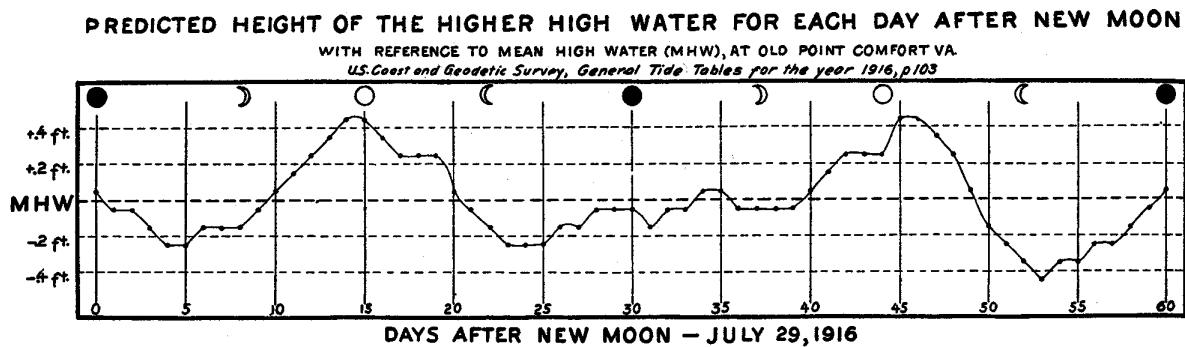


COURS DE SERIES TEMPORELLES THEORIE ET APPLICATIONS



VOLUME 1

Introduction à la théorie des processus en temps discret
Modèles ARIMA et méthode Box & Jenkins

ARTHUR CHARPENTIER
arthur.charpentier@ensae.fr

Contents

1	Introduction et notations	5
1.1	Approches temps/fréquences : un peu d'histoire	5
1.1.1	Analyse harmonique	6
1.1.2	Modèles autoregressifs et moyennes mobiles	7
1.1.3	L'approche temporelle : concept de corrélation sérielle	8
1.1.4	L'équivalence entre les deux approches temps/fréquence	8
1.2	Les développements récents	9
1.2.1	Les modèles <i>ARMA</i> , <i>ARIMA</i> et <i>SARIMA</i> : modèles linéaires	9
1.2.2	Modèles <i>ARCH</i> - volatilité stochastique	9
1.2.3	Les processus à mémoire longue	10
1.2.4	Les processus multivariés	11
1.2.5	Exemple : histoire de la prévision des modèles économiques (<i>macroéconomiques</i>)	12
1.2.6	Remarque sur les processus de comptage ou à valeurs dans un espace d'états finis	13
1.2.7	Remarque sur les données hautes fréquences	13
1.3	Théorie des processus à temps discret	14
1.3.1	Stationnarité des processus	14
1.3.2	Propriété de Markov en temps discret	16
1.4	Objectifs de l'études des séries temporelles	17
1.4.1	Description et modélisation	17
1.4.2	Prévision	17
1.4.3	Filtrage	17
1.5	Conseils bibliographiques	17
2	Propriétés des processus univariés en temps discret	19
2.1	Rappels sur les martingales à temps discret	19
2.2	Rappels sur les Chaînes de Markov	19
2.3	Notions de ' <i>processus stationnaire</i> ' et de ' <i>processus non-stationnaire</i> '	20
2.4	Fonction d'autocovariance et densité spectrale	21
2.4.1	Autocovariance et autocorrélation	21
2.4.2	Densité spectrale	21
2.4.3	Estimation de la fonction d'autocorrélation	23
2.4.4	Estimation de la densité spectrale	24
2.5	Lien entre processus en temps continu et en temps discret	25
3	Désaisonnalisation par regression linéaire	29
3.1	Présentation des données	29
3.2	Le modèle linéaire	29
3.2.1	Hypothèses sur les erreurs	29
3.2.2	Composante saisonnière du modèles	30
3.2.3	Composante tendancielle	30
3.2.4	Modèle trimestriel de Buys-Ballot (1847)	31
3.3	Estimateur des moindres carrés ordinaires (<i>mco</i>)	32
3.3.1	Solutions générales	32
3.3.2	Cas particulier : le modèle trimestriel de Buys-Ballot	33
3.3.3	Généralisation des formules de Buys-Ballot (<i>tendance linéaire</i>)	34
3.4	Application au trafic voyageur	34
3.4.1	Série agrégée par trimestre	34
3.4.2	Analyse sur données mensuelles	37
3.5	Propriétés statistiques des estimateurs	38
3.6	Prévision à un horizon h	38
3.6.1	Calcul de la prévision	38
3.6.2	Application au trafic SNCF	39

4	Désaisonnalisation par la méthode des moyennes mobiles	40
4.1	Généralités sur les moyennes mobiles	40
4.1.1	Notion d'opérateur retard L	40
4.1.2	Les moyennes mobiles	41
4.1.3	L'espace des opérateurs moyenne-mobile	45
4.2	Vecteurs propres associés à une moyenne mobile	45
4.2.1	Les séries absorbées : $\lambda = 0$	46
4.2.2	Absorbition de la composante saisonnière	46
4.2.3	Les séries invariantes : $\lambda = 1$	46
4.2.4	Transformation de suites géométriques (r^t)	47
4.2.5	Moyenne mobile différence $\Delta_p = (\mathbb{I} - L)^p$	48
4.2.6	Moyenne mobile différence saisonnière $\Delta_{p,s} = (\mathbb{I} - L^s)^p$	49
4.2.7	Moyenne mobile impaire	49
4.2.8	Moyenne mobile paire	50
4.3	Notion de 'bruit blanc'	51
4.3.1	Transformation d'un bruit blanc	51
4.4	Les procédures $X11$ et $X12$	52
4.4.1	Un algorithme simple de désaisonnalisation	52
4.4.2	L'algorithme de base de la méthode $X11$	52
4.4.3	Amélioration de la méthode : $X11 - ARIMA$ et $X12 - ARIMA$	54
4.4.4	Utilisation de la méthode $X11$ et comparaison avec les modèles $ARIMA$ saisonniers	54
4.4.5	Exemple simple inspiré de la méthode $X11$	55
4.5	Description du modèle $TRAMO/SEATS$	56
5	La prévision par lissage exponentiel	57
5.1	Principe du lissage exponentiel simple	57
5.1.1	Méthode adaptative de mise à jour (ordre 1)	58
5.1.2	Choix de la constante de lissage	58
5.1.3	Lien entre robustesse de la prévision et choix de β	58
5.1.4	Exemple d'application	59
5.2	Principe de lissage exponentiel double	61
5.2.1	Méthode adaptative de mise à jour (ordre 1)	61
5.2.2	Application de la méthode de lissage exponentiel double	62
5.3	Lissage exponentiel multiple, ou généralisé	64
5.3.1	Méthode adaptative de mise à jour (ordre 1)	66
5.4	Les méthodes de Holt-Winters (1960)	66
5.4.1	Méthode non saisonnière	66
5.4.2	La méthode saisonnière additive	67
5.5	Exemple de mise en pratique des méthodes de lissage	67
5.5.1	Présentation des données	68
5.5.2	Lissage "linéaire"	69
5.5.3	Lissage exponentiel simple	70
5.5.4	Lissage exponentiel double	70
6	Introduction aux modèles linéaires $ARIMA$	73
6.1	Rappels sur les espaces \mathcal{L}^2	73
6.1.1	Propriétés topologiques	73
6.1.2	Rappel sur les vecteurs et processus gaussiens	73
6.1.3	Regression affine dans \mathcal{L}^2	74
6.1.4	La notion d'innovation	75
6.2	Polynômes d'opérateurs retard L et avance F	77
6.2.1	Rappels sur les opérateurs retards	77
6.2.2	Inversibilité des polynômes $P(L)$	77
6.3	Compléments sur les séries stationnaires : les autocorrélations	79
6.3.1	Autocovariance et autocorrélation	79
6.3.2	Autocorrélations partielles	80
6.3.3	Densité spectrale	82
6.3.4	Autocorrélations inverses	83

