esa

COLLECTION « ÉCONOMIE ET STATISTIQUES AVANCÉES »

Série : Ecole Nationale de la Statistique et de l'Administration Economique et Centre d'Etudes des Programmes Economiques

Christian GOURIEROUX **Alain MONFORT**

Professeurs à l'ENSAE

SERIES TEMPORELLES ET MODELES DYNAMIQUES

2e édition

Economica

Dans la même série

ANTONIADIS A., BERRUYER J. et CARMONA R., *Régression non linéaire et applications.*

ARTUS P., DELEAU M. et MALGRANGE P., *Modélisation macro-économique.*

BOSQ D. et LECOUTRE J.-P., *Théorie de l'estimation fonctionnelle.*

BRILLET J.-L. *Modélisation économétrique – Principes et techniques.*

FRACHOT A. et GOURIEROUX C., *Titrisation et remboursements anticipés.*

GOURIEROUX C., *Econométrie des variables qualitatives (2e éd.).*

GOURIEROUX C., *Modèles Arch et applications financières.*

GOURIEROUX C., *Théorie des sondages.*

GOURIEROUX C. et MONFORT A., *Séries temporelles et modèles dynamiques (2e éd.).*

Statistique et modèles économétriques.

— Vol. 1 : Notions générales, estimation, prévision, algorithmes.

— Vol. 2 : Tests, régions de confiance, choix de modèles, théorie asymptotique.

GRANDMONT J.-M., *Monnaie et valeur.*

HENRIET D. et ROCHET J.-C., *Microéconomie de l'assurance.*

LAFFONT J.-J., *Cours de théorie microéconomique.*

— Vol. 1 : Fondements de l'économie publique. Cours de théorie microéconomique.

— Vol. 2 : Economie de l'incertain et de l'information.

LESOURNE J., *Economie de l'ordre et du désordre.*

LOLLIVIER S., *Exercices de microéconomie.*

MICHEL G., *L'économie française.*

MONFORT A., *Cours de probabilités (2e éd.).*

MONFORT A., *Cours de statistique mathématique (2e éd.).*

MUET P.-A., *Théories et modèles de la macroéconomie – Tome 1 : L'équilibre de courte période (4e éd.).*

PATAT J.-P., *Monnaie, institutions financières et politique monétaire (5e éd.).*

PISANI-FERRY J., STERDYNIAK H. et VILLA P., *Problèmes de macroéconomie.*

SALANIÉ B., *Théorie des contrats.*

TIROLE J., *Théorie de l'organisation industrielle – Tome 1 et Tome 2.*

VOLLE M., *Analyse des données (3e éd.).*

2154685 51

COLLECTION « ÉCONOMIE ET STATISTIQUES AVANCÉES »

Série : Ecole Nationale de la Statistique et de l'Administration Economique et Centre d'Etudes des Programmes Economiques

Christian GOURIEROUX **Alain MONFORT**

Professeurs à l'ENSAE

SERIES TEMPORELLES ET MODELES DYNAMIQUES

2e édition

8° V
110478

ECONOMICA
49, rue Héricart, 75015 Paris

DL- 30 11 1995 38002

© Ed. ECONOMICA, 1995

Tous droits de reproduction, de traduction, d'adaptation et d'exécution réservés pour tous les pays.

Avant-propos

L'étude de l'évolution temporelle d'une ou plusieurs variables a, depuis longtemps, intéressé des spécialistes de domaines très variés. Les probabilistes ont construit une importante théorie mathématique, la théorie des processus stochastiques, qui permet d'analyser avec un très grand degré de finesse les propriétés d'un ensemble de variables aléatoires indicées par le temps. Les théoriciens de la statistique ont proposé diverses approches des problèmes d'estimation, de tests, de prévision et de filtrage pour les processus stochastiques. Des praticiens appartenant à des domaines très variés (physique, automatique, économie,...) ont développé des techniques adaptées à leurs problèmes spécifiques et ont ainsi créé une grande gamme d'outils d'analyse qui n'est pas toujours reliée très clairement à la littérature théorique.

Devant cet ensemble vaste et hétéroclite, l'étudiant intéressé par l'analyse des séries économiques peut se trouver quelque peu désemparé, et l'objectif de ce livre est de l'aider le plus possible dans son apprentissage. Pour que cette aide soit efficace, il nous a semblé qu'il fallait éviter deux écueils : une abstraction trop grande et, à l'opposé, un empirisme sans aucun support théorique. Nous avons donc essayé de progresser au mieux entre ces écueils en n'hésitant pas à prendre quelque liberté avec la rigueur mathématique quand son coût devenait trop élevé ou risquait de nuire à l'intuition et, à l'inverse, en mettant en lumière les éventuelles faiblesses théoriques de techniques d'usage courant.

Par ailleurs, nous avons tenté de présenter de manière synthétique des problèmes d'origines diverses : des problèmes d'origine statistique comme la désaisonnalisation et la prévision, des problèmes d'origine économétrique comme la causalité, l'exogénéité, la cointégration, les anticipations et des problèmes issus de l'automatique comme le filtrage et le lissage de Kalman. Nous pensons en effet que ces diverses approches sont utiles à la modélisation des séries économiques et qu'elles permettent d'apporter des éclairages complémentaires à de nombreuses questions.

Le livre est organisé de la façon suivante. Le premier chapitre est consacré à une présentation générale des problèmes et des méthodes de traitement. Le deuxième chapitre montre, de façon détaillée, comment la régression linéaire peut être appliquée aux problèmes liés à la saisonnalité et à la prévision. Les méthodes de désaisonnalisation par des moyennes mobiles sont décrites dans le chapitre 3. Le chapitre 4 présente l'application des méthodes de lissage exponentiel au problème de la prévision. Le chapitre 5 fournit les outils probabilistes nécessaires à l'étude des méthodes de prévision récentes proposées par Box et Jenkins ; ces méthodes sont décrites dans le chapitre 6. Le chapitre 7 introduit la notion de séries temporelles multivariées et le chapitre 8 étudie leurs diverses représentations. Le chapitre 9 s'intéresse aux problèmes statistiques (estimations et tests) posés par les séries stationnaires. Le chapitre 10 utilise les techniques de séries temporelles pour étudier des notions économétriques de causalité, d'exogénéité et de multiplicateurs. Le chapitre 11 aborde les problèmes de modélisation économétrique spécifiques aux séries non stationnaires. Le chapitre 12 traite des anticipations. Les chapitres 13 et 14 regroupent les problèmes statistiques liés aux modèles économétriques stationnaires ou non stationnaires. Le chapitre 15 présente les modèles espace d'états et le filtrage de Kalman, tandis que le chapitre 16 présente des applications de ces techniques à l'économétrie. La deuxième édition intègre les développements récents de l'économétrie des processus non stationnaires.

CHAPITRE 1

Généralités

1.1 Définition d'une série temporelle

La mise en évidence de mécanismes économiques et la prise de décisions supposent que l'on dispose de plusieurs observations des variables importantes afin d'étudier les liaisons qui existent entre elles. L'impossibilité d'expériences contrôlées dans la plupart des domaines économiques fait que ces observations sont souvent obtenues au moyen d'enquêtes ou par exploitation de divers fichiers et sont disponibles, par exemple, une fois par an ou une fois pa trimestre. Il s'agit donc d'observations répétées, correspondant à des dates différentes. La suite des observations d'une variable $(Y_t,\ t \in \mathcal{T})$ est appelée *série temporelle* ou *série chronologique*.

Les dates d'observations sont en général équidistantes les unes des autres : c'est le cas de séries mensuelles, trimestrielles... L'année contient alors un nombre entier s d'intervalles séparant deux dates d'observations successives : $s = 12$ pour une série mensuelle, $s = 4$ pour une série trimestrielle... Les dates équidistantes sont dans la suite indexées par des entiers : $t = 1, ..., T$ où T désigne le nombre d'observations.

Signalons que dans des domaines autres qu'économiques, certains caractères (température, hygrométrie,...) sont observés de façon continue dans le temps au moyen d'appareils physiques. L'indice t est alors à valeur dans un intervalle de $\mathbb{R}$. Ce cas est très différent du précédent, puisqu'une infinité d'observations est disponible ; il ne sera pas étudié dans la suite.

Les observations peuvent correspondre à des flux ou à des niveaux. La consommation mensuelle de pétrole est un flux ; chaque observation correspond à une période. Pour que ces observations soient directement comparables, il faudrait que les longueurs des périodes soient les mêmes. Dans le cas des consommations mensuelles, cette longueur est de 28, 29, 30 ou 31 jours et il peut être nécessaire de corriger la série brute. Une façon simple d'y parvenir consiste à calculer la consommation journalière moyenne pour chacun des mois. Dans le cas d'un niveau, les observations

correspondent à des dates précises. La valeur journalière du dollar en francs à l'ouverture du marché boursier est un exemple de niveau. Notons qu'un flux peut toujours être considéré comme la différence de deux niveaux.

1.2 Exemples de séries temporelles

Donnons deux exemples de séries chronologiques qui seront utilisés dans la suite pour illustrer les diverses méthodes étudiées.

A. Indice mensuel des prix à la consommation

Tableau 1.1

Mois / Année	Janv.	Février	Mars	Avril	Mai	Juin	Juil.	Août	Sept.	Oct.	Nov.	Déc.
1970	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	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	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	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	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	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	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	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	190,3	191,7	193,4	195,5	197,4	198,9	201,5	202,5	203,8	205,7	206,8	207,8

base 100 en juillet 1970.

Source : INSEE.

B. Trafic voyageur

Les données concernent le trafic voyageur de la SNCF en deuxième classe. Elles sont exprimées en millions de voyageurs kilomètres. Les observations mensuelles portent sur la période 1963-1980.

Tableau 1.2

Mois / Année	Janv.	Fév.	Mars	Avril	Mai	Juin	Juil.	Août	Sept.	Oct.	Nov.	Déc.
1963	1 750	1 560	1 820	2 090	1 910	2 410	3 140	2 850	2 090	1 850	1 630	2 420
1964	1 710	1 600	1 800	2 120	2 100	2 460	3 200	2 960	2 190	1 870	1 770	2 270
1965	1 670	1 640	1 770	2 190	2 020	2 610	3 190	2 860	2 140	1 870	1 760	2 360
1966	1 810	1 640	1 860	1 990	2 110	2 500	3 030	2 900	2 160	1 940	1 750	2 330
1970	1 850	1 590	1 880	2 210	2 110	2 480	2 880	2 670	2 100	1 920	1 670	2 520
1968	1 834	1 792	1 860	2 138	2 115	2 485	2 581	2 639	2 038	1 936	1 784	2 391
1969	1 798	1 850	1 981	2 085	2 120	2 491	2 834	2 725	1 932	2 085	1 856	2 553
1970	1 854	1 823	2 005	2 418	2 219	2 722	2 912	2 771	2 153	2 136	1 910	2 537
1971	2 008	1 835	2 120	2 304	2 264	2 175	2 928	2 738	2 178	2 137	2 009	2 546
1972	2 084	2 034	2 152	2 522	2 318	2 684	2 971	2 759	2 267	2 152	1 978	2 723
1973	2 081	2 112	2 279	2 661	2 281	2 929	3 089	2 803	2 296	2 210	2 135	2 862
1974	2 223	2 248	2 421	2 710	2 505	3 021	3 327	3 044	2 607	2 525	2 160	2 876
1975	2 481	2 428	2 596	2 923	2 795	3 287	3 598	3 118	2 875	2 754	2 588	3 266
1976	2 667	2 668	2 804	2 806	2 976	3 430	3 705	3 053	2 764	2 802	2 707	3 307
1977	2 706	2 586	2 796	2 978	3 053	3 463	3 649	3 095	2 839	2 966	2 863	3 375
1978	2 820	2 857	3 306	3 333	3 141	3 512	3 744	3 179	2 984	2 950	2 896	3 611
1979	3 313	2 644	2 872	3 267	3 391	3 682	3 937	3 284	2 849	3 085	3 043	3 541
1980	2 848	2 913	3 248	3 250	3 375	3 640	3 771	3 259	3 206	3 269	3 181	4 008

1.3 Représentation graphique

Les séries sont habituellement présentées dans un repère orthornormé comportant en ordonnée les valeurs des observations et en abscisse les dates correspondantes. L'évolution de l'indice mensuel des prix à la consommation entre les mois de janvier 1970 et de décembre 1978 (tableau 1.1) est par exemple donnée par la figure 1.3.

L'échelle des temps choisie (unité petite) permet de mettre en évidence l'évolution à moyen terme de la série. Celle-ci semble présenter une rupture à la fin de l'année 1973, associée à la première augmentation brutale des prix du pétrole.

Définissons le taux d'accroissement de l'indice entre les dates $t-1$ et t par :

$$\delta I_t = \frac{I_t - I_{t-1}}{I_{t-1}}$$

où I_t est la valeur de l'indice pour le mois t.

Les valeurs numériques de cet accroissement se déduisent des valeurs figurant dans le tableau 1.1

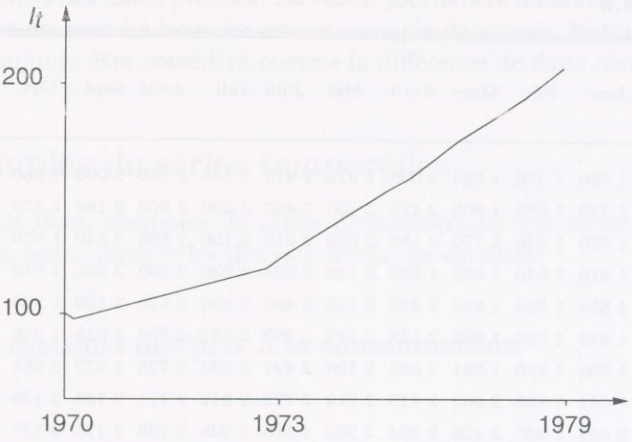

Figure 1.3

Tableau 1.4 *Accroissement relatif mensuel de l'indice des prix à la consommation entre les mois de février 1970 et décembre 1978 (en pourcentage)*

Mois / Année	Janv.	Fév.	Mars	Avril	Mai	Juin	Juil.	Août	Sept.	Oct.	Nov.	Déc.
1970		0,31	0,31	0,51	0,40	0,40	0,20	0,40	0,40	0,40	0,40	0,30
1971	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	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	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	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	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	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	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	0,48	0,74	0,89	1,09	0,97	0,76	1,21	0,60	0,64	0,93	0,53	0,48

Cette nouvelle série présente évidemment plus de changements de variation que la série initiale I_t ; un accroissement brusque de l'indice est souvent compensé par un accroissement plus faible un ou deux mois après, d'où les nombreux "pics et creux" de la figure 1.5.

