une moyenne des s_j égale à 0 puisque l'on a $S_j = 1 + s_j$.

Exemple :

- le tableau 8.8 donne les rapports des observations aux moyennes mobiles.
- on en déduit les moyennes suivantes :

$S_1' = 1.045913$	$S_2' = 1.097236$	$S_3' = 0.8539006$	$S_4' = 0.9942986$
-------------------	-------------------	--------------------	--------------------

- on calcule la moyennes des S_j' : $m_{S'} = .9978371$
- les valeurs définitives sont obtenues de façon que les S_j' soient de moyenne 1 :

$S_1 = 1.04818$	$S_2 = 1.099614$	$S_3 = 0.8557515$	$S_4 = 0.9964539$
-----------------	------------------	-------------------	-------------------

<u>règle de calcul des estimations des coefficients saisonniers du modèle multiplicatif</u>
• on calcule les rapports des observations aux moyennes mobiles ;
• on calcule la moyenne ou la médiane des rapports S_j' de chaque colonne du tableau ;
• on calcule la moyenne $m_{S'}$ de ces valeurs ;
• on obtient les estimations S_j en posant $S_j = S_j' / m_{S'}$.

2.3 Désaisonnalisation.

Les coefficients saisonniers permettent d'éliminer d'une observation les effets de la variation saisonnière correspondante. On obtient ainsi les valeurs corrigées des variations saisonnières, ou encore les valeurs désaisonnalisées.

L'avantage de cette désaisonnalisation est de permettre la comparaison de deux observations soumises à des variations saisonnières différentes.

définition : on appelle observation corrigée des variations saisonnières la valeur $x_{i,j}'$ obtenue en éliminant l'effet saisonnier sur la valeur $x_{i,j}$.

modèle additif :	$x_{i,j}' = x_{i,j} - S_j$
modèle multiplicatif :	$x_{i,j}' = x_{i,j} / S_j$

Les valeurs corrigées des variations saisonnières (expression souvent abrégée par c.v.s.) caractérisent à la fois la tendance et la variation accidentelle.

Exemple : on donne ci-dessous les quatre dernières observations de la série 2 (année 6) et les valeurs corrigées des variations saisonnières :

	1 ^{er} trimestre	2 ^e trimestre	3 ^e trimestre	4 ^e trimestre
observations :	598.00160	659.28960	533.21560	669.26750
valeur c.v.s. :	570.51396	599.56452	623.09629	671.64924

L'observation du deuxième trimestre est largement supérieure à celle du troisième, mais c'est l'inverse pour les valeurs c.v.s. : la tendance est restée croissante au troisième trimestre.

Supposons que l'observation du premier trimestre de l'année 7 soit égale à 720.15. Pour savoir si la tendance est restée à la hausse, on calcule la valeur désaisonnalisée :

$$x_{7,1}' = 720.15 / 1.04818 = 687.04771$$

et on la compare à la valeur désaisonnalisée du quatrième trimestre de l'année précédente :

$$x_{6,4}' = 671.649$$

La valeur c.v.s. $x_{7,1}'$ est supérieur à $x_{6,4}'$. La tendance est restée à la hausse si la différence est supérieure à la variation accidentelle . Il faudrait donc comparer cette différence à

l'écart type des variations accidentelles, calculé sur les données antérieures. Il semble que dans la pratique, cette comparaison ne soit guère effectuée.

3. FILTRE DE BUYS-BALLOT.

Le filtre de Buys-Ballot concerne les séries chronologiques suivant un modèle additif et dont la tendance est linéaire. Il consiste à estimer les paramètres de ce modèle suivant le critère des moindres carrés, et permet ensuite, dans la mesure où les hypothèses sont respectées, d'effectuer des prévisions. Il s'agit en fait d'une régression linéaire multiple particulière.

Lorsque la série suit le modèle multiplicatif et que la tendance est exponentielle, les logarithmes des observations vérifient les conditions précédentes. On peut alors appliquer le filtre de Buys-Ballot.

3.1 Filtre de Buys-Ballot.

Nous supposons donc que la série étudiée suit le modèle additif et que la tendance est linéaire :

$$c_t = \beta t + \alpha$$

Pour exprimer la tendance en fonction de la ligne i et de la colonne j du tableau, nous utilisons la relation et les variables t , i et j données précédemment :

$$t = (i-1)p + j$$

dans laquelle i varie de 1 à n , j de 1 à p . Le nombre total T d'observations est égal à np .