6.3.5	Complément : autocorrélogrammes de fonctions déterministes	85
6.4	Les processus autorégressifs : $AR(p)$	86
6.4.1	Réécriture de la forme $AR(p)$	86
6.4.2	Propriétés des autocorrélations - les équations de Yule-Walker	87
6.4.3	Le processus $AR(1)$	89
6.4.4	Le processus $AR(2)$	92
6.5	Les processus moyenne-mobile : $MA(q)$	94
6.5.1	Propriétés des autocorrélations	94
6.5.2	Le processus $MA(1)$	96
6.5.3	Le processus $MA(2)$	97
6.6	Les processus $ARMA(p, q)$	99
6.6.1	Propriétés des autocorrélations	99
6.6.2	Densité spectrale des processus $ARMA(p, q)$	100
6.6.3	Les processus $ARMA(1, 1)$	100
6.7	Introduction aux modèles linéaires non-stationnaires	101
6.8	Les processus $ARIMA(p, d, q)$	102
6.8.1	Processus $ARIMA$ et formes AR ou MA	104
6.9	Les modèles $SARIMA$	104
6.10	Théorème de Wold	107
6.11	Théorie spectrale et processus $ARIMA$	108
6.11.1	Théorie spectrale et notion de filtre	108
6.11.2	Le spectre d'un processus $ARMA$	109
6.11.3	Estimation de la densité spectrale d'un processus	111
7	Estimation des modèles $ARIMA$: Box-Jenkins	114
7.1	Estimation du paramètre d'intégration d	114
7.1.1	Approche empirique par l'autocorrélogramme	114
7.1.2	Tests de racine unité	115
7.1.3	Tests de racines unitaires saisonnières	121
7.1.4	Complément sur la notion de sur-différentiation	122
7.2	Estimation des ordres p et q d'un modèle $ARMA(p, q)$	123
7.2.1	Problèmes d'unicité de la représentation $ARMA$	124
7.2.2	Comportement asymptotique des moments empiriques	125
7.2.3	Méthode pratique d'estimation des ordres p et q	125
7.2.4	Cas d'un processus $MA(q)$	126
7.2.5	Cas d'un processus $ARMA(p, q)$	126
7.2.6	Propriété des estimateurs	129
7.3	Test de bruit blanc et de stationnarité	129
7.3.1	Analyse des fonctions d'autocorrélation	129
7.3.2	Statistique de Box-Pierce, ou test de "portmanteau"	130
7.3.3	Compléments : les tests de normalité	130
7.3.4	Complément : Test de rupture et de changement de tendance	132
7.4	Estimation des paramètres d'un modèle $ARMA(p, q)$	134
7.4.1	Estimation pour les modèles $AR(p)$ par la méthode des moindres carrés	134
7.4.2	Vraisemblance d'un processus $ARMA(p, q)$	134
7.4.3	Résolution du programme d'optimisation	135
7.4.4	Tests statistiques de validation du modèle	139
7.5	Choix d'un modèle	139
7.5.1	Critère de pouvoir prédictif	139
7.5.2	Critère d'information	140
7.6	Application	140
7.6.1	Identification du modèle : recherche des paramètres d, p et q	141
7.6.2	Estimation du modèle $ARIMA$	142
7.6.3	Vérification	143

8	Prévisions à l'aide des modèles <i>ARIMA</i> : Box-Jenkins	145
8.1	Prévisions à l'aide d'un modèle <i>AR</i> (p)	145
8.2	Prévisions à l'aide d'un modèle <i>MA</i> (q)	145
8.2.1	Utilisation de l'écriture <i>AR</i> (∞) du processus <i>MA</i> (q)	146
8.2.2	Utilisation de la formule de mise à jour des résultats	146
8.3	Prévisions à l'aide d'un modèle <i>ARMA</i> (p, q)	146
8.3.1	Utilisation de la forme <i>AR</i> (∞) pu processus <i>ARMA</i> (p, q)	147
8.3.2	Utilisation de la forme <i>MA</i> (∞) pu processus <i>ARMA</i> (p, q) et des formules de mise à jour	147
8.4	Prévisions dans le cas d'un processus <i>ARIMA</i> (p, d, q)	147
8.4.1	Utilisation de l'approximation <i>AR</i>	148
8.4.2	Utilisation de l'approximation <i>MA</i>	148
8.5	Intervalle de confiance de la prévision	149
8.6	Prévision pour certains processus <i>AR</i> et <i>MA</i>	149
8.6.1	Prévision pour un processus <i>AR</i> (1)	149
8.6.2	Prévision pour un processus <i>MA</i> (1)	150
8.6.3	Prévision pour un processus <i>ARIMA</i> (1, 1, 0)	152
8.7	Application	153
8.7.1	Exemple de prévision : cas d'école	153
8.7.2	Exemple d'application : cas pratique	155
9	Applications de la méthode de Box & Jenkins	156
9.1	Application à un portefeuille d'assurance-vie	156
9.1.1	Modélisation de la série	156
9.1.2	Estimation de modèles <i>ARMA</i>	157
9.2	Application de la série des taux d'intérêt à 3 mois	161
9.2.1	Modélisation de la série	162
9.2.2	Estimation des paramètres d'une modélisation <i>ARIMA</i> (1, 1, 1)	162
9.2.3	Estimation des paramètres d'une modélisation <i>ARIMA</i> (2, 1, 2)	163
9.2.4	Estimation des paramètres d'une modélisation <i>ARIMA</i> (4, 1, 4)	164
9.2.5	Estimation des paramètres d'une modélisation <i>ARIMA</i> (8, 1, 2)	165
9.2.6	Estimation des paramètres d'une modélisation <i>ARIMA</i> (8, 1, 4)	165
9.2.7	Choix du modèle	166
9.3	Application à des données simulées	166
9.4	Modélisation du trafic autoroutier	170
9.4.1	Modélisation du trafic sur l'autoroute <i>A7</i>	170
9.4.2	Modélisation du trafic sur l'autoroute <i>A13</i>	171
9.5	Modélisation du nombre de victimes sur les routes	173
9.6	Modélisation du taux de croissance du <i>PIB</i> américain	176