Une fois éliminés ces "pics et creux" à très court terme, on voit apparaître une évolution à moyen terme comprenant quatre phases principales : la première, jusqu'au mois d'octobre 1973 correspond à un taux d'accroissement à peu près fixe voisin de 0,4 % par mois ; cet accroissement a alors augmenté brusquement jusqu'au milieu de l'année 1974, puis

après une décroissance s'est stabilisé à partir de 1976 à un niveau proche de 0,7 %, supérieur à celui du début de la période.

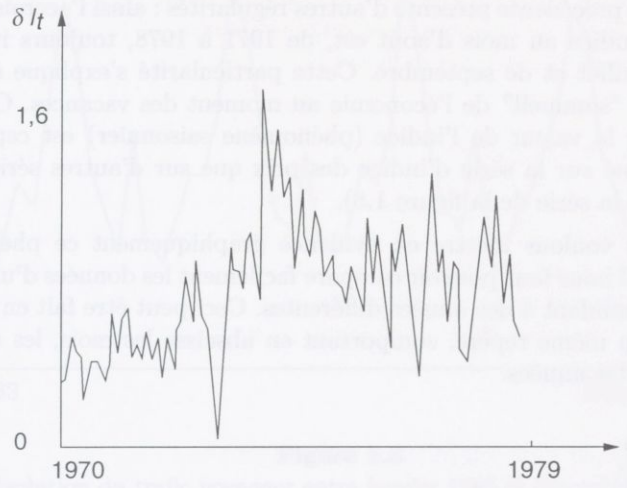

Figure 1.5

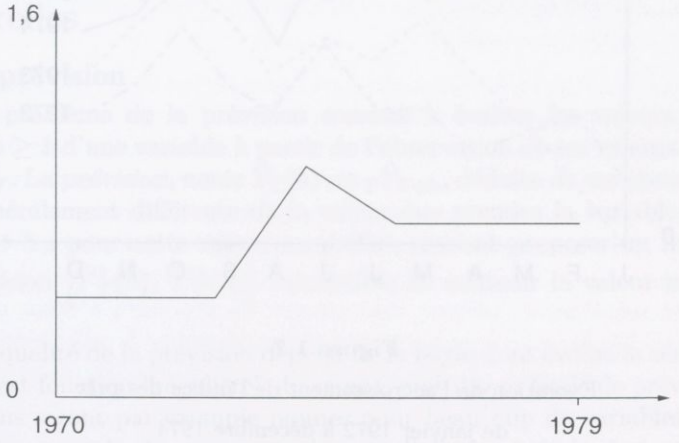

Figure 1.6
Evolution à moyen terme de l'accroissement relatif de l'indice des prix

La quasi-stabilité de δI_t sur la période 1970-1972 peut également être mise en évidence en présentant la série I_t dans un repère semi-logarithmique (ordonnée Log I_t ; abscisse t). En effet, la série vérifie la relation approchée : $\dfrac{I_t - I_{t-1}}{I_{t-1}} = \delta_0$ où δ_0 est une constante et satisfait donc ap-

proximativement : $I_t = I_1(1+\delta_0)^{t-1}$. Prenant le logarithme de cette égalité, nous voyons que dans un repère semi-logarithmique l'évolution à moyen terme sera sur la période 1970-1972 approximativement linéaire.

La série précédente présente d'autres régularités : ainsi l'accroissement relatif de l'indice au mois d'août est, de 1971 à 1978, toujours inférieur à ceux de juillet et de septembre. Cette particularité s'explique évidemment par le "sommeil" de l'économie au moment des vacances. Cet effet du mois sur la valeur de l'indice (phénomène saisonnier) est cependant moins marqué sur la série d'indice des prix que sur d'autres séries (voir par exemple la série de la figure 1.8).

Si nous voulons mettre en évidence graphiquement ce phénomène périodique, il nous faut pouvoir compare facilement les données d'un même mois correspondant à des années différentes. Ceci peut être fait en présentant dans un même repère, comportant en abscisse les mois, les données de chacune des années.

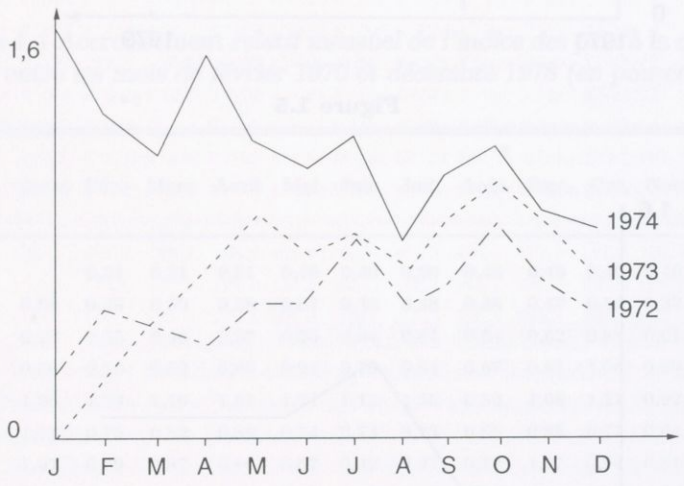

Figure 1.7

Evolution de l'accroissement de l'indice des prix

de janvier 1972 à décembre 1974

Cette représentation graphique fait mieux apparaître les similitudes de forme au cours de l'année. Ainsi entre mai et décembre les trois années présentent des sommets pour les mêmes mois.

Des remarques similaires peuvent être faites pour l'autre série. Le graphe des données relatives au trafic SNCF (figure 1.8) fait ainsi apparaître une évolution à moyen terme approximativement constante et une périodicité de 12 nettement plus marquée que dans la série précédente.

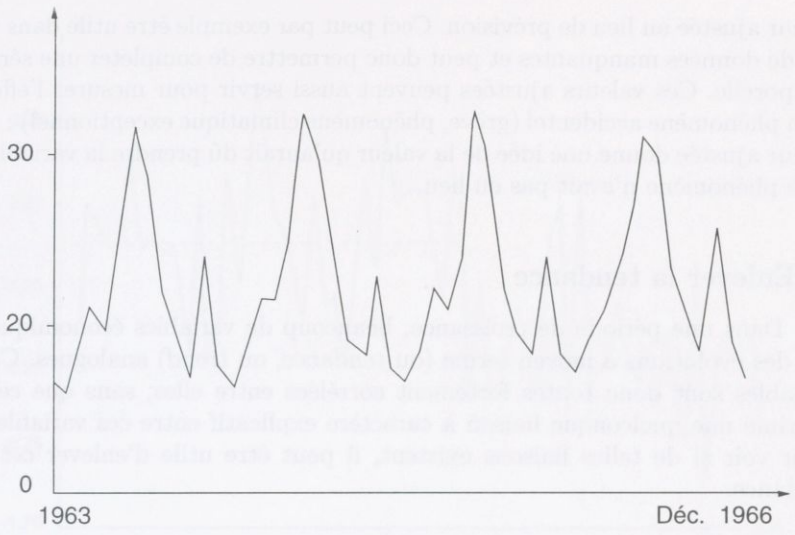

Figure 1.8
Evolution du trafic voyageur entre janvier 1963 et décembre 1966

1.4 Quelques problèmes posés par les séries temporelles

A. La prévision

Le problème de la prévision consiste à évaluer les valeurs futures Y_{T+h}, $h \geq 1$ d'une variable à partir de l'observation de ses valeurs passées $Y_1, \ldots Y_T$. La *prévision*, notée $\hat{Y}_T(h)$ ou ${}_T\hat{Y}_{T+h}$, déduite de ces observations sera généralement différente de la valeur que prendra la variable à l'instant $T+h$; pour cette raison on préfère souvent proposer un *intervalle de prévision* $\left[\hat{Y}^1_T(h),\ \hat{Y}^2_T(h)\right]$ susceptible de contenir la valeur inconnue Y_{T+h}.

La qualité de la prévision dépend de la façon dont évolue la série. Plus la série est fonction "régulière" du temps, plus il sera facile de prévoir. Les prévisions seront par exemple bonnes pour beaucoup de variables économiques en période de croissance, lorsque l'allure générale de la série est linéaire ou exponentielle. En revanche, les diverses méthodes de prévision ne permettent pas de prévoir un changement d'évolution dû à une modification des structures économiques, que rien dans le passé ne permettait de supposer.

La qualité de la prévision dépend de l'*horizon* h et est généralement meilleure, lorsque h est petit.

Les méthodes servant à la prévision peuvent également être utilisées pour évaluer une valeur passée de la variable. On emploie alors le terme de

valeur ajustée au lieu de prévision. Ceci peut par exemple être utile dans le cas de données manquantes et peut donc permettre de compléter une série temporelle. Ces valeurs ajustées peuvent aussi servir pour mesurer l'effet d'un phénomène accidentel (grève, phénomène climatique exceptionnel) ; la valeur ajustée donne une idée de la valeur qu'aurait dû prendre la variable, si ce phénomène n'avait pas eu lieu.

B. Enlever la tendance

Dans une période de croissance, beaucoup de variables économiques ont des évolutions à moyen terme (ou *tendance*, ou *trend*) analogues. Ces variables sont donc toutes fortement corrélées entre elles, sans que ceci exprime une quelconque liaison à caractère explicatif entre ces variables. Pour voir si de telles liaisons existent, il peut être utile d'enlever cette tendance.

C. Correction des variations saisonnières

Nous avons observé sur la sértie du trafic SNCF que les valeurs prises pour un mois donné étaient généralement du même côté de la moyenne annuelle ; par exemple les valeurs associées au mois d'août sont au-dessus de celle-ci. On appelle série corrigée des *variations saisonnières* ou *série désaisonnalisée* notée Y_{CVS}, la série obtenue en retranchant de la série initiale cet effet saisonnier.

Cette série corrigée des variations saisonnières est d'un grand intérêt pour l'interprétation des séries temporelles. Supposons, par exemple, qu'une mesure économique ait été prise au mois de juillet afin de réduire l'augmentation des prix. L'observation d'un taux d'accroissement des prix plus faible en août qu'en juillet permet-il d'en déduire un effet de cette mesure ? Il faut se garder de conclure trop rapidement ; l'observation des valeurs prises par la série corrigée des variations saisonnières, où l'effet de la saison est éliminé, permet directement de répondre à cette question. Signalons d'autre part que pour une raison similaire à celle décrite en B, il peut être nécessaire de désaisonnaliser les séries avant de chercher les liaisons explicatives pouvant exister entre variables.

D. Détection de rupture

Suite à des changements de politique économique ou à des modifications profondes des relations structurelles entre variables, les séries peuvent dans certains cas présenter des ruptures soit de niveau, soit de pente. Nous donnons ci-dessous deux séries. La première relative au taux d'accroissement du PNB espagnol montre une baisse brusque du niveau correspondant aux "crises pétrolières" successives.

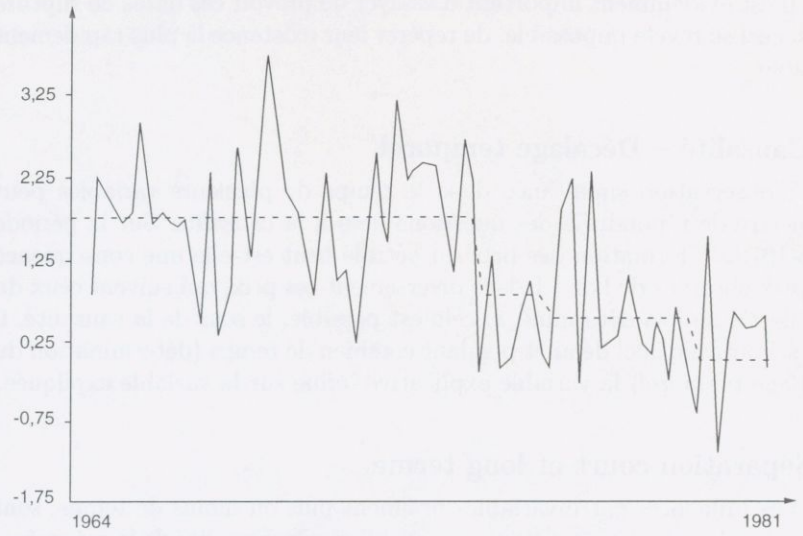

Figure 1.9
PNB espagnol (accroissement)

La seconde fournit les évolutions de taux de change de la livre par rapport au dollar. On y observe diverses ruptures de pente.

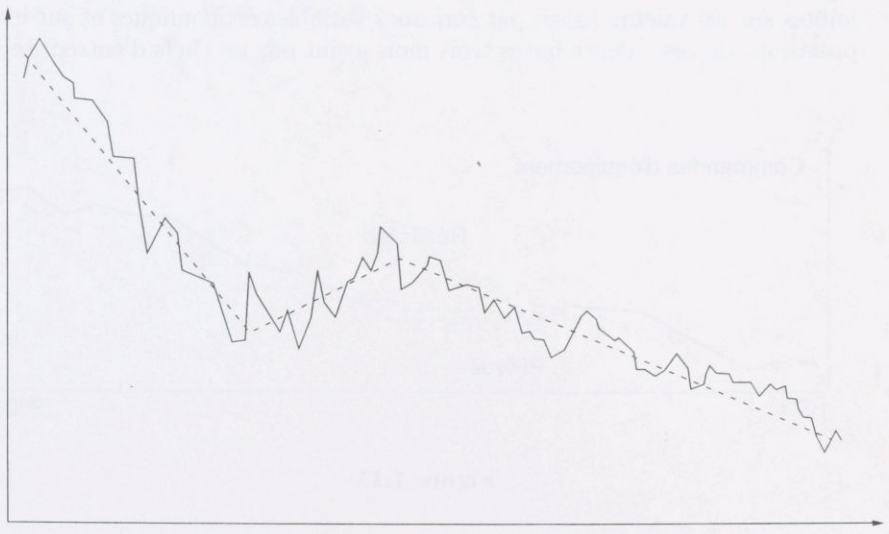

Figure 1.10
Taux de change livre/dollar

Il est évidemment important d'essayer de prévoir ces dates de rupture ou, si ceci se révèle impossible, de repérer leur existence le plus rapidement possible.