Le modèle complet est le suivant :

$\text{pour tout } i = 1, \dots, n \quad \text{pour tout } j = 1, \dots, p \quad x_{i,j} = \beta [(i-1)p + j] + \alpha + \sigma_j + \varepsilon_{i,j}$
--

Les coefficients en caractères grecs sont des coefficients théoriques qu'il s'agit d'estimer : on retrouve ici la notation employée dans le chapitre précédent. Les observations de la variable expliquée sont notées ici $x_{i,j}$, et les variables explicatives sont le temps t et p variables particulières qui n'apparaissent pas directement dans la formule et dont les coefficients de régression sont les variations saisonnières σ_j .

Le critère des moindres carrés consiste à déterminer les paramètres b , a , s_1 , s_2 , ..., s_p de façon à minimiser la somme des carrés des différences $e_{i,j} = x_{i,j} - [(b(i-1)p + j) + a + s_j]$ entre la valeur observée $x_{i,j}$ et la valeur estimée par le modèle $b[(i-1)p + j] + a + s_j$.

Les valeurs obtenues sont des estimations des paramètres théoriques β , α , σ_j . Le coefficient de corrélation du modèle est le coefficient de corrélation entre les valeurs observées $x_{i,j}$ et les valeurs estimées $x_{i,j}'$.

On calculera les estimations des paramètres à l'aide des formules suivantes :

$$b = \frac{2}{p(n^2 - 1)} \left[\sum_{i=1}^n i m_{i\bullet} - \frac{n(n+1)}{2} m \right]$$

$$\begin{aligned} a &= m - b(n p + 1)/2 \\ s_j &= m_{\bullet j} - m - b [j - (p + 1)/2] \end{aligned}$$

avec les notations suivantes :

m : moyenne de la totalité des observations

$m_{i\bullet}$: moyenne des observations de la ligne i

$m_{\bullet j}$: moyenne des observations de la colonne j

définition : les termes $e_{i,j} = x_{i,j} - [(b(i-1)p + j) + a + s_j]$ sont appelés résidus.

Comme en régression, la variance des résidus s^2 dépend du coefficient de corrélation r et de la variance σ_x^2 des observations $x_{i,j}$ de la variable expliquée :

$$s^2 = \sigma_x^2 (1 - r^2)$$

L'interprétation du coefficient de corrélation est délicate dans le cas des séries chronologiques. Il est souvent très élevé (une valeur de 0.98 n'est pas rare) sans que l'on puisse en déduire directement que l'ajustement obtenu est satisfaisant. Cette particularité est due au fait que la variable temps est ordonnée. Dès lors, il suffit que la tendance soit croissante ou décroissante pour que le coefficient de corrélation soit élevée. Mais il ne donne aucun renseignement sur la nature de la tendance, qui peut être exponentielle ou linéaire etc.

Exemple : le modèle additif et l'hypothèse de linéarité de la tendance de la série 1 sont confirmés par la représentation graphique (figure 8.8).

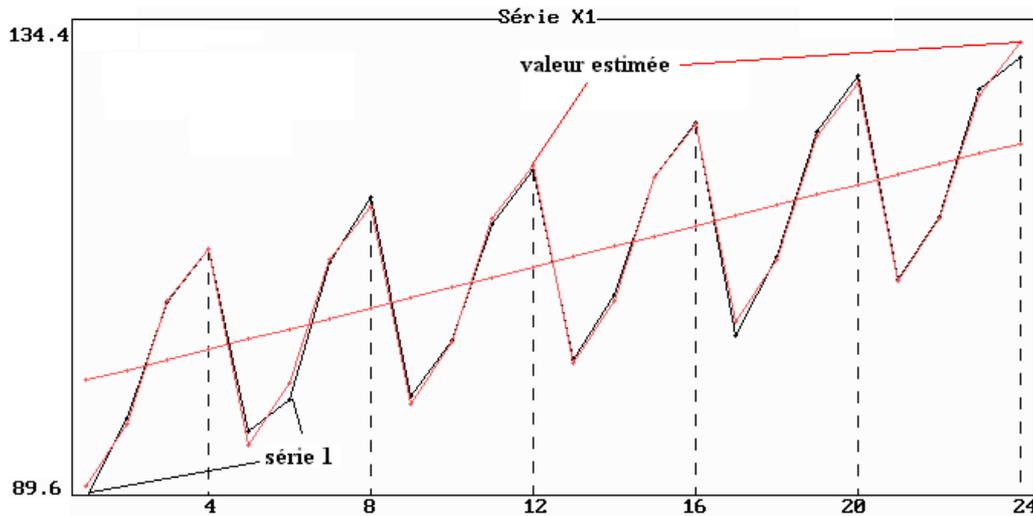


Figure 8.8 : représentation simultanée de la série 1, de la tendance et des valeurs estimées par le modèle