”La statistique est la première des sciences inexactes.”
Edmond et Jules de Goncourt, Journal

1 Introduction et notations

L'étude des séries temporelles, ou séries chronologiques, correspond à l'analyse statistique d'observations régulièrement espacées dans le temps. Elles ont été utilisées en astronomie (*'on the periodicity of sunspots'*, 1906), en météorologie (*'time-series regression of sea level on weather'*, 1968), en théorie du signal (*'Noise in FM receivers'*, 1963), en biologie (*'the autocorrelation curves of schizophrenic brain waves and the power spectrum'*, 1960), en économie (*'time-series analysis of imports, exports and other economic variables'*, 1971)...etc.

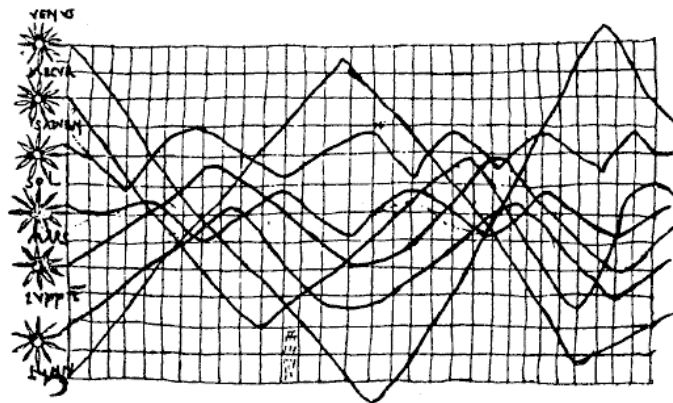
1.1 Approches temps/fréquences : un peu d'histoire

De façon générale, quand on parle de séries stationnaires, on a en tête une représentation de la forme X_t , où $t \in \mathbb{Z}$, représentant les observations (*potentielles*) du processus, dont on peut définir un ensemble d'autocovariance $\gamma(t, s) = \mathbb{E}([X_t - \mu][X_s - \mu])$ qui ne dépend que la distance entre t et s , $\gamma(t, s) = \gamma(t + h, s + h)$ pour tout $h \in \mathbb{Z}$ (*notion 'faible' de stationnarité*). On demande généralement à cette autocovariance $\gamma(t, s)$ de tendre vers 0 quand la différence entre t et s tend vers l'infini : la covariance entre des éléments très éloignés dans la série tend vers 0.

Cette approche, basée sur l'utilisation des corrélations, correspond à l'analyse de type '**temporelle**' : elle consiste à étudier les corrélations croisées de fonctions de la série (X_t). Ces méthodes sont généralement paramétriques de type moyenne-mobiles (*moving average MA*) ou autorégressives (*AR*) - voire les deux (*ARMA*). Toutes ces méthodes consistants à estimer des paramètres peuvent généralement être vus comme des généralisations de la régression linéaire.

L'autre approche également utilisée est celle basée sur l'étude des '**fréquences**'. Cette vision est une généralisation des méthodes utilisées en analyse de Fourier. L'idée est ici d'approximer une fonction analytique par une somme pondérée de fonctions sinus ou cosinus.

Historiquement, ce sont les astronomes qui les premiers ont travaillé sur des séries chronologiques. La reproduction ci-dessous est tiré d'un manuscrit du X^e siècle, représentant l'inclinaison des orbites des planètes en fonction du temps. C'est en particulier grâce à ce genre de données que Kepler a pu énoncer ses lois sur le mouvement des planètes.



Ces visualisations graphiques ont permis, grâce aux différents outils mathématiques mis en place au XVIII^e et XIX^e siècles, de mettre en place les premières techniques d'étude des séries chronologiques¹, parmi lesquelles, l'analyse harmonique.

¹En fait, comme le note Bernstein dans *Against the Gods (the remarkable story of risk)*, les grecs ou les hébreux ont observés des phénomènes cycliques (*par exemple*), mais ils n'ont jamais pensé à faire de la prévision. Il a fallu attendre la Renaissance pour que l'avenir ne soit plus qu'une question de chance ou un fruit du hasard. Y compris au XVIII^e siècle, prévoir des phénomène futurs pouvait faire croire à une tentative de rivaliser avec les dieux : Halley remarqua que la même comète fut aperçue en 1531, en 1607 et en 1682 (*cette comète avait été observée d'ailleurs depuis 240 avant J.C.*), et il prévoit qu'on la reverra en 1758 (*ce fut effectivement le cas, au grand émoi de toute l'Europe, puisque tous les 76 ans, la comète, dite de Halley, arrive en vue de la terre*).

1.1.1 Analyse harmonique

Les astronomes ont été les premiers à utiliser l'analyse de Fourier pour des séries chronologiques. Leur but était de détecter des saisonnalités cachées au sein de leurs données. Ainsi, Lagrange a utilisé ces méthodes pour détecter de la périodicité cachée en 1772 et en 1778. Un demi-siècle plus tard, en 1847, Buys et Ballot, dans '*Les changements périodiques de températures*' ont proposé des méthodes pour étudier la périodicité de données astronomiques. Toutefois, il a fallu attendre 1889 pour que Sir Arthur Shuster introduise le périodogramme, qui constitue la base des méthodes spectrales d'analyse de séries chronologiques.