E. Causalité – Décalage temporel

L'observation simultanée dans le temps de plusieurs variables peut permettre de répondre à des questions liées à la *causalité*. Sur la période 1975-1979, la formation des prix du pétrole brut est-elle une conséquence des prix observés de l'or ? Est-ce inversement ces prix, qui suivent ceux du pétrole ? Une fois déterminé, si cela est possible, le sens de la causalité, il faut savoir avec quel délai et pendant combien de temps (détermination du *décalage temporel*) la variable explicative influe sur la variable expliquée.

F. Séparation court et long terme

Les influences entre variables prennent plus ou moins de temps, sont plus ou moins persistantes. L'un des problèmes importants de la macroéconométrie est de séparer ces relations persistantes (dites de long terme) de celles qui ne le sont pas. Ces dernières s'interprètent souvent en terme d'ajustement.

G. Etude des anticipations des agents

L'analyse des séries temporelles permet aussi d'étudier comment les agents réagissent vis-à-vis du temps. Ainsi on dispose souvent de données jointes sur les valeurs prises par certaines variables économiques et sur les prévisions de ces valeurs faites trois mois avant par les chefs d'entreprise.

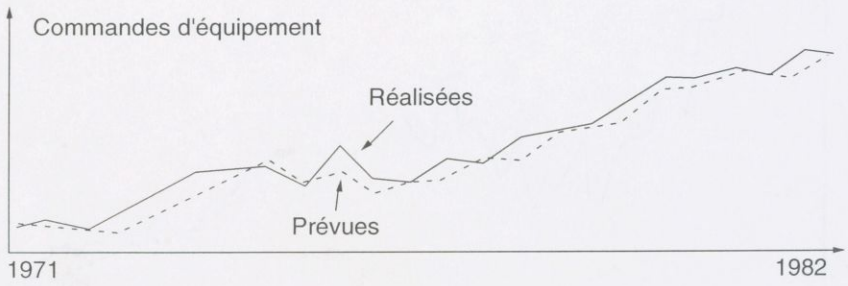

Figure 1.11

On peut alors se demander s'ils anticipent bien le niveau de la variable, son évolution ; on peut essayer de comprendre comment ils calculent implicitement leurs anticipations...

1.5 Spécificité des méthodes de traitement

Diverses méthodes ont été mises au point pour résoudre les problèmes précédents.

Le caractère optimal de chacune de ces méthodes dépend du problème considéré : désaisonnalisation ou prévision par exemple, et également de la série dont on dispose. Pour illustrer ce fait, nous pouvons considérer trois critères proposés par Lovell (1963) que devrait raisonnablement satisfaire une méthode de correction des variations saisonnières :

$$X_{CVS} + Y_{CVS} = (X+Y)_{CVS} \quad \forall X,\ Y \tag{I}$$

$$\lambda X_{CVS} = (\lambda X)_{CVS} \quad \forall X,\ \forall \lambda \in R^{+} \tag{II}$$

$$X_{CVS}\, Y_{CVS} = (XY)_{CVS} \quad \forall X,\ Y \tag{III}$$

La première de ces propriétés serait très importante en pratique, puisqu'elle permettrait d'obtenir la série agrégée corrigée de variations saisonnières à partir des séries de base désaisonnalisées. Elle permettrait également d'obtenir la série désaisonnalisée d'un niveau à partir de la série désaisonnalisée du flux associé et inversement.

La seconde propriété exprime l'indépendance de la méthode de désaisonnalisation vis-à-vis des changements d'unités : désaisonnaliser une série de consommation de pétrole exprimée en m^3 ou exprimée en barils conduira au même résultat.

La troisième propriété permettrait par exemple d'obtenir le taux de chômage désaisonnalisé en faisant le rapport du nombre ce chômeurs désaisonnalisé par le nombre d'actifs désaisonnalisé. Ces trois propriétés, qui a priori semblent naturelles, conduisent cependant à des méthodes inutilisables en pratique.

Les deux premières conditions impliquent la linéarité du procédé d'ajustement saisonnier ; il existe donc des réels $a_{t\tau}$ tels que : $X_{CVS,\ t} = \sum_{\tau} a_{t\tau} X_{\tau}$.

Si nous écrivons maintenant la condition (III) :

$$\left[\sum_{\tau} a_{t\tau} X_{\tau}\right]\left[\sum_{\tau} a_{t\tau} Y_{\tau}\right] = \sum_{\tau} a_{t\tau} X_{\tau} Y_{\tau}, \quad \forall X_{\tau},\ Y_{\tau},$$

nous en déduisons qu'il existe une valeur t_0 telle que :

$$\forall \tau \neq t_0,\ a_{t\tau} = 0,$$

et

$$a_{tt_0} = 0 \text{ ou } 1.$$

La valeur désaisonnalisée serait égale soit à une valeur décalée, soit à zéro, ce qui est peu intéressant.

Bien que chaque méthode ne puisse posséder de bonnes propriétés pour toutes les séries existantes, il sera toujours possible de trouver des méthodes satisfaisantes pour chaque série étudiée. Généralement, les méthodes usuelles sont déduites d'un modèle décrivant la série chronologique ; elles posséderont certaines propriétés d'"optimalité" pour les séries compatibles avec le modèle. Plus la série s'écartera du modèle supposé, plus la méthode correspondante sera mauvaise et les résultats entachés de distorsions graves.

1.6 Modélisation d'une série temporelle

La résolution des divers problèmes que nous avons évoqués s'appuie sur des modèles décrivant la façon dont la série évolue. Il est commode de distinguer trois types de modèles : les modèles d'ajustement, les modèles autoprojectifs et les modèles explicatifs. Nous allons présenter brièvement ces trois types de modèles et nous indiquerons au passage comment ils seront utilisés dans la suite du livre.

A. Modèles d'ajustement

1. Le principe

L'observation de données réelles fait généralement apparaître diverses régularités ; si nous considérons par exemple la série du trafic voyageur entre janvier 1963 et décembre 1966 (figure 1.8), nous constatons que le mouvement à moyen terme (tendance) est approximativement constant ; si nous retranchons cette tendance de la série initiale nous obtenons une courbe approximativement périodique et nulle en moyenne. On peut donc penser faire l'hypothèse que la série a été générée par un modèle du type :

$$Y_t = a + S_t + u_t,$$

où a est une constante inconnue représentant la tendance et S_t est une fonction périodique du temps (par exemple sinusoïdale) de période 12, nulle en moyenne, qu'on appellera *mouvement saisonnier* ; u_t, appelé *partie irrégulière*, aura le statut de variable aléatoire centrée. Notons que cette partie irrégulière est souvent petite par rapport aux deux autres, ce qui ne signifie pas qu'elle est négligeable ; au contraire, c'est souvent dans cette partie que l'on trouvera les fluctuations les plus intéressantes sur le plan économique.

La décomposition précédente est classique ; on y ajoute quelquefois une quatrième partie, appelée le *cycle*, représentant des mouvements périodiques à moyen terme.

D'une manière plus générale, on peut proposer un modèle du type :

$$y_t = f(t,\ u_t),$$

où f est une fonction indexée par un nombre fini de paramètres inconnus et où u_t est une variable aléatoire centrée sur laquelle diverses hypothèses

additionnelles peuvent être faites ; un tel modèle est appelé *modèle d'ajustement.*

2. Ajustement global et ajustement local

Les hypothèses faites sur les variables aléatoires u_t induisent des méthodes d'estimation de la fonction f, ayant de bonnes propriétés d'optimalité.

Ainsi, dans le chapitre II on propose une méthode d'estimation optimale dans le cas où les u_t ont mêmes variances et sont non corrélées : c'est la méthode des moindres carrés ordinaires. On propose également une méthode dans le cas où les u_t ont même variance et sont corrélés. Ces méthodes font jouer un rôle équivalent à toutes les observations et on peut dire qu'elles conduisent à des *ajustement globaux.* Les résultats des estimations obtenues peuvent être utilisés de diverses façons et, en particulier, pour la désaisonnalisation et la prévision.

Dans certains cas on peut vouloir fonder les modèles et les méthodes d'estimation sur des critères permettant de faire jouer des rôles différents aux diverses observations. Ainsi dans le chapitre 3 sur les moyennes mobiles, on définit des séries qui sont "*localement*" assimilables à des polynômes de bas degrés, c'est-à-dire qui sont invariantes par un ajustement polynomial local faisant intervenir un petit nombre d'observations. Cette définition permet d'introduire une classe importante de méthodes de désaisonnalisation.

Dans le chapitre 4 consacré aux méthodes de prévision par lissage on verra également que ces méthodes peuvent être justifiées par un *ajustement local* de diverses fonctions (polynômes, fonctions périodiques, exponentielles...) autour de la valeur présente ; dans ces ajustements les observations jouent un rôle exponentiellement décroissant quand on remonte dans le passé, d'où le nom de lissage exponentiel donné à ces méthodes.

3. Ajustement additif ou multiplicatif

La fonction d'ajustement $f(t,\ u_t)$ est souvent additive, c'est-à-dire qu'elle se décompose en :

$$f(t,\ u_t) = g(t) + u_t,$$

on dit que le modèle est un modèle d'ajustement *additif.*

Si la fonction $f(t,\ u_t)$ s'écrit :

$$f(t,\ u_t) = g(t) \cdot u_t,$$

on dit que le modèle est un modèle d'ajustement *multiplicatif.*

Notons que, lorsque les diverses variables sont positives, on passe d'un modèle multiplicatif à un modèle additif à l'aide de la fonction logarithme.

4. Ajustement déterministe et ajustement aléatoire

On a vu qu'un modèle d'ajustement additif s'écrit :

$$Y_t = g(t) + u_t,$$

où $g(t)$ est une fonction du temps inobservable, mais déterministe.

Il est également possible de supposer que $g(t)$, $t \in \mathbb{Z}$, est un processus aléatoire indépendant du processus u_t, $t \in \mathbb{Z}$. Une classe particulière de modèles de ce type, appelés *modèles à composantes non observables*, est étudiée dans le chapitre 15.

B. Modèles autoprojectifs

Dans un modèle autoprojectif, on suppose que Y est une fonction de ses valeurs passées et d'une perturbation aléatoire u_t :

$$Y_t = f(Y_{t-1},\ Y_{t-2}, ...,\ u_t).$$

Une classe de tels modèles, particulièrement utiles pour la prévision, sont les modèles ARIMA étudiés à partir du chapitre V, d'abord dans le cas univarié, puis dans le cas multivarié. Comme on le verra, cette classe permet d'atteindre une gamme de modèles très variés à l'aide d'un nombre de paramètres relativement limité. En outre, il est possible de proposer des méthodes, dites méthodes d'identification, permettant de choisir dans cet ensemble de modèles celui qui semble le mieux adapté aux données dont on dispose. Une fois ce modèle choisi, on peut en estimer les paramètres et déterminer les prévisions optimales.

Parmi les méthodes de prévision ainsi mises en évidence on trouve, en particulier, les méthodes de lissage exponentiel. L'avantage essentiel de la méthode identification – estimation – prévision fondée sur les modèles ARIMA, appelée quelquefois méthode de Box et Jenkins, est qu'elle permet de sélectionner la méthode de prévision optimale dans un vaste ensemble de possibilités, tandis que les méthodes classiques (ajustement global ou lissage) comportent une plus grande part d'arbitraire dans le choix de la fonction utilisée pour l'ajustement.

De plus, on verra que dans le cas multivarié ces modèles permettent de donner des réponses aux problèmes liés à la causalité, à la séparation court et long terme, aux comportements d'anticipation des agents (chapitres 10 à 13).

C. Modèles explicatifs

Dans cette dernière catégorie de modèles la variable Y_t est exprimée en fonction d'un vecteur de variables observables, dites exogènes, X_t et d'une perturbation aléatoire u_t :

$$Y_t = f(X_t,\ u_t),$$

X_t est soit déterministe, soit aléatoire ; dans ce dernier cas, les processus $\{X_t\}$ et $\{u_t\}$ ont certaines propriétés d'indépendance ou de non corrélation.

Ces modèles sont les modèles de base de l'économétrie et nous les considérerons essentiellement pour faire le lien entre eux et les modèles autoprojectifs.

1. Modèle explicatif statique

Dans un modèle explicatif statique les variables X_t ne contiennent pas de valeurs passées de Y_t et les u_t sont indépendants entre eux. Par exemple, un modèle de ce type sera :

$$Y_t = a + bX_t + u_t, \ t = 1, ..., \ T,$$

où les X_t sont indépendants de tous les u_t et où les u_t sont centrés et indépendants entre eux.

2. Modèle explicatif dynamique

Un modèle explicatif peut être dynamique soit parce que les u_t sont autocorrélés, soit parce que X_t contient des valeurs passées de Y_t, c'est-à-dire des variables dites "endogènes retardées".

– Perturbations autocorrélées

Une façon commode de prendre en compte l'autocorrélation des u_t est de supposer que la série des u_t satisfait un modèle autoprojectif. L'approche autoprojective permet donc de proposer une classe de modèles pour les perturbations, c'est-à-dire pour la composante représentant notre ignorance, et apparaît complémentaire de la formalisation reliant Y et X qui, elle, est souvent fondée sur la connaissance des mécanismes économiques.

– Variables endogènes retardées

La théorie économique fournit souvent des indications sur l'identité des variables à faire intervenir dans un modèle ; en revanche elle donne rarement des précisions sur les décalages temporels pertinents. Dans l'étude de ce problème difficile et crucial l'approche autoprojective peut aussi, comme on le verra dans le chapitre 10, être utile.

Ces modèles explicatifs dynamiques seront étudiés en détail dans la suite du livre. On distinguera en particulier les modèles stationnaires et non stationnaires (chapitres 11, 13, 14), on s'intéressera aux modèles dynamiques fondés sur la théorie des anticipations (chapitre 12) et on fera le lien entre les modèles dynamiques et la théorie du filtrage de Kalman (chapitres 15 et 16).