Nous donnons les résultats partiels suivants :

années	moyennes annuelles m_i	produits $i m_i$
1	102.57846	102.57846
2	106.88766	213.77531
3	110.81528	332.44585
4	115.15820	460.63280
5	118.96325	594.81625
6	122.77737	736.66425

On en déduit le modèle estimé par le filtre de Buys-Ballot :

$b = 1.011173$	$a = 100.2237$
$s_1 = -10.43134$	$s_2 = -5.27912$
$s_3 = 5.77288$	$s_4 = 9.93759$
$r = 0.99791$	$s^2 = 0.5641465$

Tableau 9.8 : paramètres du modèle linéaire additif estimés par le filtre de Buys-Ballot sur la série 1

L'ajustement peut être considéré comme très précis puisque le coefficient de corrélation entre les observations $x_{i,j}$ et les estimations $x_{i,j}'$ est égal à 0.99791.

La série 1 n'est pas une série réelle : elle a été obtenue par simulation du modèle linéaire additif avec comme paramètres théoriques :

$$x_t = \beta t + \alpha + \sigma_t + \varepsilon_t$$

$$\beta = 1, \alpha = 100, \varepsilon_t \approx N(0,1), \sigma_1 = -10, \sigma_2 = -5, \sigma_3 = 5, \sigma_4 = 10.$$

Les valeurs estimées sont visiblement très proches des valeurs théoriques. La méthode statistique permet donc de retrouver à partir des observations les valeurs des paramètres utilisés pour générer les données. C'est le moins qu'on pouvait en attendre.

3.2 Validation du modèle linéaire et prévision.

Pour effectuer la prévision $x_{i,j}^p$ de la série chronologique à l'instant $t = (i-1)p + j$, on remplace dans le modèle théorique les paramètres par leurs estimations :

$$x_{i,j}^p = b [(i-1)p + j] + a + s_j$$

Ces prévisions ponctuelles peuvent être complétées par des prévisions par intervalle de confiance, comme en régression, mais nous ne donnerons pas les formules trop complexes pour être utilisées facilement.

Exemple : Les valeurs estimées par le modèle sont données par la formule :

$$x_{i,j} = 1.011173 [(i-1)4 + j] + 100.2237 + s_j$$

avec :

$s_1 = -10.43134$	$s_2 = -5.27912$	$s_3 = 5.77288$	$s_4 = 9.93759$
-------------------	------------------	-----------------	-----------------

Les prévisions concernant l'année 7 sont les suivantes :

$x_{7,1}^p = 115.07169$	$x_{7,2}^p = 121.23509$	$x_{7,3}^p = 133.29826$	$x_{7,4}^p = 138.47414$
-------------------------	-------------------------	-------------------------	-------------------------

Les prévisions que l'on peut effectuer après l'estimation des paramètres ne sont justifiées que dans la mesure où les hypothèses du modèle sont respectées.

Une hypothèse fondamentale pour la prévision et souvent négligée est que les conditions dans lesquelles la série chronologique évolue sont les mêmes à la date de la prévision que dans le passé. Il faut noter que cette condition n'est pas toujours vérifiée, par l'effet de la prévision elle-même : par exemple, un hypermarché qui prévoit une baisse de son chiffre d'affaires va prendre des mesures de réduction de coût, augmenter sa publicité etc., de façon à augmenter ses ventes : les décisions vont donc à l'encontre de la prévision.

Il est indispensable en outre de contrôler statistiquement le modèle. Pour cela, on étudie les résidus $e_{i,j}$. Ces résidus possèdent les propriétés mathématiques habituelles en régres-

sion : ils sont centrés, indépendants de la variable explicative (ici le temps). Les résidus de chaque trimestre sont en outre de moyenne nulle.

Pour que le modèle soit valide, leur répartition doit être proche de la loi normale. Ils ne doivent présenter pas d'évolution particulière dans le temps, ce que l'on peut vérifier par une représentation graphique ou par des tests (tests de Durbin et Watson, sur le coefficient d'autocorrélation de rang 1).