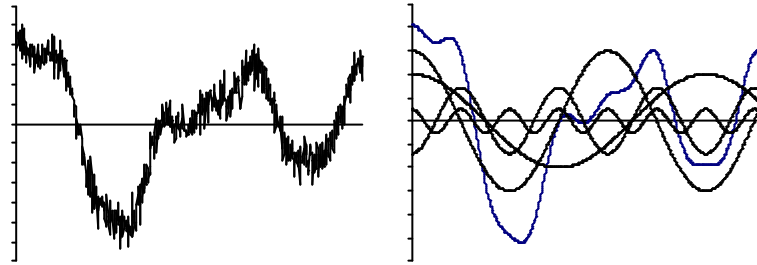
L'idée est la suivante : on recherche un modèle sous-jacent de la forme

$$Y_t = \sum \rho_j \cos [\omega_j t - \theta_j] + \varepsilon_t = \sum [\alpha_j \cos (\omega_j t) + \beta_j \sin (\omega_j t)] + \varepsilon_t,$$

où (ε_t) est une suite de variables aléatoires indépendantes identiquement distribuées, qui correspondront à un 'bruit blanc' (*cette notion serait longuement développée par la suite*).

Le facteur ρ_j (ou $\sqrt{\alpha_j^2 + \beta_j^2}$) correspond à l'amplitude de la j -ème composante périodique, et indique le poids de cette composante au sein de la somme.

Exemple 1 *Considérons la série temporelle ci-dessous à gauche. Une fois enlevé le bruit, nous obtenons une série qui peut être décrite comme une somme pondérée de fonctions sinusoïdales*



En l'occurrence, la série de gauche peut être vue comme la somme d'un bruit et de 4 fonctions sinusoïdales (d'amplitudes ρ_j différentes).

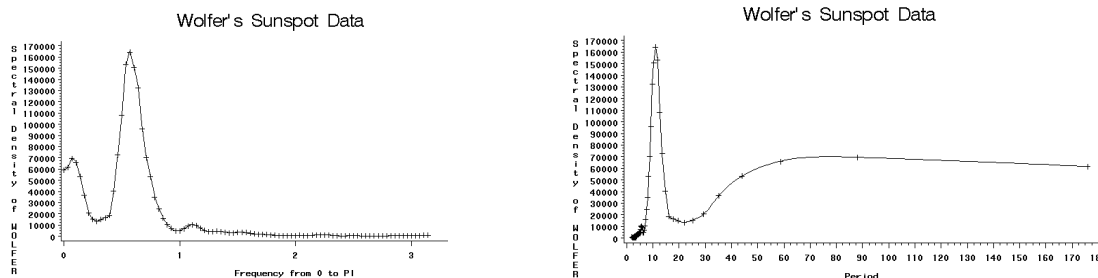
A partir d'un échantillon Y_0, \dots, Y_{T-1} , et en considérant les fréquences $\omega_j = 2\pi j/T$, le **périodogramme** est défini par

$$I(\omega_j) = \frac{2}{T} \left[\left(\sum Y_t \cos(\omega_j t) \right)^2 + \left(\sum Y_t \sin(\omega_j t) \right)^2 \right] = \frac{T}{2} (a^2(\omega_j) + b^2(\omega_j)).$$

Il est alors possible de montrer que $2I(\omega_j)/T$ est un estimateur consistant de ρ_j^2 (*au sens où cet estimateur converge en probabilité quand le nombre d'observations augmente*). Cette convergence a été longuement étudiée par Yule en 1927.

Exemple 2 *En considérant la série chronologique du nombre de taches solaires*

Wolf a obtenu le périodogramme représenté ci-dessous (à gauche en fréquence ω , à droite en période $f = 2\pi/\omega$),



et a ainsi pu montrer qu'il y avait une périodicité de 11 années dans le cycle solaire. Ce résultat a été retrouvé par la suite avec davantage de données, et un second cycle de l'ordre de 160 années a également été détecté.

En 1924, Whittaker et Robinson ont utilisé cette théorie sur la brillance de l'étoile T-Ursa Major, observée sur 600 jours, et ont montré que la brillance pouvait être modélisée (*presque parfaitement*) à l'aide de 2 fonctions harmoniques, de périodes respectives 24 et 29 jours.

Si cette théorie a donné de très bons résultats en astronomie, son application en économie a conduit à des résultats nettement moins concluants. En 1921 et 1922, Beveridge a utilisé cette théorie sur le prix du blé (*wheat prices and rainfall in western europe*). La série présentait tellement de pics qu'au moins 20 périodicités étaient possibles... et plus encore si l'on commençait à prendre en compte de facteurs économiques ou météorologiques.

Si les phénomènes astronomiques permettent d'utiliser cette théorie, c'est parce que des cycles parfaitement réguliers sont observés. Toutefois, cette méthode s'est révélée plus compliquée à mettre en oeuvre en sciences humaines.

1.1.2 Modèles autoregressifs et moyennes mobiles

Deux articles en 1927 ont ouvert une autre voie : l'article de Yule (*on the method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers*) et celui de Slutsky (*the summation of random causes as the source of cyclical processes*).

Yule a introduit dans la littérature les **modèles autorégressifs**, en considérant des modèles de la forme

$$Y_t = \alpha Y_{t-1} + \beta Y_{t-2}.$$

Etant données deux valeurs initiales, cette suite présente un comportement saisonnier, fonction des paramètres α et β . Yule remarque qu'en fait, le comportement dépend des racines (*complexes*) de l'équation $z^2 - \alpha z - \beta = 0$, et plus particulièrement de leur position par rapport au disque unité. Si leur module est inférieur à 1, alors on observe un comportement sinusoïdal amorti. En fait, la forme générale des solutions sera

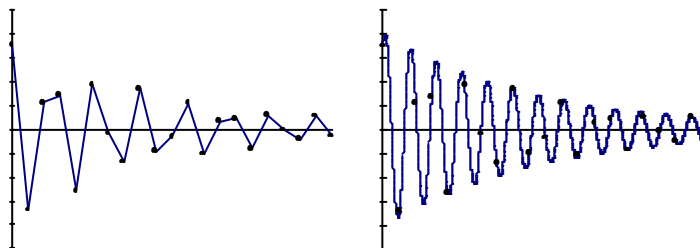
$$Y_t = A\rho^t \cos(\omega t - \theta), \text{ lorsque } 0 < \rho < 1.$$

Le modèle autorégressif proposé par Yule est le suivant

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t, \quad (1)$$

où (ε_t) correspond à un '**bruit blanc**' : un bruit blanc correspond à un processus 'indépendant' (*ou, plus faiblement, non corrélé avec son passé*). Néanmoins, des hypothèses plus fortes doivent parfois être faites : on veut que ce bruit soit également indépendant du passé de la variable Y_t , i.e. ε_t indépendant de Y_{t-h} pour tout $h \geq 1$, et on parle alors d'**innovation** du processus (Y_t) .