CHAPITRE 2

Désaisonnalisation par la méthode de la régression linéaire

2.1 Ecriture d'un modèle linéaire

On suppose que la série brute est la somme des deux composantes déterministes (la tendance Z_t et la saisonnalité S_t) et d'une composante aléatoire (la perturbation u_t) :

$$X_t = Z_t + S_t + u_t.$$

Pour compléter la description du modèle, il reste à expliciter les formes fonctionnelles des composantes déterministes et à faire des hypothèses sur la loi des perturbations u_t, $t = 1, ... T$. Chacune des composantes déterministes est écrite comme combinaison linéaire de fonctions connues du temps :

$$Z_t = Z_t^1 b_1 + Z_t^2 b_2 + ... + Z_t^k b_k = \sum_{i=1}^{k} Z_t^i b_i,$$

$$S_t = S_t^1 c_1 + S_t^2 c_2 + ... + S_t^l c_l = \sum_{j=1}^{l} S_t^j c_j \ ;$$

$b_1 ... b_k$, $c_1 ... c_l$ sont des constantes inconnues qu'il faudra estimer à partir des observations. Les perturbations u_t sont choisies de moyennes nulles $Eu_t = 0 \ \forall t$, de même variance : $Vu_t = \sigma^2 \ \forall t$, et non corrélées $\text{Cov}(u_t, \ u_\tau) = 0, \ \forall t \neq \tau$.

Finalement, le modèle est :

$$\boxed{\begin{array}{c} X_t = \sum_{i=1}^{k} Z_t^i \ b_i + \sum_{j=1}^{l} S_t^j \ c_j + u_t \quad t = 1, \cdots, \ T \\ \text{avec } Eu_t = 0, Vu_t = \sigma^2, \ \text{Cov} \ (u_t, \ u_\tau) = 0 \text{ si } t \neq \tau \end{array}} \tag{2.1}$$

La partie déterministe apparaît comme une fonction linéaire des $k+l$ paramètres b et c, d'où le nom de modèle linéaire donné à la formule (2.1).

Ce modèle peut être écrit en terme de vecteurs. Appelons X, $Z^1, ..., Z^k$, $S^1, ..., S^l$, u les vecteurs de $\mathbb{R}^T$, dont les composants sont respectivement X_t, $Z_t^1, ..., Z_t^k$, $S_t^1, ..., S_t^l$, u_t.

Le système d'équations (2.1) signifie que le vecteur aléatoire X admet la décomposition :

$$\boxed{\begin{array}{c} X = \sum_{i=1}^{k} Z^i b_i + \sum_{j=1}^{l} S^j c_j + u, \\ Eu = 0,\ Vu = \sigma^2 I\ d. \end{array}}$$

La tendance $\sum_{i=1}^{k} Z^i b_i$ est un élément du sous-espace vectoriel $\mathcal{Z}$ engendré par les vecteurs $Z^1, ..., Z^k$. Lorsque ces k vecteurs sont indépendants, ce sous-espace est de dimension k. De même, la saisonnalité $\sum_{j=1}^{l} S^j c_j$ est un élément du sous-espace $\mathcal{S}$ engendré par $S^1, ..., S^l$. Ce sous-espace est de dimension l dès que les vecteurs $S^1, ..., S^l$ constituent un système libre. Ces conditions sur les dimensions de $\mathcal{Z}$ et $\mathcal{S}$ sont supposées satisfaites dans la suite.

Les fonctions Z^i et S^j dépendent de la façon dont le caractère X varie avec le temps. Considérons, par exemple la série trimestrielle d'un indice de production industrielle (voir annexe 2).

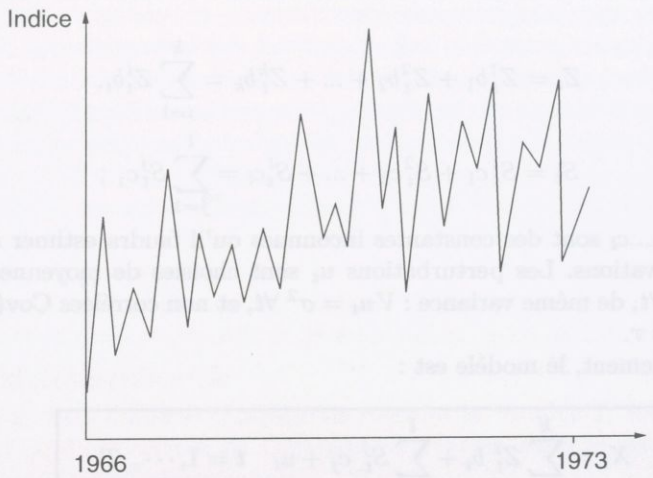

Figure 2.2

Sur la période considérée la croissance à moyen terme est approximativement linéaire ; il est alors naturel d'approcher la tendance par une droite : $Z_t = b_1 + b_2 t$. Le sous-espace $\mathcal{Z}$ de dimension 2 est engendré par les séries $Z_t^1 = 1$ et $Z_t^2 = t$. Une fois retranché de la série brute le mouvement à moyen terme, la courbe obtenue (figure 2.3) présente des pics et des creux d'amplitudes comparables à intervalle régulier. Cet intervalle de quatre observations correspond ici à l'année. La composante saisonnière S_t pourra être choisie sous la forme d'une fonction périodique de période quatre. Elle prendra quatre valeurs (ou coefficients saisonniers) c_1, c_2, c_3, c_4 associées à chacun des trimestres : $S_t = c_j$, si la date t correspond au trimestre j.

On peut alors écrire : $S_t = S_t^1 c_1 + S_t^2 c_2 + S_t^3 c_3 + S_t^4 c_4$ où S_t^j, fonction indicatrice du trimestre j, prend la valeur 1 si la date t correspond à ce trimestre, la valeur 0 sinon.

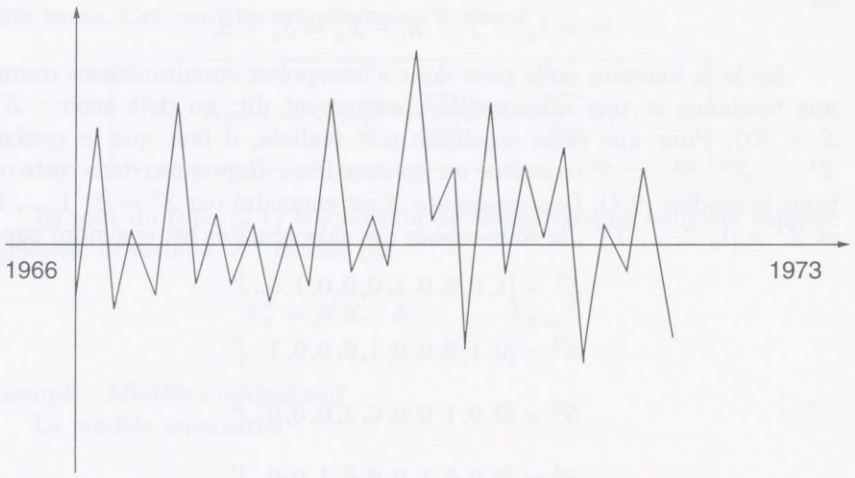

Figure 2.3

En résumé on peut, pour décrire cette série particulière, proposer en première approximation un modèle du type :

$$\boxed{X_t = b_1 + b_2 t + S_t^1 c_1 + S_t^2 c_2 + S_t^3 c_3 + S_t^4 c_4 + u_t} \qquad (2.4)$$

Généralement, on choisit pour la tendance une fonction de forme lisse, présentant peu de changements de variations ; souvent ce sera un polynôme de degré faible. La modélisation la plus simple de la saisonnalité consiste, comme dans l'exemple précédent, à prendre un effet fixe de la saison. D'autres formes peuvent être proposées : une variation lente de chacun des coefficients saisonniers est décrite en prenant ceux-ci fonctions polynomiales du temps ; les coefficients des polynômes sont alors supposés différents pour chaque saison (voir exercice 2.6).

La décomposition de la saisonnalité peut d'autre part inclure des variables prenant des valeurs non nulles pour plus d'une saison. Par exemple, l'une de ces variables pourrait être le nombre de jours de travail dans la saison considérée. L'effet des vacances, des grèves, des week-ends pourrait également être représenté de cette façon.

2.2 Unicité de la décomposition

Les formes retenues pour les composantes déterministes doivent conduire à une décomposition unique de la série initiale en tendance et saisonnalité. Il ne doit pas exister deux couples distincts $[Z,\ S]$ et $[Z^*,\ S^*]$ tels que :

$$\forall t = 1, ...,\ T, \quad Z_t + S_t = Z_t^* + S_t^*,$$

ou

$$\forall t = 1, ...,\ T, \quad Z_t - Z_t^* = S_t^* - S_t.$$

Seule la fonction nulle peut donc s'interpréter simultanément comme une tendance et une saisonnalité ; autrement dit, on doit avoir : $\mathcal{Z} \cap \mathcal{S} = \{0\}$. Pour que cette condition soit réalisée, il faut que le système $Z^1, ...,\ Z^k,\ S^1, ...,\ S^l$ constitue un système libre. Reprenons dans cette optique le modèle (2.4). Le sous-espace $\mathcal{Z}$ est engendré par $Z^1 = [1,\ 1, ...,\ 1]'$ et $Z^2 = [1,\ 2, ...,\ T]'$; le sous-espace des saisonnalités est engendré par :

$$S^1 = [1,0,0,0,1,0,0,0,1,0...]'$$

$$S^2 = [0,1,0,0,0,1,0,0,0,1...]'$$

$$S^3 = [0,0,1,0,0,0,1,0,0,0...]'$$

$$S^4 = [0,0,0,1,0,0,0,1,0,0...]'$$

L'unicité de la décomposition n'est pas satisfaite. Les six vecteurs $Z^1,\ Z^2,\ S^1,\ S^2,\ S^3,\ S^4$ sont en effet liés entre eux par la relation : $Z^1 = S^1 + S^2 + S^3 + S^4$.

Pour rétablir l'unicité de la décomposition, il semble naturel d'imposer aux effets saisonniers de se compenser sur une année, ce qui conduit à introduire la contrainte :

$$\boxed{c_1 + c_2 + c_3 + c_4 = 0.} \tag{2.5}$$

Le sous-espace des saisonnalités $\mathcal{S}^*$ contient alors les fonctions $\mathcal{S}$ telles que :

$$S_t = S_t^1 c_1 + S_t^2 c_2 + S_t^3 c_3 + S_t^4 c_4, \text{ avec } c_1 + c_2 + c_3 + c_4 = 0.$$

Remplaçons c_4 par $-c_1 - c_2 - c_3$, on obtient :

$$S_t = \left(S_t^1 - S_t^4\right) c_1 + \left(S_t^2 - S_t^4\right) c_2 + \left(S_t^3 - S_t^4\right) c_3.$$

Le sous-espace $\mathcal{S}^*$, de dimension 3, est engendré par :

$$S^{1*} = S^1 - S^4,\ S^{2*} = S^2 - S^4,\ S^{3*} = S^3 - S^4.$$

Le modèle composé de (2.4) – (2.5) est appelé *modèle trimestriel de Buys-Ballot.*

2.3 Transformations de la série brute

L'intérêt principal d'une description linéaire de la série réside dans la simplicité du modèle ; ceci permettra d'en faire une étude mathématique approfondie. Choisir une telle écriture peut sembler restrictif, cependant l'hypothèse de linéarité est moins forte qu'il n'y paraît. De nombreux modèles peuvent être rendus linéaires après transformation de la série brute. Ces modèles transformés s'écrivent :

$$\boxed{f(X_t,\ X_{t-1},...,\ X_{t-p}) = \sum_{i=1}^{k} Z_t^i b_i + \sum_{j=1}^{l} S_t^j c_j + u_t} \qquad (2.6)$$

Ils sont du type (2.1) à condition de choisir comme nouvelle variable expliquée la variable X^* définie par :

$$X_t^* = f(X_t,\ X_{t-1},...,\ X_{t-p}).$$

Exemple : Modèle multiplicatif

Le modèle trimestriel :

$$\text{Log } X_t = b_1 + b_2 t + S_t^1 c_1 + S_t^2 c_2 + S_t^3 c_3 + S_t^4 c_4 + u_t,\quad X_t > 0,\quad \forall t,$$

où S_t^j est la fonction indicatrice du trimestre et où les coefficients saisonniers satisfont $c_1 + c_2 + c_3 + c_4 = 0$ est du type (2.6). La série transformée de la série brute est $X_t^* = \text{Log } X_t$. Il s'agit d'un modèle multiplicatif et chacune des composantes expliquant la série initiale X peut être mise en évidence en appliquant à l'équation la fonction exponentielle. On a : $X_t = Z_t^* S_t^* u_t^*$ avec :

$$Z_t^* = \exp(b_1 + b_2 t),\quad S_t^* = \exp\left(S_t^1 c_1 + S_t^2 c_2 + S_t^3 c_3 + S_t^4 c_4\right),$$

et

$$u_t^* = \exp\,(u_t).$$

L'évolution à moyen terme a une forme exponentielle croissante, si $b_2 > 0$, décroissante sinon ; si $b_2 > 0$, les effets saisonniers croissent en valeur absolue avec t (voir figure 2.7).

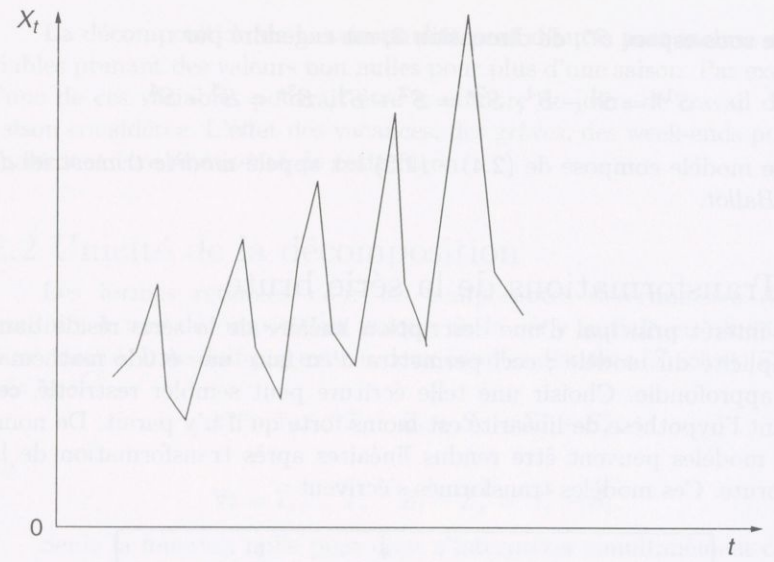

Figure 2.7
Modèle multiplicatif ($b_2 > 0$)

Exemple : Modèle logistique

Considérons le modèle donné par :

$$\text{Log}\,\frac{1-X_t}{X_t} = b_1 + b_2 t + \sum_{j=1}^{4} S_t^j c_j + u_t.$$

Pour que ce modèle ait un sens, il faut que X_t soit compris entre 0 et 1, quel que soit t.