Rappelons enfin que la prévision $x_{i,j}^p$ est la prévision de la moyenne des observations pour t fixé : l'intervalle de confiance donné par certains logiciels est celui de cette moyenne, et non de la valeur individuelle à l'instant t .

4. LISSAGE EXPONENTIEL.

4.1 Généralités sur le lissage exponentiel.

Le lissage exponentiel est une classe de méthodes de lissage de séries chronologiques dont l'objectif est la prévision à court terme. Ces méthodes sont fondées sur une hypothèse fondamentale : chaque observation à l'instant t dépend des observations précédentes et d'une variation accidentelle, et cette dépendance est plus ou moins stable dans le temps.

L'estimation x_t^p de la série à l'instant t connaissant x_{t-1} est donc obtenue par une formule de la forme :

$$x_t^p = \alpha_{t-1} x_0 + \alpha_{t-2} x_1 + \alpha_{t-3} x_2 + \dots + \alpha_0 x_{t-1}$$

On peut généraliser cette formule au cas où seule l'observation x_{t-h} est connue :

$$x_t^p = \alpha_{t-h} x_0 + \alpha_{t-h-1} x_1 + \alpha_{t-h-2} x_2 + \dots + \alpha_0 x_{t-h}$$

Dans le premier cas, on peut prévoir à l'horizon 1 : sachant x_t , on prévoit x_{t+1} , l'observation suivante, et dans le cas général, on peut prévoir à l'horizon h : sachant x_t , on prévoit x_{t+h} . On peut penser que plus l'horizon h est faible, meilleure est la prévision.

Les méthodes de lissage exponentiel consistent à choisir les coefficients α_j en fonction de la série étudiée.

- Dans le cas où la série est stationnaire (on ne distingue pas de tendance à la hausse ni à la baisse), on utilise le lissage exponentiel simple. Les coefficients α_j sont de la forme :

$$\alpha_j = \alpha^j / (1 - \alpha)$$

où α est la constante de lissage choisie de façon empirique entre 0 et 1. La valeur $1/\alpha$ est parfois appelée âge moyen du lissage.

- Lorsque la série présente une tendance linéaire par morceaux (la tendance peut être considérée comme linéaire sur une suite de quelques observations), et n'est soumise à aucune variation saisonnière, on effectue les prévisions à l'aide du double lissage exponentiel. En effet, en appliquant deux fois le lissage exponentiel simple, on détermine une suite de droites de tendance qui ajustent les observations.

- Dans le cas d'une série soumise à des variations saisonnières, on utilise souvent le modèle de Holt et Winters, que nous expliquons rapidement dans le paragraphe 4.2.

Dans les lissages exponentiels simple et double, il y a donc une seule constante à fixer. Le choix peut être empirique, c'est-à-dire effectué par l'utilisateur en fonction de la connaissance qu'il a de la série. Lorsque la valeur x_t ne dépend guère que des 3 ou 4 dernières observations, on peut choisir la constante α proche de 1. Inversement, si les observations antérieures gardent longtemps une influence sur la valeur x_t , on choisira α proche de 0. On peut aussi déterminer la constante de façon à minimiser la somme des carrés des erreurs commises pour l'horizon h fixé. Certains programmes proposent cette option.

Exemple : la figure 9.8 ci-dessous montre l'ajustement obtenu à l'horizon 1 du cours de l'action Alcatel par un double lissage exponentiel.