Exemple 3 La série définie par $Y_t = 1.8Y_{t-1} - 0.8Y_{t-2}$, représentée ci-dessous à gauche, peut être écrite également $Y_t = -2 \times 0.9^t \cos(4t - 1/2)$, $t \in \mathbb{Z}$,



Slutsky a introduit les **moyennes mobiles** la même année que Yule a introduit les processus autorégressifs... mais son article, écrit en 1927 en russe n'a été traduit qu'en 1937 en anglais. Pour cela, il a utilisé des nombres générés par la lotterie officielle, et a réussi à générer une série décrivant le cycle économique en Angleterre, de 1855 à 1877. La forme générale était la suivante,

$$Y_t = \mu_0 \varepsilon_t + \mu_1 \varepsilon_{t-1} + \dots + \mu_q \varepsilon_{t-q}, \quad (2)$$

où (ε_t) est un bruit blanc, correspondant à ces nombres générés par la lotterie officielle : on obtient des variables indépendantes entre elles (*cf tables de nombres aléatoires*), mais surtout indépendantes du cycle économique. Cette écriture a suggéré d'élargir la relation (1) sous une forme proche de (2), à savoir

$$\phi_0 Y_t + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} = \varepsilon_t.$$

Les processus introduits par Yule deviendront les processus $AR(p)$ et ceux introduits par Slutsky les processus $MA(q)$. L'analogie entre les deux processus sera même poussée plus loin lorsqu'il sera montré que les processus $AR(p)$ et $MA(q)$ sont respectivement des processus $MA(\infty)$ et $AR(\infty)$, sous certaines conditions.

1.1.3 L'approche temporelle : concept de corrélation sérielle

Si l'approche spectrale repose sur l'utilisation du spectre (*ou du périodogramme*), l'approche temporelle repose sur l'autocorrélogramme, ou plus généralement sur l'utilisation de la corrélation sérielle. Poynting est le premier à introduire cette idée, en 1884, en étudiant la relation entre le mouvement du prix du blé, et les importations de coton et de soie. Le coefficient de corrélation sérielle a été défini par Hooker en 1901, dans une étude sur le taux de mariage en Angleterre, et l'indice du commerce. Etant données deux séries temporelles, (X_t) et (Y_t) , la **covariance sérielle** est définie par $c_k(X, Y) = cov(X_t, Y_{t+k})$ et la **corrélation sérielle** sera alors $r_k(X, Y) = c_k(X, Y) / c_0(X, Y)$.

Le coefficient d'autocorrélation est alors obtenu en considérant $\rho_k = corr(X_t, X_{t+k}) = r_k(X, X)$. Les années 30 ont alors vu l'éclosion des résultats de base dans le domaine des séries chronologiques, sous l'impulsion de Khintchine, Cramer, Wold, Kolmogorov, Wiener...etc. Ces auteurs ont développé une théorie des séries temporelles, en considérant qu'une série chronologique est une réalisation d'un processus aléatoire.

1.1.4 L'équivalence entre les deux approches temps/fréquence

Dans un premier temps, l'analyse harmonique a été généralisée pour passer d'une somme de Fourier à une intégrale de Fourier

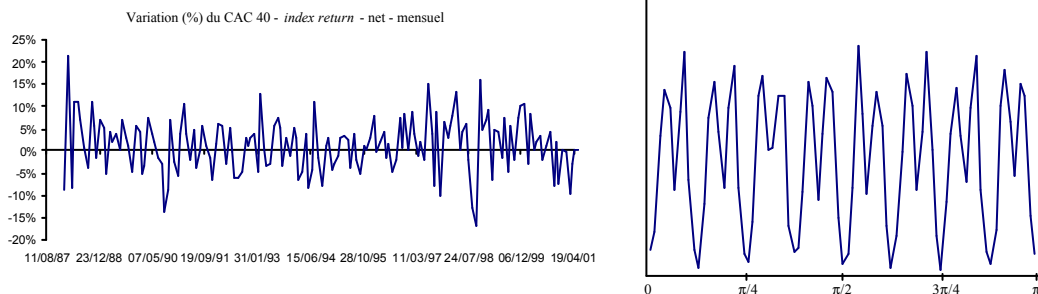
$$Y_t = \int_0^\pi [\cos(\omega t) dA(\omega) + \sin(\omega t) dB(\omega)].$$

Cette simple idée de lissage du périodogramme a permis de contourner les problèmes qu'avait pu observer Beveridge lorsqu'il cherchait des périodicités cachées dans des disciplines autres que l'astronomie.

La synthèse entre ces deux branches (*la première travaillant en temps, avec des autocorrélations, et la seconde travaillant sur le spectre de la série*) a été faite dans les années 30, en parallèle aux Etats-Unis par Norbert Wiener ('*generalised harmonic analysis*', 1930) et en Union Soviétique par Khintchine ('*korrelationstheorie der stationaren stochastischen prozesse*', 1934). Leur résultat est de mettre en avant une relation bijective entre la fonction d'autocovariance d'un processus stationnaire, et sa densité spectrale :

$$g(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} \gamma(h) \cos(\omega h) \quad \text{ou} \quad \gamma(h) = \int_0^\pi \cos(\omega h) g(\omega) d\omega, \quad \text{où} \quad \gamma(h) = cov(X_t, X_{t-h}).$$

Et si l'analogie entre autocorrélogramme et densité spectrale existe d'un point de vue théorique, il est possible de mettre en avant le même genre de relation entre les autocorrélations empiriques et le périodogramme empirique. Les graphiques ci-dessous représentent les variations de l'indice $CAC40$ en données mensuelles, à gauche, et le périodogramme associé en fréquence (ω) à droite,



1.2 Les développements récents

L'étude des séries temporelles semble avoir atteint sa maturité au cours des années 70 où des développements significatifs sont apparus. En 1965, Cooley et Tukey ont beaucoup aidé l'étude spectrale des séries grâce à leur article '*an algorithm for the machine calculation of complex Fourier series*', introduisant la **Fast Fourier Transform (FFT)**. Cet algorithme a permis de calculer rapidement des périodogrammes. A la même époque, en 1970, Box et Jenkins ont publié leur ouvrage '*Time series analysis, forecasting and control*', montrant que l'étude des séries temporelles à l'aide de processus de type *ARMA* (*obtenus en associant les écritures (1) et (2) des processus AR et MA*) pouvait s'appliquer à de nombreux domaines, et pouvait être facilement implémentée informatiquement².