La variable X_t s'écrit :

$$X_t = \frac{1}{1 + \exp\,(b_1 + b_2 t + \sum_{j=1}^{4} S_t^j c_j + u_t)},$$

et son évolution a approximativement la forme indiquée dans la figure 2.8, lorsque b_2 est négatif.

Ce modèle, dit logistique, est adapté à l'étude de la consommation de certains biens durables. X_t représente alors la proportion de ménages possédant ce bien à l'instant t.

2.4 Estimateur des moindres carrés ordinaires (m.c.o.)

Considérons le modèle linéaire (2.1)

$$X_t = \sum_{i=1}^{k} Z_t^i b_i + \sum_{j=1}^{l} S_t^j c_j + u_t,\ t = 1, ..., T.$$

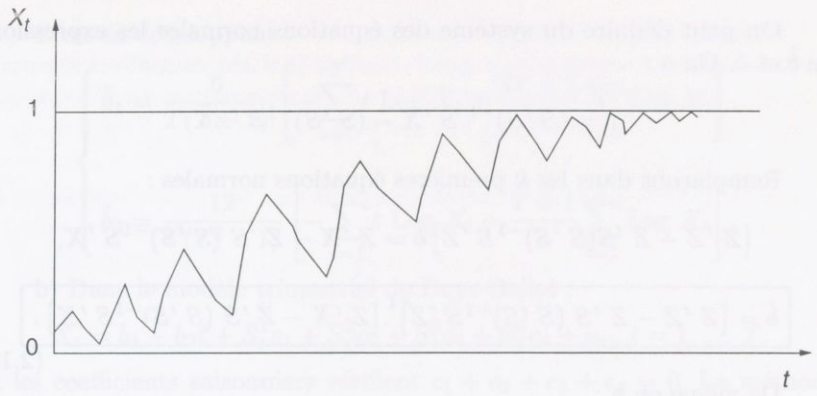

Figure 2.8
Modèle logistique

La méthode des m.c.o. consiste à choisir comme estimations $\hat{b}_i$, $\hat{c}_j$ des paramètres b_i, c_j, les valeurs qui rendent minimal le carré de la distance entre la série et sa partie déterministe :

$$\sum_{t=1}^{T} u_t^2 = \sum_{t=1}^{T} \left[X_t - \sum_{i=1}^{k} Z_t^i b_i - \sum_{j=1}^{l} S_t^j c_j \right]^2 .$$

Notons $\hat{b} = \left[\hat{b}_1, ..., \hat{b}_k\right]'$, $\hat{c} = [\hat{c}_1, ..., \hat{c}_l]'$ les estimations des vecteurs des paramètres et Z, S les matrices de tailles $(T,\ k)$ et $(T,\ l)$ dont les colonnes sont respectivement $Z^1, ...\ Z^k$ et $S^1, ..., S^l$. Les solutions $\hat{b}$ et $\hat{c}$ du problème de minimisation satisfont le système des $k + l$ *équations normales* :

$$\begin{cases} Z'Z\hat{b} + Z'S\hat{c} = Z'X, \\ S'Z\hat{b} + S'S\hat{c} = S'X, \end{cases}$$

(le symbole $'$ désigne la transposition des matrices).

Ce système peut encore s'écrire :

$$\begin{bmatrix} Z'Z & Z'S \\ S'Z & S'S \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} Z'X \\ S'X \end{bmatrix} .$$

Lorsque les vecteurs $Z^1, ..., Z^k$, $S^1, ..., S^l$ sont linéairement indépendants, la matrice du premier membre est inversible ; le système admet alors la solution unique :

$$\boxed{\begin{bmatrix} \hat{b} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} Z'Z & Z'S \\ S'Z & S'S \end{bmatrix}^{-1} \begin{bmatrix} Z'X \\ S'X \end{bmatrix}} \qquad (2.9)$$

On peut déduire du système des équations normales les expressions de $\hat{b}$ et $\hat{c}$. On a :

$$\hat{c} = (S'S)^{-1} S'X - (S'S)^{-1} S'Z\hat{b}.$$

Remplaçons dans les k premières équations normales :

$$\left[Z'Z - Z'S(S'S)^{-1}S'Z\right]\hat{b} = Z'X - Z'S\,(S'S)^{-1}S'X,$$

$$\boxed{\hat{b} = \left[Z'Z - Z'S\,(S'S)^{-1}S'Z\right]^{1}\left[Z'X - Z'S\,(S'S)^{-1}S'X\right].} \tag{2.10}$$

De même on a :

$$\boxed{\hat{c} = \left[S'S - S'Z\,(Z'Z)^{-1}Z'S\right]^{-1}\left[S'X - S'Z\,(Z'Z)^{-1}Z'X\right].} \tag{2.11}$$

Remarque 2.12 :

En l'absence de saisonnalité ($l = 0$), il faut supprimer dans le système des équations normales les blocs relatifs à S et $\hat{c}$; l'équation (2.9) se réduit alors à : $\hat{b} = (Z'\ Z)^{-1}Z'X$.

Remarque 2.13 :

La méthode que nous venons de décrire suppose que les vecteurs $Z^1, ..., Z^k, S^1, ..., S^l$ constituent un système libre ; elle ne s'applique donc pas directement au modèle de Buys-Ballot et nous verrons au paragraphe 2.5b comment l'adapter à ce modèle.

2.5 Applications

a) Considérons le modèle multiplicatif sans saisonnalité défini par :

$$\text{Log } X_t = b_1 + b_2 t + u_t,\ t = 1, ..., T.$$

La matrice Z a deux colonnes respectivement égales à $Z^1 = [1, ..., l]'$ et $Z^2 = [1, 2, ..., T]'$. Les estimateurs $\hat{b}_1$ et $\hat{b}_2$ sont donnés par :

$$\begin{bmatrix}\hat{b}_1\\ \hat{b}_2\end{bmatrix} = (Z'Z)^{-1}Z'X^*, \text{ avec } X^* = \begin{bmatrix}\text{Log } X_1\\ :\\ \text{Log } X_T\end{bmatrix},$$

$$\begin{bmatrix}\hat{b}_1\\ \hat{b}_2\end{bmatrix} = \begin{bmatrix} T & \dfrac{T(T+1)}{2} \\ \dfrac{T(T+1)}{2} & \dfrac{T(T+1)(2T+1)}{6}\end{bmatrix}^{-1}\begin{bmatrix}\displaystyle\sum_{t=1}^{T}\text{Log } X_t\\ \displaystyle\sum_{t=1}^{T} t\ \text{Log } X_t\end{bmatrix}.$$

D'où en développant :

$$\begin{cases} \hat{b}_1 = \dfrac{6}{T(T-1)} \left[-\displaystyle\sum_{t=1}^{T} t \text{ Log } X_t + \frac{2T+1}{3} \sum_{t=1}^{T} \text{Log } X_t \right], \\ \hat{b}_2 = \dfrac{12}{T(T^2-1)} \left[-\displaystyle\sum_{t=1}^{T} t \text{ Log } X_t + \frac{T+1}{3} \sum_{t=1}^{T} \text{Log } X_t \right]. \end{cases}$$

b) Dans le modèle trimestriel de Buys-Ballot :

$$X_t = b_1 + b_2 t + S_t^1 c_1 + S_t^2 c_2 + S_t^3 c_3 + S_t^4 c_4 + u_t, \ t = 1, ..., T,$$

et les coefficients saisonniers vérifient $c_1 + c_2 + c_3 + c_4 = 0$. La méthode des m.c.o. doit alors être appliquée en tenant compte de cette égalité. Les estimateurs $\hat{b}_1$, $\hat{b}_2$, $\hat{c}_1$, $\hat{c}_2$, $\hat{c}_3$, $\hat{c}_4$ sont obtenus comme solutions du problème de minimisation sous contrainte :

$$\begin{cases} \min\limits_{b,\ c} \displaystyle\sum_{t=1}^{T} \left[X_t - b_1 - b_2 t - \sum_{j=1}^{4} S_t^j c_j \right]^2 \\ \text{sous } c_1 + c_2 + c_3 + c_4 = 0. \end{cases}$$

Comme $\displaystyle\sum_{j=1}^{4} S_t^j = 1 \ \ \forall t$, ce problème s'écrit aussi :

$$\begin{cases} \min\limits_{b_2,\ \delta_j} \displaystyle\sum_{t=1}^{T} \left[X_t - b_2 t - \sum_{j=1}^{4} S_t^j \delta_j \right]^2 \\ \text{avec } b_1 = \dfrac{1}{4} \displaystyle\sum_{j=1}^{4} \delta_j \text{ et } c_j = \delta_j - \frac{1}{4} \sum_{j=1}^{4} \delta j, \ \forall j. \end{cases}$$

Ce dernier programme montre que les estimateurs $\hat{b}_1$, $\hat{b}_2$, $\hat{c}_j$ $j = 1, ..., 4$ peuvent être calculés en utilisant une démarche en deux étapes. On commence d'abord par chercher $\hat{b}_2$, $\hat{\delta}_j$ $j = 1, ..., 4$ en minimisant :

$$\sum_{t=1}^{T} \left[X_t - b_2 t - \sum_{j=1}^{4} S_t^j \delta_j \right]^2$$

c'est-à-dire en régressant par les m.c.o. *sans* contrainte X_t sur $Z_t^2 = t$, S_t^1, S_t^2, S_t^3, S_t^4. On obtient ensuite $\hat{b}_1$ et $\hat{c}_j$ en utilisant les égalités :

$$\boxed{\begin{cases} \hat{b}_1 = \dfrac{1}{4} \displaystyle\sum_{j=1}^{4} \hat{\delta}_j, \\ \hat{c}_j = \hat{\delta}_j - \dfrac{1}{4} \displaystyle\sum_{j=1}^{4} \hat{\delta}_j = \hat{\delta}_j - \hat{b}_1, \quad j = 1, ..., 4. \end{cases}} \tag{2.14}$$

Dans un premier temps, nous allons déterminer $\hat{b}_2$, $\hat{\delta}_j$. Afin de simplifier les calculs, nous supposons que le nombre T d'observations correspond à un nombre entier N d'années : $T = 4N$. L'observation du j^e trimestre de la n^e année est indexée par $t = 4(n-1) + j$ $n = 1, ..., N$, $j = 1, ..., 4$.

Nous désignerons par :

$\tilde{x}_n$ = la moyenne des observations de X sur les quatre trimestres de l'année n,

$\bar{x}_j$ = la moyenne des observations de X relatives au j^e trimestre,

$\bar{x}$ = la moyenne de toutes les observations de X.

Remarquons que : $\bar{x} = \dfrac{1}{N}\sum_{n=1}^{N} \tilde{x}_n = \dfrac{1}{4}\sum_{j=1}^{4} \bar{x}_j$.

Les estimateurs des divers paramètres sont : (voir exercice 2.9)

$$\boxed{\hat{\delta}_j = \bar{x}_j - (j + 2(N-1))\hat{b}_2 \quad j = 1, ..., 4} \tag{2.15}$$

$$\boxed{\hat{b}_2 = 3\frac{\displaystyle\sum_{n=1}^{N} n\tilde{x}_n - \frac{N(N+1)}{2}\bar{x}}{N(N^2-1)}} \tag{2.16}$$

Les estimateurs $\hat{b}_1$ et $\hat{c}_j$ s'obtiennent ensuite à partir des formules (2.14) et (2.15).

$$\begin{cases} \hat{b}_1 = \dfrac{1}{4}\displaystyle\sum_{j=1}^{4} \hat{\delta}_j = \bar{x} - \left[\frac{5}{2} + 2(N-1)\right]\hat{b}_2, \\ \hat{c}_j = \hat{\delta}_j - \hat{b}_1 = \bar{x}_j - \bar{x} - \hat{b}_2\left(j - \dfrac{5}{2}\right). \end{cases}$$

Remarquons que les estimateurs $\hat{b}_1$ et $\hat{b}_2$ des paramètres de la tendance peuvent être obtenus en effectuant la régression simple des moyennes annuelles $\tilde{x}_n$ sur l'indice du milieu de l'année exprimé en trimestres : $\tilde{t}_n = 4(n-1) + \dfrac{5}{2}$ (voir exercice 2.1).

2.6 Propriétés statistiques des estimateurs

Les estimateurs des paramètres obtenus en 2.4 sont des variables aléatoires approchant les vraies valeurs inconnues b et c. Il faut alors étudier leurs propriétés statistiques, en particulier avoir une idée de l'erreur commise en effectuant ces approximations.

A. Moments d'ordre 1

Sous l'hypothèse $Eu_t = 0$, les estimateurs des m.c.o. sont sans biais, c'est-à-dire égaux en moyenne aux valeurs à estimer : $\forall b_i,\ c_j$:

$$E\hat{b}_i = b_i,\ i = 1, ...,\ k, \quad \text{et} \quad E\hat{c}_j = c_j,\ j = 1, ...,\ l.$$

Il n'y a donc en moyenne, ni surestimation, ni sous-estimation des valeurs inconnues.

En particulier :

$$\hat{Z}_t = \sum_{i=1}^{k} \hat{b}_i Z_t^i \quad \text{et} \quad \hat{S}_t = \sum_{j=1}^{l} \hat{c}_j S_t^j,$$

estiment sans biais les valeurs inconnues de la tendance et de la saisonnalité à l'instant t.

B. Moment d'ordre 2

Une fonction linéaire des paramètres $d = \sum_{i=1}^{k} \beta_i b_i + \sum_{j=1}^{l} \gamma_j c_j$, où les β_i et γ_j sont des nombres connus est donc estimée sans biais par $\hat{d} = \sum_{i=1}^{k} \beta_i \hat{b}_i + \sum_{j=1}^{l} \gamma_j \hat{c}_j$.