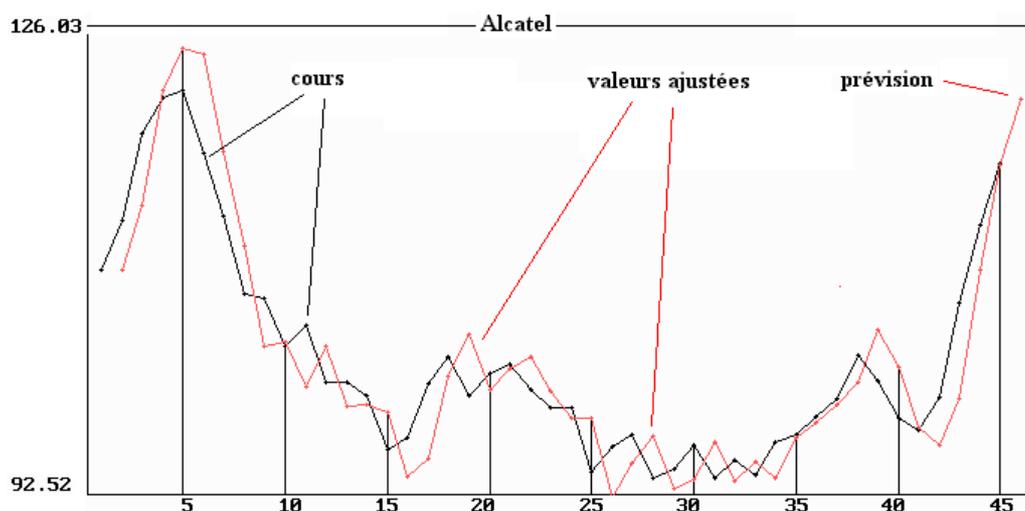


Figure 9.8 : cours du titre Alcatel et prévision à l'horizon 1 par double lissage exponentiel ($\alpha = 0.65$)

La constante de lissage a été déterminée de façon à minimiser la somme des carrés des erreurs ($\alpha = 0.65$). La valeur prévue suit la série avec retard.

Le cours prévu à l'instant $t = 46$ (8 mars 1999) est égal à 122.22 € ; le cours moyen observé de la séance est de 118 € : la prévision n'est pas très satisfaisante. Les méthodes de prévision ne donnent pas de bons résultats sur les cours en Bourse (sinon, cela se saurait !).

4.2 Modèle de Holt et Winters.

La méthode de Holt et Winters permet en effet d'effectuer des prévisions sur des séries chronologiques assez irrégulières et soumises ou non à des variations saisonnières suivant un modèle additif ou multiplicatif.

Elle consiste en trois lissages exponentiels simultanés. On définit donc trois paramètres, notés α , β et γ . A chaque instant t , elle donne une estimation :

- de la tendance
- du coefficient saisonnier correspondant
- de la valeur observée.

On peut choisir les coefficients arbitrairement : faibles si l'on considère que la valeur à l'instant t dépend d'un grand nombre d'observations antérieures, élevés dans le cas contraire. On peut aussi en calculer les valeurs optimales, en minimisant la somme des carrés des différences entre les valeurs observées et estimées. On procède ensuite aux prévisions, en considérant que la tendance suit un modèle linéaire additif ou multiplicatif à très court terme.

Exemple : *on étudie ici la série constituée du nombre trimestriel de naissances dans la région Centre. La figure 10.8 montre une tendance assez irrégulière, avec des variations saisonnières de période 4.*

Le filtre de Buys-Ballot n'est visiblement pas adapté dans ce cas particulier, et nous appliquons ici le modèle de Holt et Winters. Nous choisissons comme modèle un modèle dont la tendance à court terme est linéaire (ici, on se limite à examiner une succession de quelques points), et les variations de période 4 additives. Nous fixons l'horizon à 4 pour disposer des prévisions de toute l'année 1986.

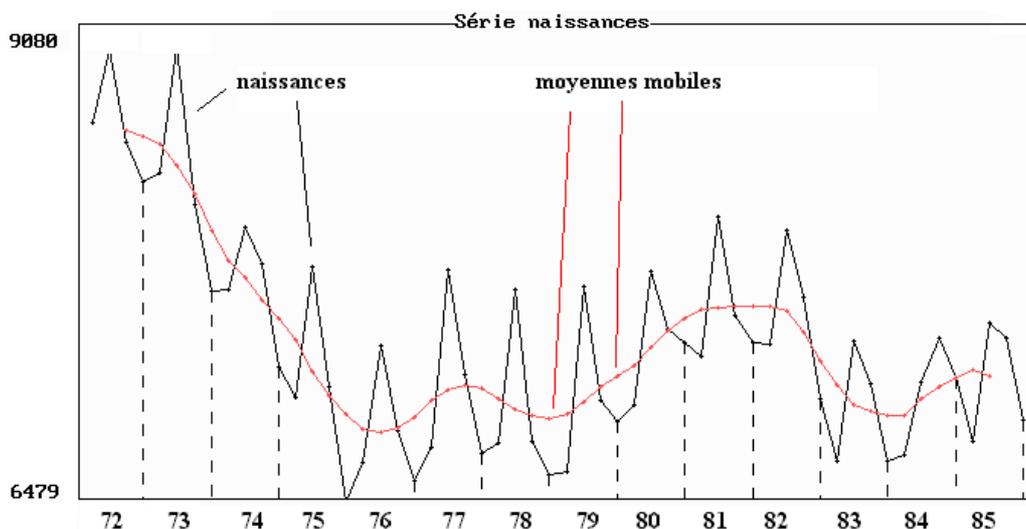


Figure 10.8 : nombre de naissances dans la région Centre de 1972 à 1985
Moyennes mobiles de longueur 4.