1.2.1 Les modèles ARMA, ARIMA et SARIMA : modèles linéaires

Les modèles *ARMA* sont un mélange des modèles (1) et (2) proposés par Yule et Slutsky. Un processus (X_t) est un processus *ARMA* (p, q) s'il existe un bruit blanc (ε_t) (*c'est à dire un processus stationnaire tel que ε_t et ε_{t-k} soient indépendants, pour tout k , pour tout t*) tel que

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \text{ pour tout } t.$$

Sous certaines conditions, ces processus sont stationnaires. Comme nous le verrons par la suite, ces processus peuvent s'écrire sous la forme

$$\Phi(L) X_t = \Theta(L) \varepsilon_t, \text{ où } \Phi(L) = \mathbb{I} - \phi_1 L - \dots - \phi_p L^p \text{ et } \Theta(L) = \mathbb{I} + \theta_1 L + \dots + \theta_q L^q,$$

L représentant l'**opérateur retard**, au sens où $LX_t = X_{t-1}$, et avec la convention $L^p = L \circ L^{p-1}$, soit $L^p X_t = X_{t-p}$: la série (Y_t) telle que $Y_t = L^p X_t$ est alors la série (X_t) retardée de p périodes.

Parallèlement, on dira qu'un processus non-stationnaire est **intégré** d'ordre 1, si en le différenciant une fois, on obtient un processus stationnaire : (X_t) (*non-stationnaire*) sera dit intégré d'ordre 1 si le processus (Y_t) définit $Y_t = \Delta X_t = X_t - X_{t-1} = (1 - L) X_t$ est stationnaire. On dira, par extension, que (X_t) est intégré d'ordre d si (X_t) est non-stationnaire, ..., (Y_t) où $Y_t = (1 - L)^{d-1} X_t$, est non-stationnaire, et (Z_t) où $Z_t = (1 - L)^d X_t$, est stationnaire. On appellera alors processus *ARIMA* (p, d, q) un processus (X_t) pouvant se mettre sous la forme

$$\Pi(L) X_t = \Phi(L) (1 - L)^d X_t = \Theta(L) \varepsilon_t, \text{ où } (\varepsilon_t) \text{ est un bruit blanc.}$$

Pour les données réelles, on notera que $d = 1, 2$ ou 3 (*au maximum*). Cela signifie que (Y_t) définit comme différence d'ordre d du processus (X_t) , soit $Y_t = (1 - L)^d X_t$, "suit un processus *ARMA* (p, q) "³.

On parlera d'ailleurs de présence de racine unité : 1 est alors racine du polynôme autorégressif $\Pi(z)$. Par généralisation, on peut considérer le cas où $\exp(2i\pi/s)$ est racine du polynôme autorégressif : c'est à dire que $\Pi(L) = (1 - L^s) \Phi(L)$. On dira alors que l'on est présence d'une racine unité saisonnière, qui engendreront les modèles *SARIMA*.

Les modèles intégrés sont très présents dans les séries économiques, par exemple les séries d'indices boursiers, d'indice de production, d'indice de prix.... Les modèles *SARIMA* sont également très présents dès lors que les séries sont très saisonnières (*avec une forte périodicité trimestrielle, annuelle... etc*).

Remarque 1 Parmi les transformations usuelles des variables, la transformation par $(1 - L)$ est parmi les plus utilisées : on ne considère alors plus la série brute (X_t) mais la variation (brute) $Y_t = X_t - X_{t-1}$. Dans le cas où X_t est un prix (par exemple un indice boursier, CAC40 ou SP500), on considère également souvent la variable obtenue comme différence des logarithmes des prix $Z_t = \log X_t - \log X_{t-1}$, qui est alors le rendement ou le taux de croissance ("return" en anglais).

1.2.2 Modèles ARCH - volatilité stochastique

Dans les années 80, des développements ont été apportés dans l'étude de la **non-linéarité** de certaines séries, et sur leur modélisation. En 1982, Engle a introduit la classe des modèles *ARCH* (*autorégressifs conditionnellement hétéroscédastiques*⁴). Ces modèles ont été introduits pour palier à une observation empirique qui n'était pas prise

²Sur les méthodes de prévision en économie, il peut être intéressant de se reporter à '*The past, present and future of macroeconomic forecasting*' de Francis Diebold (1997).

³Ceci n'est qu'une notation : comme nous le verrons par la suite, les processus *ARIMA* sont un peu plus compliqués que les processus *ARMA* puisqu'il faut prendre en compte des conditions initiales : (Y_t) ne suit qu'asymptotiquement un processus *ARMA* (p, q) .

⁴Pour rappel, un modèle économétrique est dit homoscedastique si la variance des erreurs (*centrées*) $\mathbb{E}(\varepsilon_t^2)$ est constante - quelque soit la période d'étude. Dans le cas contraire, on parlera d'hétéroscédasticité. Les modèles sont ici conditionnellement hétéroscédastiques car $E(\varepsilon_t^2 | \varepsilon_{t-1})$ dépend de t .

en compte par les modèles : la volatilité conditionnelle d'une série (Y_t) n'a aucune raison d'être constante. Dans les modèles $AR(1)$, la variance conditionnelle de X_t sachant X_{t-1} est constante : $V(X_t|X_{t-1}) = \sigma^2$ où $V(\varepsilon_t) = \sigma^2$ (*notion d'homoscédasticité*). Engle a cherché un modèle dans lequel la variance conditionnelle de X_t sachant X_{t-1} dépendrait de X_{t-1} , et plus particulièrement, $V(X_t|X_{t-1}) = [\alpha + \beta X_{t-1}^2] \sigma^2$. Pour cela, il a considéré les modèles de la forme

$$X_t = \varepsilon_t \sqrt{h_t}, \text{ où } h_t = \alpha_0 + \alpha_1 X_{t-1}^2.$$