L'erreur commise en approchant d par $\hat{d}$ peut être mesurée par l'erreur quadratique moyenne $E(\hat{d} - d)^2$. Comme $d = E\hat{d}$, cette erreur est égale à la variance de $\hat{d}$, notée $V\hat{d}$. Explicitons cette variance :

$$\begin{aligned} V\hat{d} &= V\left[\sum_{i=1}^{k} \beta_i \hat{b}_i + \sum_{j=1}^{l} \gamma_j \hat{c}_j\right] \\ &= \sum_{i=1}^{k} \beta_i^2 V\hat{b}_i + \sum_{j=1}^{l} \gamma_j^2 V\hat{c}_j + 2\sum_{i=1}^{k}\sum_{j=1}^{l} \beta_i \gamma_j \operatorname{Cov}(\hat{b}_i,\ \hat{c}_j) \\ &\quad + 2\sum_{i=1}^{k}\sum_{i'=1}^{i-1} \beta_i \beta_{i'} \operatorname{Cov}(\hat{b}_i,\ \hat{b}_{i'}) + 2\sum_{j=1}^{l}\sum_{j'=1}^{j-1} \gamma_i \gamma_{j'} \operatorname{Cov}(\hat{c}_j,\ \hat{c}_{j'}). \end{aligned}$$

Le calcul de cette variance suppose connues les variances des estimateurs et leurs covariances respectives. Ces quantités sont généralement présentées sous forme matricielle dans la matrice de variance covariance : $V\begin{bmatrix} \hat{b} \\ \hat{c} \end{bmatrix}$. Cette matrice carrée, symétrique, de taille $k + l$ est composée de quatre blocs :

$$V\begin{bmatrix} \hat{b} \\ \hat{c} \end{bmatrix} = \begin{pmatrix} V\hat{b} & \operatorname{Cov}\ (\hat{b},\ \hat{c}) \\ \operatorname{Cov}\ (\hat{b},\ \hat{c})' & V\hat{c} \end{pmatrix}.$$

$V\hat{b}$ matrice carrée de taille k, a pour terme général d'ordre $(i,\ i')$: Cov $(\hat{b}_i,\ \hat{b}_{i'})$; en particulier les termes de la diagonale ne sont autres que les variances $V\hat{b}_i$. Les autres blocs Cov $(\hat{b},\ \hat{c})$ et $V\hat{c}$ de taille $(k,\ l)$ et $(l,\ l)$, ont pour termes Cov $(\hat{b}_i,\ \hat{c}_j)$ et Cov $(\hat{c}_j,\ \hat{c}_{j'})$.

Si nous notons $\beta = [\beta_1, ...,\ \beta_k]'$ et $\gamma = [\gamma_1, ...,\ \gamma_l]'$, $V\hat{d}$ s'écrit :

$$\boxed{V\hat{d} = [\beta',\ \gamma']\, V \begin{bmatrix} \hat{b} \\ \hat{c} \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix}} \tag{2.18}$$

Lorsque les perturbations u_t sont de moyenne nulle, de même variance σ^2 et sont non corrélées, on peut facilement montrer que la matrice de variance-covariance de $\begin{bmatrix} \hat{b} \\ \hat{c} \end{bmatrix}$ est donnée par :

$$\boxed{V \begin{bmatrix} \hat{b} \\ \hat{c} \end{bmatrix} = \sigma^2 \begin{bmatrix} Z'Z & Z'S \\ S'Z & S'S \end{bmatrix}^{-1}} \tag{2.19}$$

Rappelons également que la principale justification de la méthode des moindres carrés ordinaires est le théorème de Gauss-Markov qui indique que $\hat{d}$ est l'estimateur de d de variance minimale parmi les estimateurs sans biais, fonctions linéaires des observations X_t.

Exemple 2.20 :

Reprenons le modèle de Buys-Ballot étudié en 2.5.b), on a :

$$V \begin{bmatrix} \hat{b}_2 \\ \hat{\delta} \end{bmatrix} = \sigma^2 \begin{bmatrix} Z^{2\prime}Z^2 & Z^{2\prime}\, S \\ S'Z^2 & S'S \end{bmatrix}^{-1} ;$$

en particulier, l'expression

$$V\hat{\delta} = \sigma^2 \left[S'S - S'Z^2 (Z^{2'}\ Z^2)^{-1} Z^{2'}\ S \right]^{-1}$$

permet de calculer les variances et covariances des coefficients saisonniers estimés $\hat{c}$.

C. Estimation de la variance des perturbations

La matrice obtenue en (2.19) ne peut être calculée numériquement, puisque son expression dépend du paramètre inconnu σ^2. Celui-ci peut être estimé sans biais de la manière suivante.

Soit $\hat{u}_t$ le *résidu* d'estimation à l'instant t : $\hat{u}_t = X_t - \hat{Z}_t - \hat{S}_t$; $\hat{u}_t$ est une approximation aléatoire de la perturbation u_t.

La quantité :

$$\boxed{s^2 = \frac{\sum_{t=1}^{T} \hat{u}_t^2}{T-k-l},} \tag{2.21}$$

est un estimateur sans biais de σ^2. La matrice de variance-covariance est alors estimée par :

$$\boxed{\hat{V}\begin{bmatrix}\hat{b}\\ \hat{c}\end{bmatrix} = s^2\begin{bmatrix}Z'Z & Z'S\\ S'Z & S'S\end{bmatrix}^{-1},} \tag{2.22}$$

et de même la variance d'une fonction linéaire des estimateurs :

$$\hat{d} = \sum_{i=1}^{k} \beta_i \hat{b}_i + \sum_{j=1}^{l} \gamma_i \hat{c}_j,$$

est estimée par :

$$\hat{V}\hat{d} = [\beta',\ \gamma']\,\hat{V}\begin{bmatrix}\hat{b}\\ \hat{c}\end{bmatrix}\begin{bmatrix}\beta\\ \gamma\end{bmatrix}.$$

D. Intervalle de confiance

Au lieu de proposer pour d une valeur approchée $\hat{d}$, il est souvent préférable d'en fournir un encadrement : $\hat{d}_1 < d < \hat{d}_2$. L'intervalle de confiance $\left[\hat{d}_1,\ \hat{d}_2\right]$ est aléatoire et ne contient d qu'avec une certaine probabilité. Lorsque les perturbations sont supposées indépendantes, suivant une même loi normale $N(0,\ \sigma^2)$, un tel intervalle encadrant la valeur d avec la probabilité 0.95 est donné par :

$$\boxed{\left[\hat{d}_1 = \hat{d} - \Phi\sqrt{\hat{V}\hat{d}}\ ;\ \hat{d}_2 = \hat{d} + \Phi\sqrt{\hat{V}\hat{d}}\right]} \tag{2.23}$$

où Φ est le nombre ayant la probabilité 5 % d'être dépassé en module dans une loi de Student à $T-k-l$ degrés de liberté. Si T est grand, Φ vaut approximativement 2.

2.7 Utilisations de la méthode de régression

A. Recherche de valeurs aberrantes

Certaines observations peuvent ne pas paraître compatibles avec le reste des données. Savoir si ces observations sont vraiment "aberrantes",

c'est-à-dire si elles n'ont pas été engendrées de la même façon que les autres données est un point important. Traiter comme les autres observations des données aberrantes risque, en effet, d'affecter sensiblement l'analyse.

Ces valeurs atypiques peuvent apparaître pour plusieurs raisons :

– elles peuvent être dues à des erreurs de mesure, de saisie ou de calcul. On peut dans un tel cas espérer les corriger en examinant la façon, dont les données ont été recueillies et traitées ;

– elles peuvent refléter des phénomènes accidentels (une grève par exemple). Les dates auxquelles se sont produits de tels événements peuvent facilement être retrouvées. Ainsi des valeurs atypiques apparaissent souvent dans les séries économiques françaises en 1963 (grève des mineurs), en 1968 (grève générale), en 1974 (première crise pétrolière)... Ces phénomènes étrangers à l'évolution habituelle de la série sont généralement éliminés.

On peut éventuellement considérer comme valeurs aberrantes des données engendrées de la même façon que les autres données, mais tirées dans la queue de la distribution. Ainsi dans un échantillon d'individus renouvelé chaque année pourrait à une certaine date figurer un milliardaire. Le traitement d'un tel cas n'est pas clair ; souvent on élimine ces valeurs et on effectue l'analyse sur les données obtenues après l'élimination.

Il y a diverses manières de détecter les valeurs aberrantes.

1) Jusqu'à maintenant nous avons considéré une décomposition de X_t de la forme :

$$X_t = Z_t + S_t + u_t,\ t = 1, ..., T, \tag{2.24}$$

dans laquelle ne figurent pas de valeurs aberrantes. Supposons que le modèle (2.24) décrive bien la série X. Sous l'hypothèse de normalité des perturbations, u_t est dans 95 % des cas compris entre -2σ et $+2\sigma$. Approchant toutes ces quantités, nous voyons que $\hat{u}_t$ devrait généralement être compris entre $-2s$ et $+2s$. Les dates où ces résidus se trouvent à l'extérieur de cet intervalle peuvent correspondre à des valeurs aberrantes.

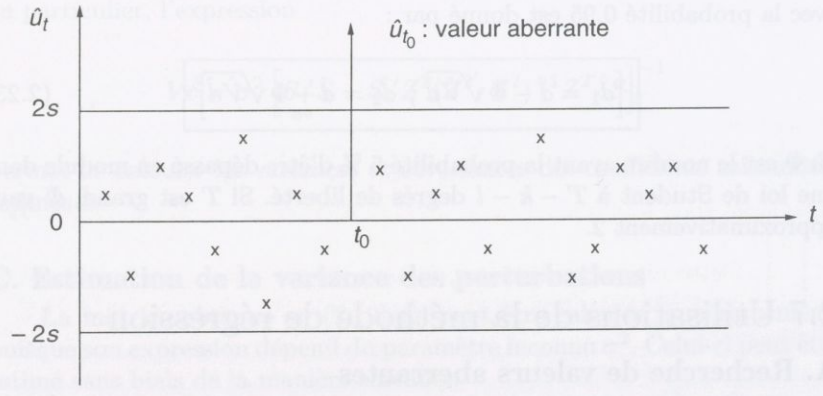

Figure 2.25
Valeurs aberrantes

Cette approche peut facilement être améliorée pour tenir compte de la loi des résidus ; désignons par A la matrice :

$$A = \left(Id - [Z,\ S] \begin{bmatrix} Z'Z & Z'S \\ S'Z & S'S \end{bmatrix}^{-1} \begin{bmatrix} Z' \\ S' \end{bmatrix} \right),$$

les résidus sont de moyennes nulles et ont pour matrice de variance-covariance : $V\hat{u} = \sigma^2 A$. Le résidu "studentisé" est donné par $\dfrac{\hat{u}_t}{s\sqrt{a_{tt}}}$, où a_{tt} et le t^e élément de la diagonale de A.

Asymptotiquement, s'il n'y a pas de valeurs aberrantes, le résidu studentisé suit une loi normale centrée réduite et est, dans approximativement 95 % des cas, inférieur à 2 en module ; on peut donc accepter l'hypothèse d'existence d'une valeur atypique à la date t si : $\left| \dfrac{\hat{u}_t}{s\sqrt{a_{tt}}} \right| > 2$ et la rejeter dans le cas contraire.

2) Une autre approche consiste à utiliser un modèle intégrant l'éventualité d'une valeur aberrante à l'instant t_0. Pour cela on introduit habituellement une variable indicatrice de cette date :

$$X_t = Z_t + S_t + \alpha \mathbb{1}_{t=t_0} + u_t. \tag{2.26}$$

où $\mathbb{1}_{t=t_0}$ vaut 0 si $t \neq t_0$, 1 sinon.

L'absence de valeur aberrante se traduit par : $H_0(\alpha = 0)$. Nous pouvons tester cette hypothèse par un test de Student classique.

Sous l'hypothèse générale, l'estimateur des m.c.o. de α n'est autre que la valeur $X_{t_0} - \hat{Z}_{t_0} - \hat{S}_{t_0}$. On voit donc que la somme des carrés des résidus sous l'hypothèse générale n'est autre que celle obtenue en appliquant les m.c.o. au modèle (2.24) pour les données X_t, $t \neq t_0$; nous la noterons $SCR^2_{(-t_0)}$. Sous l'hypothèse nulle la somme des carrés des résidus SCR^2 est obtenue en appliquant les m.c.o. à (2.24) pour les données complètes. Le module de la statistique de Student est donc égal à :

$$\left[\frac{SCR^2 - SCR^2_{(-t_0)}}{SCR^2_{(-t_0)}/(T-k-l-1)} \right]^{1/2} = \frac{|\hat{\alpha}|}{\sqrt{\hat{V}\hat{\alpha}}}.$$

Une valeur de cette statistique supérieure à la valeur critique à 95 % de la loi de Student à $(T-k-l-1)$ degrés de liberté conduit à accepter l'existence d'une valeur aberrante à la date t_0.

B. Désaisonnalisation

La série corrigée des variations saisonnières $X_t^{CVS} = X_t - S_t$ peut être estimée par :

$$\boxed{\hat{X}_t^{CVS} = X_t - \hat{S}_t = X_t - \textstyle\sum_{j=1}^{l} S_t^j \hat{c}_j.}$$

L'erreur quadratique moyenne e_t due à l'approximation de X_t^{CVS} par $\hat{X}_t^{CVS}$ est estimée par :

$$\hat{E}\left(X_t^{CVS} - \hat{X}_t^{CVS}\right)^2 = \hat{E}\left(\hat{S}_t - S_t\right)^2 = \hat{V}\left(\sum_{j=1}^{l} S_t^j \hat{c}_j\right)$$

$$= \left[S_t^1, ..., S_t^l\right] \hat{V}\hat{c} \begin{bmatrix} S_t^1 \\ \vdots \\ \vdots \\ S_t^l \end{bmatrix}.$$

C. Prévision

Les estimateurs $\hat{b}_i$, $\hat{c}_j$ peuvent servir à prévoir une valeur de X non encore observée : X_{T+h} $h \geq 1$.