Les paramètres optimaux que l'on obtient en minimisant la somme des carrés de la totalité des résidus sont : $\alpha = 0.9$, $\beta = 0.9$, $\gamma = 0.1$.

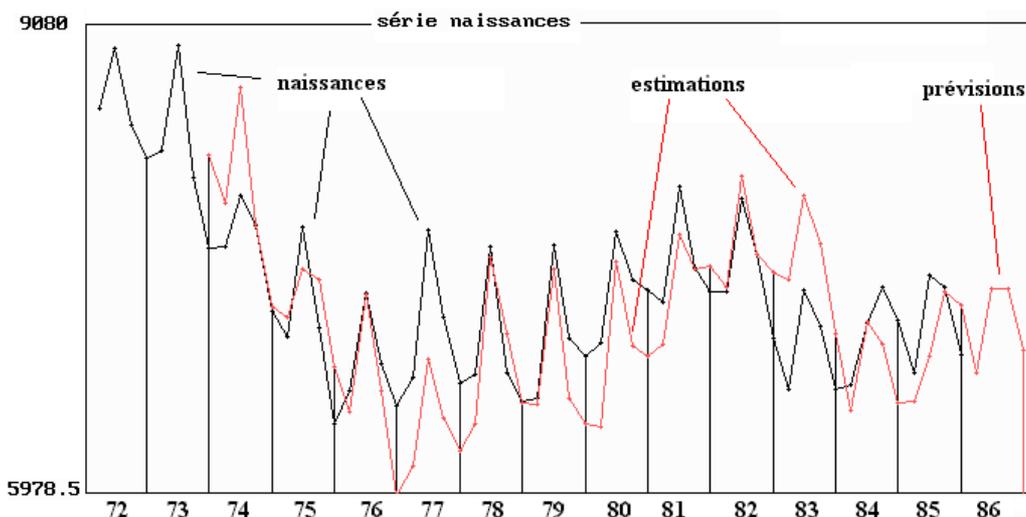


Figure 11.8 : nombre de naissances dans la région Centre de 1972 à 1985
Estimations par le modèle de Holt et Winters et prévision pour 1986

Le tableau 10.8 permet de comparer prévisions et observations de l'année 1986 : les naissances aux deux premiers trimestres ne sont pas très bien estimées, contrairement aux deux dernières.

	1 ^{er} trimestre	2 ^e trimestre	3 ^e trimestre	4 ^e trimestre
observations	7018	7720	7415	6986
prévisions	6824	7410	7405	6984

Tableau 10.8 : nombre de naissances dans la région Centre observés et prévus en 1986

CONCLUSION

L'analyse des séries chronologiques est un des objectifs fondamentaux de la statistique. Nous insistons sur le fait, que quelle que soit la méthode utilisée, il faut être vigilant sur les prévisions effectuées qui peuvent être dans certains cas totalement aberrantes (cf. Bensaïer et Bleuse-Trillon, p. 123, à propos du modèle de Holt et Winters).

Nous conseillons fortement aux praticiens de se limiter aux méthodes qu'ils connaissent lorsqu'ils effectuent leurs prévisions. Le logiciel qu'ils utilisent doit être sûr, contrôlé : il nous est arrivé d'obtenir des résultats différents sur les mêmes données et par la même méthode en employant deux logiciels différents, ou d'aboutir des prévisions manifestement fausses puisque différentes des valeurs données dans les ouvrages de référence.

TABLE DES MATIERES

1. DESCRIPTION D'UNE SÉRIE CHRONOLOGIQUE.	1
1.1 Description de la tendance.....	2
1.2 Description simultanée des variations saisonnières et de la tendance.....	6
2. MODÉLISATION ET DÉSAISONNALISATION.....	9
2.1 Modèle additif de série chronologique.....	9
2.2 Modèle multiplicatif de série chronologique.....	12
2.3 Désaisonnalisation.....	16
3. FILTRE DE BUYS-BALLOT.....	17
3.1 Filtre de Buys-Ballot.....	17
3.2 Validation du modèle linéaire et prévision.....	20
4. LISSAGE EXPONENTIEL.....	21
4.1 Généralités sur le lissage exponentiel.....	21
4.2 Modèle de Holt et Winters.....	23
CONCLUSION.....	25