Cette classe de modèle, appelée $ARCH(1)$ a été généralisée sous la forme $ARCH(p)$,

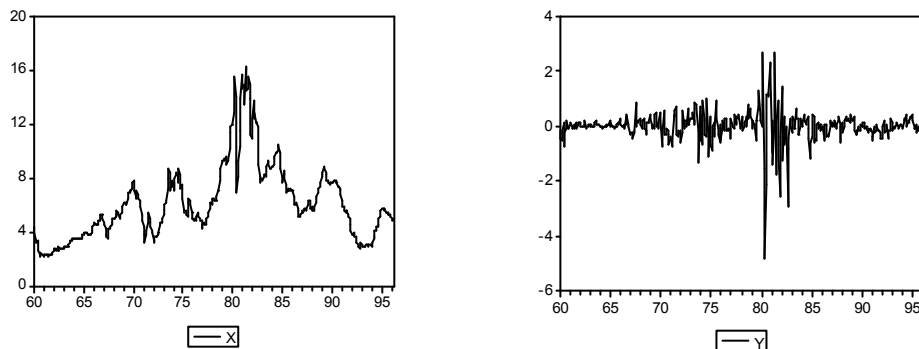
$$X_t = \varepsilon_t \sqrt{h_t}, \text{ où } h_t = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_p X_{t-p}^2.$$

Cette forme pour h_t a permis l'analogie entre les modèles AR et les modèles $ARCH$. De plus, cette classe de modèles $ARCH$ a été généralisée de la même façon que les $ARMA$ généralisent les AR , en considérant des fonctions h_t de la forme

$$h_t = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \varepsilon_{t-j},$$

générant ainsi les modèles $GARCH$.

Exemple 4 Le graphique ci-dessous à gauche correspond à des taux d'intérêt (X_t), et à droite, à la variation de ces taux d'intérêt $Y_t = X_t - X_{t-1}$,

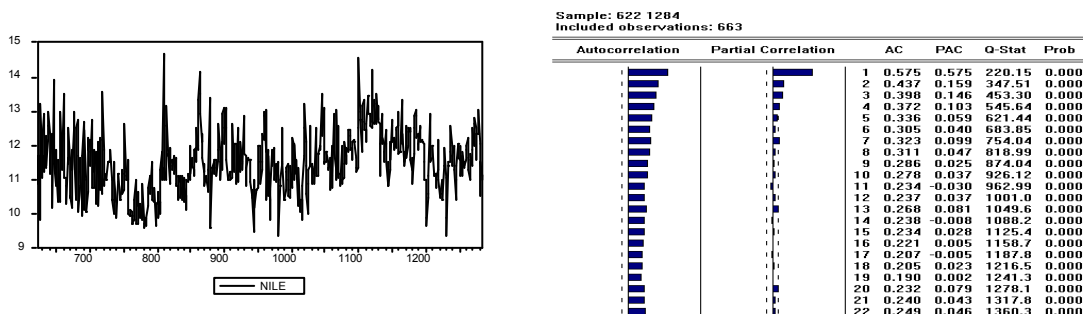


Les longues périodes de fortes volatilité (*volatility clustering*) sont une des caractéristiques des modèles $ARCH$, et c'est, entre autres, pour cela que les modèles $ARCH$ ou $GARCH$ sont énormément utilisés dans les modèles financiers.

1.2.3 Les processus à mémoire longue

D'autres avancées ont été faites sur la **mémoire longue** de certaines séries. Les processus stationnaires de type AR ont un autocorrélogramme qui converge vers 0 de façon exponentielle ($\gamma(h) = \alpha\beta^h$). Les processus à mémoire longue seront caractérisés par une décroissance de leur autocorrélogramme suivant une fonction puissance ($\gamma(h) = \gamma h^\delta$).

Exemple 5 Par exemple, le graphique ci-dessous à gauche correspond au niveau minimum du Nil entre 622 et 1284,



Bien que la série soit stationnaire, les autocorrélations $\gamma(h) = \text{cov}(X_t, X_{t+h})$ sont encore significativement non-nulles après 60 ans (graphique de droite). Ce type de comportement sera appelé 'mémoire longue'.

Plusieurs classes de processus appartiennent à cette série,

(i) les processus self-similaires, introduits par Kolmogorov en 1958 et développés par Mandelbrot (1965) : ces processus sont caractérisés par l'existence d'une constante H (dite de *self-similarité*) telle que, pour tout constante c , la distribution de Y_{ct} soit égale à celle de $c^H Y_t$. On retrouve dans cette classe les processus de Levy.

(ii) les processus *FARMA*, généralisation des modèles *ARIMA* décrits par Box et Jenkins. Ces modèles *ARIMA* étaient obtenus en considérant que les différences premières $\Delta^d X_t$ (où $\Delta X_t = X_t - X_{t-1}$, $\Delta^2 X_t = \Delta(\Delta X_t)$...etc) suivent un processus *ARMA* (p, q) . On parle alors de processus *ARMA* intégré. Les processus *FARIMA* ont été obtenus en considérant, formellement, les cas où d n'est pas entier, compris entre $-1/2$ et $1/2$. Cette généralisation, proposée par Granger en 1980, repose sur la manipulation des séries d'opérateurs retard (L) , et sur le développement en série entière de $(1 - L)^d$.

(iii) l'agrégation de processus *AR* (1) a également été proposée par Granger en 1980 et cette classe de processus a été étudiée par Gourieroux et Gonçalves en 1988. On considère des processus vérifiant, pour tout $t \geq 0$,

$$X_{i,t} = \Phi_i X_{i,t-1} + C_i \varepsilon_t + \eta_{i,t} \text{ pour } i = 1, 2, \dots$$

1.2.4 Les processus multivariés

Enfin, d'autres développements ont été fait dans l'étude des processus multivariés. Si l'on se place uniquement en dimension 2, on comprend que la généralisation des processus univariés à une dimension supérieur est relativement compliquée.