Si on suppose que le modèle est encore vrai à l'instant $T+h$, on a :

$$X_{T+h} = \sum_{i=1}^{k} Z_{T+h}^i b_i + \sum_{j=1}^{l} S_{T+h}^i c_j + u_{T+h},$$

avec $Eu_{T+h} = 0$, $Vu_{T+h} = \sigma^2$, $\text{Cov}(u_{T+h}, u_t) = 0$, $t = 1, ..., T$, et la valeur X_{T+h} peut être approchée par :

$$\boxed{\hat{X}_T(h) = \sum_{i=1}^{k} Z_{T+h}^i \hat{b}_i + \sum_{j=1}^{l} S_{T+h}^j \hat{c}_j.} \tag{2.27}$$

On peut montrer que $\hat{X}_T(h)$ est, au sens de l'erreur quadratique moyenne, la meilleure prévision fonction linéaire des valeurs observées $X_1, ..., X_T$ de la variable et satisfaisant la condition de sans biais :

$$E\left(\hat{X}_T(h) - X_{T+h}\right) = 0.$$

L'estimation de la variance de l'erreur commise en effectuant cette approximation est :

$$\hat{e}_h = \hat{E}\left(\hat{X}_T(h) - X_{T+h}\right)^2$$

$$= \hat{V}\left[\sum_{i=1}^{k} Z_{T+h}^i \hat{b}_i + \sum_{j=1}^{l} S_{T+h}^j \hat{c}_j - u_{T+h}\right]$$

$$= \hat{V}\hat{d} + s^2,$$

où $\hat{d}$ est l'estimateur de $d = \sum_{i=1}^{k} Z_{T+h}^i b_i + \sum_{j=1}^{l} S_{T+h}^j c_j$.

Lorsque les perturbations suivent des lois normales, on obtient un intervalle de prévision à 95 % pour X_{T+h} en prenant :

$$\left[\hat{X}_T(h) - \Phi\hat{e}_h^{1/2},\ \hat{X}_T(h) + \Phi\hat{e}_h^{1/2}\right],$$

où Φ est défini comme en (2.23).

2.8 Autocorrélation des erreurs

Dans les graphes précédents nous avons supposé que les perturbations étaient non corrélées. Il est cependant possible que les valeurs prises par une série temporelle à des instants successifs soient liées entre elles. Une telle liaison peut être prise en compte par l'intermédiaire des perturbations.

A. Autocorrélation d'ordre 1

Supposons par exemple que les perturbations satisfassent :

$$u_t = \rho u_{t-1} + \varepsilon_t, \tag{2.28}$$

où les variables ε_t sont indépendantes de même loi $N(0,\ \sigma^2)$ et où ρ est un réel compris entre -1 et 1. Nous supposerons pour simplifier que la relation (2.28) est valable pour $t \in \mathbb{Z}$, c'est-à-dire vraie aussi pour des valeurs non observées correspondant aux indices non compris entre 1 et T. La perturbation peut alors s'écrire (voir aussi le chapitre V pour une étude complète des modèles généraux incluant 2.28)

$$u_t = \rho u_{t-1} + \varepsilon_t = \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2 u_{t-2},$$

$$u_t = \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \ldots\ \rho^i\varepsilon_{t-i} + \ldots,$$

$$u_t = \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}.$$

On déduit de cette écriture les moments de u_t :

$$Eu_t = \sum_{i=0}^{\infty} \rho^i E(\varepsilon_{t-i}) = 0,$$

$$Vu_t = V\left(\sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}\right) = \sum_{i=0}^{\infty} \rho^{2i} V(\varepsilon_{t-i}) = \sigma^2 \sum_{i=0}^{\infty} \rho^{2i},$$

$$Vu_t = \frac{\sigma^2}{1-\rho^2},$$

et pour tout $h > 0$:

$$\text{Cov}\ (u_t,\ u_{t-h}) = \text{Cov}\ \left(\sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i},\ \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i-h}\right) = \sigma^2 \frac{\rho^h}{1-\rho^2},$$

(La justification des interversions des signes Σ et E est donnée dans le chapitre 5, annexe 5.A).

La matrice de variance-covariance des perturbations est dans ce cas non scalaire ; elle est donnée par :

$$Vu = \frac{\sigma^2}{1-\rho^2} \begin{pmatrix} 1 & \rho & \dots & \rho^{T-1} \\ \rho & 1 & & \\ \vdots & & & \rho \\ \rho^{T-1} & \dots & \rho & 1 \end{pmatrix} = \sigma^2 \Omega. \tag{2.29}$$

Considérons de nouveau le modèle linéaire :

$$X = \sum_{i=1}^{k} Z^i b_i + \sum_{j=1}^{l} S^i c_i + u, \tag{2.30}$$

où les vecteurs Z^i et S^j sont linéairement indépendants et où les perturbations satisfont (2.28) ; la méthode des moindres carrés ordinaires décrite à la section précédente fournit des estimateurs sans biais des coefficients, mais ces estimateurs ne sont pas les plus précis possible. La précision peut être améliorée en employant l'estimateur des moindres carrés généralisés défini par :

$$\begin{bmatrix} \tilde{b} \\ \tilde{c} \end{bmatrix} = \left[(Z,\ S)' \Omega^{-1} (Z,\ S)\right]^{-1} (Z,\ S)' \Omega^{-1} X.$$

La matrice Ω peut facilement être inversée ; on a :

$$\Omega^{-1} = \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 \\ -\rho & 1+\rho^2 & -\rho & & 0 \\ 0 & -\rho & 1+\rho^2 & & \\ \vdots & & & & -\rho \\ 0 & 0 & & -\rho & 1 \end{pmatrix}.$$

Cet estimateur des moindres carrés généralisés est identique à l'estimateur des moindres carrés ordinaires appliqué au modèle :

$$H\ X = \sum_{i=1}^{k} H Z^i b_i + \sum_{j=1}^{l} H S^j c_j + \nu,$$

où H est une matrice vérifiant $H \Omega H' = Id$ ou $H\ 'H = \Omega^{-1}$. On obtient une telle matrice H en prenant :

$$H = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & 0 \\ -\rho & 1 & 0 & 0 \\ 0 & -\rho & 1 & 0 \\ \vdots & \vdots & & \\ 0 & 0 & -\rho & 1 \end{pmatrix}.$$

Soit H^* la matrice obtenue en enlevant la première ligne de H. L'estimateur des moindres carrés généralisés est peu différent de l'estimateur des moindres carrés ordinaires appliqué à :

$$H^*X = \sum_{i=1}^{k} H^*Z^i b_i + \sum_{j=1}^{l} H^*S^j c_j + w,$$

c'est-à-dire à l'estimateur obtenu en régressant :

$$X_t - \rho X_{t-1} \text{ sur } Z_t^i - \rho Z_{t-1}^i,\ S_t^j - \rho S_{t-1}^j,$$

pour $t = 2, \ldots T$. Cette méthode approchée a été recommandée par Cochran Orcutt [1949].

Les deux méthodes d'estimation précédentes, méthode des moindres carrés généralisés et méthode de Cochran-Orcutt, ne peuvent être directement appliquées, car elles supposent le paramètre ρ connu ; en général ρ est inconnu et il est nécessaire d'employer une méthode en deux étapes.

Dans une première étape, on estime le paramètre ρ. Ce paramètre ρ s'interprète comme la corrélation entre u_t et u_{t-1} et il est généralement estimé par la corrélation empirique entre les résidus $\hat{u}_t$ et $\hat{u}_{t-1}$ déduits de la méthode des *moindres carrés ordinaires* :

$$\hat{\rho} = \frac{\sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^{T} \hat{u}_t^2}.$$

Dans un deuxième temps on emploie la méthode des moindres carrés généralisés ou la méthode de Cochran-Orcutt après avoir remplacé le paramètre ρ inconnu par son estimation $\hat{\rho}$.

B. Y a-t-il autocorrélation d'ordre 1 ?

Avant de décrire la série chronologique par un modèle avec autocorrélation des erreurs, il faut savoir si l'emploi d'un tel modèle, est nécessaire. Pour cela il faut pouvoir tester l'hypothèse $\rho = 0$ (absence de corrélation et donc emploi de la méthode des moindres carrés ordinaires) contre l'hypothèse contraire (corrélation et emploi de la méthode des moindres carrés généralisés).

Ce test est habituellement effectué selon la méthode proposée par Durbin et Watson [Durbin-Watson (1950) (1951)] par :

$$DW = \frac{\sum_{t=2}^{T} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{T} \hat{u}_t^2},$$

où $\hat{u}_t$ est le résidu d'estimation déduit de la méthode des moindres carrés ordinaires.

Si le nombre d'observations est grand, on a :

$$DW \# 1 + \frac{\sum_{t=2}^{T} \hat{u}_t^2}{\sum_{t=1}^{T} \hat{u}_t^2} - 2\frac{\sum_{t=2}^{T} \hat{u}_t \, \hat{u}_{t-1}}{\sum_{t=1}^{T} \hat{u}_t^2} \# 1 + 1 - 2\hat{\rho},$$

$$DW \# 2(1 - \hat{\rho}).$$

La loi de cette statistique, lorsque $\rho = 0$, dépend des valeurs des variables Z^i et S^j. Il est cependant possible d'encadrer DW par deux statistiques dont les lois ne dépendent pas de ces variables. Ceci conduit à des tests faisant intervenir deux seuils critiques $d_l < d_u$, fonctions du nombre d'observations T et du nombre de variables explicatives k' non compris la constante (voir la table statistique jointe). Ainsi dans le modèle (2.30), $k' = k + l - 1$. Lorsqu'on veut tester l'absence de corrélation contre l'hypothèse de corrélation positive, on rejette l'hypothèse de non-corrélation si $DW < d_l$, on accepte cette hypothèse si $DW > d_u$; dans le cas $d_l < DW < d_u$, on ne peut pas conclure.

La procédure précédente peut être adaptée pour tester l'hypothèse de non-corrélation contre l'hypothèse d'une corrélation négative ; il suffit de remplacer DW par $4 - DW$. Si $4 - DW < d_l$ on rejette l'hypothèse de non-corrélation ; si $4 - DW > d_u$ on accepte cette hypothèse ; on ne peut conclure si $d_l < 4 - DW < d_u$.

Un inconvénient de ce test est évidemment qu'il ne permet pas toujours de choisir entre les deux hypothèses.

Une autre méthode pour tester l'hypothèse ($\rho = 0$) contre l'alternative ($\rho \neq 0$) pourrait être fondée sur la comparaison des signes des résidus $\hat{u}_t$ et $\hat{u}_{t-1}$ $t = 2,\ 4,\ 6...$ Les diverses combinaisons de signes et le nombre de fois où elles apparaissent sont données dans la table ci-après. La somme $T_{11} + T_{12} + T_{21} + T_{22}$ est égale à la partie entière de $T/2$.

L'hypothèse de non-corrélation pourrait alors être testée en appliquant un test du khi-deux à cette table de contingence 2×2. Lorsque l'hypothèse d'indépendance des signes est acceptée, on applique la méthode des moindres carrés ordinaires ; si cette hypothèse est rejetée on utilise la méthode des moindres carrés généralisés ou la méthode de Cochran-Orcutt.

Valeurs critiques à 5 % du test de Durbin-Watson

T	$k'=1$		$k'=2$		$k'=3$		$k'=4$		$k'=5$	
	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,36	1,50	1,30	1,57	1,21	1,65	1,14	1,74	1,07	1,83
31	1,36	1,50	1,30	1,57	1,23	1,65	1,16	1,74	1,09	1,83
32	1,37	1,50	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82
33	1,38	1,51	1,32	1,58	1,23	1,65	1,19	1,73	1,13	1,81
34	1,39	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,15	1,81
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80
36	1,41	1,52	1,35	1,59	1,29	1,65	1,24	1,73	1,18	1,80
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,80
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,21	1,79
39	1,43	1,54	1,38	1,60	1,33	1,66	1,27	1,72	1,22	1,79
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
55	1,53	1,60	1,49	1,64	1,45	1,68	1,41	1,72	1,38	1,77
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,50	1,70	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,52	1,70	1,49	1,74	1,46	1,77
75	1,60	1,65	1,57	1,68	1,54	1,71	1,51	1,74	1,49	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
85	1,62	1,67	1,60	1,70	1,57	1,72	1,55	1,75	1,52	1,77
90	1,63	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78
95	1,64	1,69	1,62	1,71	1,60	1,73	1,58	1,75	1,56	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78

		Signe de $\hat{u}_t$	
		+	−
Signe de $\hat{u}_{t-1}$	+	T_{11}	T_{12}
	−	T_{21}	T_{22}

C. Prévision

Lorsque les perturbations sont autocorrélées les formules de prévision (2.27) doivent être modifiées pour tenir compte des liaisons entre les observations de deux dates différentes. Supposons par exemple que nous disposions des observations $X_1, ..., X_T$ et que nous souhaitions prévoir la variable correspondant à la date $T+1$:

$$X_{T+1} = \sum_{i=1}^{k} Z_{T+1}^{i} b_i + \sum_{j=1}^{l} S_{T+1}^{j} c_j + u_{T+1},$$

avec $u_{T+1} = \rho u_T + \varepsilon_{T+1}$, où ε_{T+1} est indépendant de $\varepsilon_1, ..., \varepsilon_T$ et suit une loi $N(0,\ \sigma^2)$.

On a :

$$X_{T+1} = \rho X_T + \sum_{i=1}^{k} \left(Z_{T+1}^{i} - \rho Z_T^{i}\right) b_i + \sum_{j=1}^{l} \left(S_{T+1}^{j} - \rho S_T^{j}\right) c_j + \varepsilon_{T+1},$$

et la prévision optimale de X_{T+1} (lorsque les paramètres sont connus) est égale à la partie prédéterminée :

$$\rho X_T + \sum_{i=1}^{k} \left(Z_{T+1}^{i} - \rho Z_T^{i}\right) b_i + \sum_{j=1}^{l} \left(S_{T+1}^{j} - \rho S_T^{j}\right) c_j.$$

Lorsque les paramètres sont inconnus, une prévision de X_{T+1} sera :

$$\begin{aligned} \hat{X}_T(1) &= \hat{\rho} X_T + \sum_{i=1}^{k} \left(Z_{T+1}^{i} - \hat{\rho} Z_T^{i}\right) \hat{b}_i + \sum_{j=1}^{l} \left(S_{T+1}^{j} - \hat{\rho} S_T^{j}\right) \hat{c}_j \\ &= \sum_{i=1}^{k} Z_{T+1}^{i} \hat{b}_i + \sum_{j=1}^{l} S_{T+1}^{j} \hat{c}_j + \hat{\rho} \hat{u}_T, \end{aligned} \tag{2.31}$$

où $\hat{u}_T = X_T - \sum_{i=1}^{k} Z_T^i \hat{b}_i - \sum_{j=1}^{l} S_T^j \hat{c}_j$.

D. Autocorrélation d'ordre supérieur

Dans le modèle linéaire, la perturbation représente notamment les effets des variables omises. Si ces variables présentent elles-mêmes une saisonnalité, celle-ci va apparaître dans l'autocorrélation.

Considérons par exemple une série trimestrielle (période 4) ; il peut alors sembler préférable d'introduire l'autocorrélation entre les perturbations par :

$$u_t = \rho u_{t-4} + \varepsilon_t, \ -1 < \rho < 1.$$

La démarche décrite dans le cas de l'autocorrélation d'ordre 1 peut facilement être généralisée.

Supposons que nous disposions de $4T$ observations. La matrice de variance-covariance de u est :

$$Vu = \sigma^2 \Omega \otimes Id,$$

où Ω est la matrice définie par (2.29), Id est la matrice identité d'ordre 4 et $\otimes$ est le produit de Kronecker des matrices [1]. Donc $(Vu)^{-1} = \frac{1}{\sigma^2}\Omega^{-1} \otimes Id$ et l'estimateur des moindres carrés généralisés est donné par :

$$\begin{bmatrix} \tilde{b} \\ \tilde{c} \end{bmatrix} = \left[(Z,\ S)'(\Omega^{-1} \otimes Id)(Z,\ S)\right]^{-1} (Z,\ S)'(\Omega^{-1} \otimes Id)X.$$

Cet estimateur peut être approché par l'estimateur des moindres carrés ordinaires obtenu en régressant $X_t - \rho X_{t-4}$ sur $Z_t - \rho Z_{t-4}$, $S_t - \rho S_{t-4}$ pour $t = 5, ..., T$.

---

[1] Si A est une matrice $(p \times q)$ d'éléments a_{ij}, si B est une matrice $(g \times h)$, le produit de Kronecker de A par B, écrit $A \otimes B$, est la matrice $(pg \times qh)$ admettant la décomposition par blocs :

$$A \otimes B = \begin{bmatrix} a_{11}B & \ldots\ldots & a_{1q}B \\ \vdots & & \vdots \\ a_{p1}B & \ldots\ldots & a_{pq}B \end{bmatrix}.$$

Lorsque A et B sont carrés et inversibles, $A \otimes B$ est carrée inversible et

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$$

Ces méthodes d'estimation supposent connue la corrélation ρ et il sera nécessaire d'estimer celle-ci dans une première étape. On pourra par exemple l'estimer par :

$$\hat{\rho} = \frac{\sum_{t=5}^{T} \hat{u}_t \hat{u}_{t-4}}{\sum_{t=4}^{T} \hat{u}_t^2}.$$

Finalement nous pouvons remarquer que le test de l'hypothèse $\rho = 0$ fondé sur les signes se généralise immédiatement à condition de comparer les signes de $\hat{u}_t$ et $\hat{u}_{t-4}$ et que la technique de prévision est identique à condition de considérer $X_{T+1} - \rho X_{T-3}$.

2.9 Deux inconvénients de la méthode des m.c.o.

A. Mise à jour

Les estimateurs $\hat{b}$, $\hat{c}$ dépendent du nombre d'observations effectuées ; dans cette section nous notons $\hat{b}^T$, $\hat{c}^T$ leurs expressions relatives à la période $1, ..., T$. Lorsqu'on dispose d'une observation supplémentaire X_{T+1}, ces estimateurs sont transformés en $\hat{b}^{T+1}$, $\hat{c}^{T+1}$. Cependant, il n'y a pas de relation très simple (exercice 2.7) entre $\left(\hat{b}^{T+1}, \hat{c}^{T+1}\right)$ et $\left(\hat{b}^T, \hat{c}^T\right)$ (voir cependant le chapitre 16 pour la définition des moindres carrés récursifs).

B. Changement de régimes

Les bonnes propriétés statistiques de la méthode des m.c.o. supposent que le modèle linéaire :

$$X_t = \sum_{i=1}^{k} Z_t^i b_i + \sum_{j=1}^{l} S_t^j c_j + u_t,$$

fournisse une description correcte de la série brute. Il arrive cependant que la forme fonctionnelle liant Z^i, S^i et X change au cours de la période et qu'un modèle préférable soit par exemple :

$$X_t = \begin{cases} \sum_{i=1}^{k} Z_t^i b_i + \sum_{j=1}^{l} S_t^j c_j + u_t, & si\ t \leq t_0, \\ \sum_{i=1}^{k^*} Z_t^{i*} b_i^* + \sum_{j=1}^{l^*} S_t^{j*} c_j^* + u_t, & si\ t > t_0. \end{cases} \tag{2.32}$$

Dans ce cas une désaisonnalisation effectuée en utilisant le modèle (2.28) conduit à de mauvais résultats. Il est donc nécessaire de déceler les changements de régime en observant la façon dont évolue X_t.

Les changements de régime, s'ils sont décelés [voir Monfort (1982, p. 263] doivent être pris en compte dans l'estimation. Considérons par exemple, le modèle (2.32) correspondant à un changement de régime ; notons A_t la variable valant 1, si $t \leq t_0$ et 0, sinon. On a :

$$X_t = \sum_{i=1}^{k} Z_t^i A_t b_i + \sum_{i=1}^{k^*} Z_t^{i*}(1 - A_t) b_i^* + \sum_{j=1}^{l} S_t^j A_t c_j + \sum_{j=1}^{l^*} S_t^{j*}(1 - A_t) c_j^* + u_t.$$

(2.32) définit donc un modèle linéaire dans lequel les nouvelles variables explicatives sont $Z^i A$, $Z^{i*}(1 - A)$, $S^j\ A$, $S^{j*}(1 - A)$ et les paramètres peuvent être estimés par la méthode des m.c.o. ou des m.c.g. appliquée à ce nouveau modèle.

2.10 Application au "trafic voyageur"

A titre d'exemple, nous utilisons pour décrire cette série mensuelle la formulation la plus classique comportant une tendance linéaire : $b_2 t + b_1$, et une saisonnalité de période 12 :

$$\sum_{j=1}^{12} c_j S_t^j, \text{ avec } \sum_{j=1}^{12} c_j = 0,$$

où S_t^j est la fonction indicatrice du mois j.

Les paramètres b_1 et b_2 de la tendance et les 12 coefficients saisonniers c_j ont été estimés par les moindres carrés ordinaires, calculés à partir des observations de la période 1963-1979. Les 12 dernières observations relatives à l'année 1980 ont été retirées de cette estimation de manière à pouvoir comparer sur cette année les réalisations avec les prévisions déduites du modèle estimé.

Dans un premier temps, on a calculé les estimateurs non contraints de b_2 et des $\delta_j = b_1 + c_j$, $j = 1, ..., 12$. Les estimations sont les suivantes :

$\hat{b}_2$	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\delta}_3$	$\hat{\delta}_4$	$\hat{\delta}_5$	$\hat{\delta}_6$	$\hat{\delta}_7$	$\hat{\delta}_8$	$\hat{\delta}_9$	$\hat{\delta}_{10}$	$\hat{\delta}_{11}$	$\hat{\delta}_{12}$
6,2	1552	1444	1639	1893	1809	2210	2579	2262	1728	1645	1483	2122

L'estimation de b_1 s'en déduit en prenant la moyenne des douze derniers chiffres : $\hat{b}_1 = 1864$.

Les coefficients saisonniers estimés $\hat{c}_j$ sont :

Janv.	Fév.	Mars	Avril	Mai	Juin	Juillet	Août	Sept.	Oct.	Nov.	Déc.
-310	-420	-225	29	-55	346	715	398	-136	-219	-381	258

La somme de ces coefficients est évidemment égale à 0. Comme on pouvait s'y attendre, les coefficients saisonniers correspondant aux vacances (avril, juin, juillet, août, décembre) sont positifs, ce qui traduit l'augmentation du trafic durant ces périodes.

La qualité de l'ajustement peut être étudiée en reportant sur un même graphique les valeurs observées de la série X_t et les valeurs ajustées $\hat{X}_t$ (voir figure 2.33) ou en examinant l'évolution des résidus d'estimation (figure 2.34). Cette dernière montre que l'allure générale des résidus ne correspond pas exactement à une suite de variables aléatoires centrées non corrélées de même variance et donc que la forme retenue n'est pas parfaitement adaptée. En particulier, la figure 2.34 donnant l'évolution des résidus de 1963 à 1979 semble indiquer une convexité de la tendance ; celle-ci pourrait être prise en compte en employant un modèle additif, où la tendance serait modélisée par un polynôme de degré deux ou en utilisant une version mensuelle du modèle multiplicatif décrit plus haut. On peut aussi remarquer que l'hypothèse de stabilité des coefficients saisonniers pourrait être remise en cause au vu de la figure 2.34.

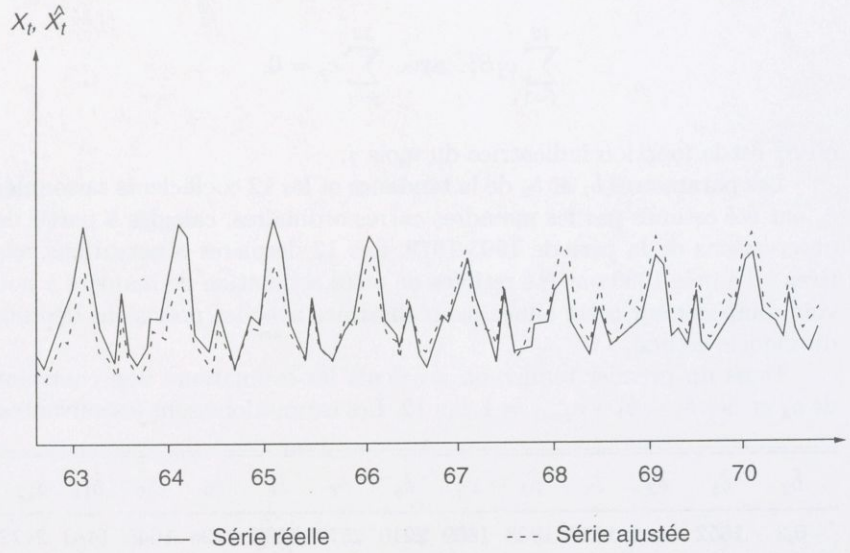

Figure 2.33
Comparaison sur la période 1963-1970 des séries réelle et ajustée

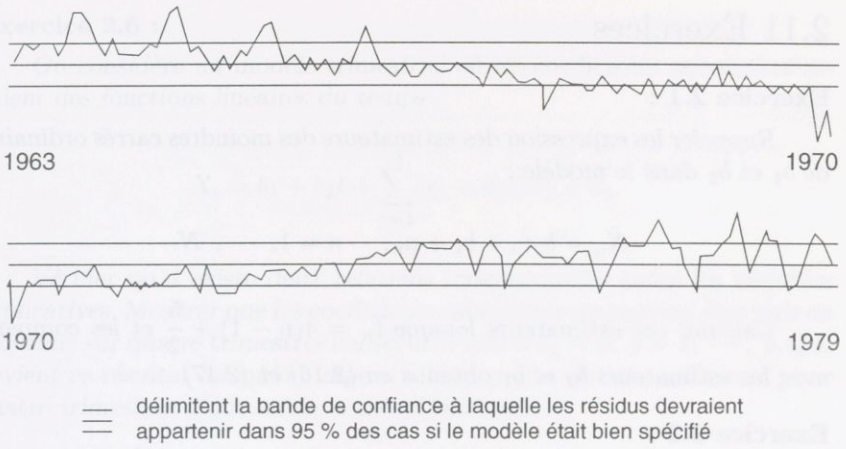

Figure 2.34
Evolution des résidus d'estimation

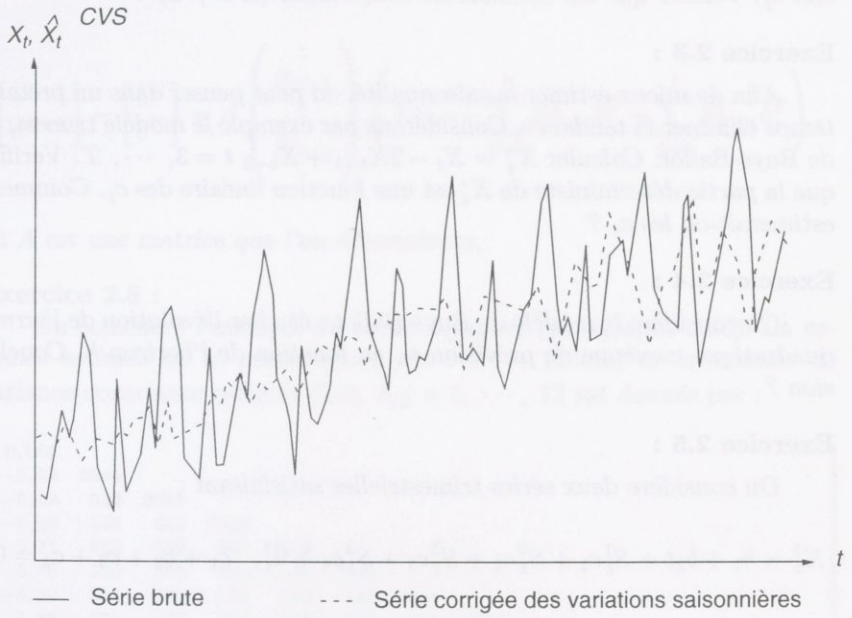

Figure 2.35
Série corrigée des variations saisonnières sur la période 1972-1979

La série désaisonnalisée s'obtient en retranchant de la série initiale les estimations des coefficients saisonniers. Elle est représentée dans la figure 2.35 avec la série brute. On voit que la désaisonnalisation a pour effet d'amortir les pics et les creux de la série initiale.

L'écart relatif moyen entre la prévision et la réalisation pour les douze mois de l'année 1980 est de 6,9 %.