(i) les modèles *VAR* - vecteurs autorégressifs - sont une généralisation des modèles *AR* en dimension n . Si l'on considère par exemple un couple Z_t de deux variables (X_t, Y_t) que l'on souhaite expliquer par leur passé, on obtient un modèle de la forme

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} \phi_1 & \alpha_1 \\ \beta_1 & \psi_1 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}, \text{ soit } Z_t = A_1 Z_{t-1} + U_t,$$

où la matrice A_t est composée des coefficients autoregressifs 'usuels' (ϕ_1 et ψ_1) mais aussi des notions relatives à la notion de causalité, X_t dépendant de Y_{t-1} , et Y_t dépendant de X_{t-1} .

(ii) la cointégration est une notion relative au comportement des plusieurs variables intégrées, et à la relation qui les unit à long terme : on considère (X_t) et (Y_t) non-stationnaires, et intégrées d'ordre d , satisfaisant une relation du type

$$X_t = \alpha + \beta Y_t + \varepsilon_t.$$

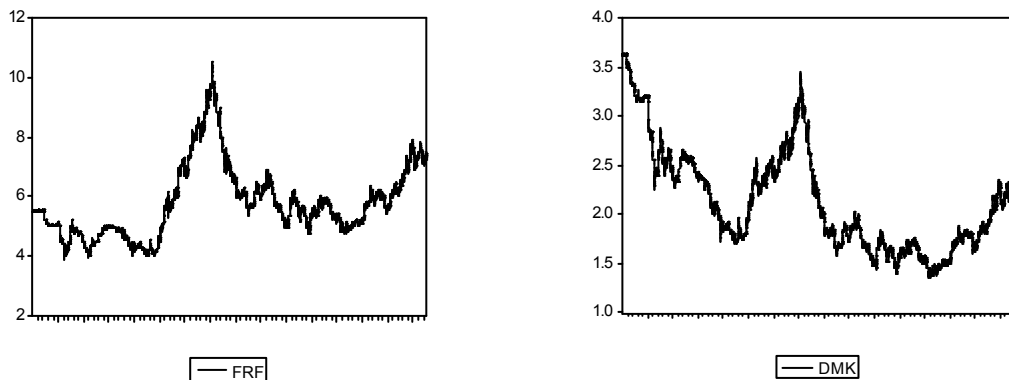
Plus formellement, si le vecteur (Z_t) est intégré d'ordre d , on dira que les séries $Z_t^1, Z_t^2, \dots, Z_t^n$ sont cointégrées si et seulement s'il existe une relation linéaire non-nulle des composantes qui soient intégrées d'ordre strictement inférieur à d

(iii) le modèle 'filtre de Kalman'. Ce modèle est un cas particulier d'une classe plus large de modèles, les **modèles espace d'états**, de la forme

$$\begin{cases} Z_{t+1} = A_t Z_t + \varepsilon_t \\ Y_t = C_t Z_t + \eta_t, \end{cases}$$

où (Y_t) est le vecteur que l'on étudie, (Z_t) est un vecteur aléatoire (=état) inconnu, A_t et C_t sont des matrices déterministes, et (ε_t, η_t) est un bruit blanc normal. L'idée est d'estimer récursivement Z_t en fonction de Y_0, \dots, Y_t .

Exemple 6 *Considérons un entrepreneur américain, investissant dans différents pays. Afin de faire de la prévision de ses résultats, il est nécessaire de prévoir les taux de change des différentes devises : cette prévision doit se faire sur le couple $(r_t^{USD/FRF}, r_t^{USD/DMK})$ et non pas devise par devise.*



En effet, deux analyses faites en parallèle, et indépendamment, peuvent aboutir à des résultats impossibles : il serait incohérent, dans l'exemple ci-dessus, de prévoir que le taux de change du Franc va augmenter, et que le taux de change du Deutsche Mark va baisser.

1.2.5 Exemple : histoire de la prévision des modèles économiques (macroéconomiques)

La théorie économique inspirée de Keynes reposait sur l'utilisation de 'prévisions conditionnelles' : sous certaines hypothèses, les prévisions d'une ou plusieurs variables étaient faites conditionnellement à des comportements, au sein de modèles structurels. Plus particulièrement, dès 1936, Keynes proposait par exemple de lier la consommation C_t au revenu disponible R_t , sous la forme $C_t = \alpha R_t + \beta$: une prévision de R_t permettait de prévoir C_t . Brown avait proposé un modèle légèrement différent dès 1952, en intégrant le fait que les individus ont des habitudes de consommation, entraînant une inertie importante : $C_t = \alpha R_t + \beta + \gamma C_{t-1}$. Ces 'prévisions structurelles' ont toutefois cessé de faire référence à partir des années 70.

Les prévisions non-structurelles ont alors pu prendre en compte les différents cycles observés en économie (1977 : 'Business cycle modeling without pretending to have too much a priori theory' de Sargent et Sims) : des prévisions de séries économiques peuvent se faire sans nécessairement avoir de modèle structurel derrière. Les modèles utilisés sont toutefois relativement anciens puisqu'ils sont inspirés des modèles de Slutsky et Yule, tous deux datant de 1927, basés sur la notion de modèle autorégressif. La publication de l'ouvrage de Box et Jenkins en 1970 permettra une avancée rapide avec l'utilisation des modèles *ARMA*.

Toutefois, le lacune de la théorie de Box et Jenkins est qu'elle ne prend pas en compte des effets croisés de dépendance entre variables. Pour effectuer de la prévision d'un ensemble de variables, a priori liées, il convient d'effectuer une prévision 'globale' : la théorie des modèles *VAR* (modèles autorégressifs vectoriels) a été introduite en économie sous l'impulsion de Sims en 1980, qui a travaillé sur des systèmes d'équations où toutes les variables sont alors endogènes (contrairement aux équations structurelles de Keynes). Cette théorie avait toutefois été étudiée dès les années 70 par Granger par exemple, qui avait travaillé sur la notion 'simple' de causalité entre variables.

Toutefois, la présence d'un certain nombre de variables non-stationnaires a posé un certain nombre de problèmes : Granger a alors introduit la notion de cointégration en 1981 : cette notion dit que deux variables X et Y peuvent suivre une tendance stochastique, mais la différence (ou le spread) $X - Y$ peut être stationnaire. Cette notion sera à l'origine des modèles à tendance commune, permettant de travailler sur des systèmes d'équations où certaines variables sont cointégrées. En particulier, dès 1978, Hall se posait la question de savoir si la consommation par habitant n'était pas une martingale, ce qui conduirait à écrire $C_t = C_{t-1} + \varepsilon_t$ où ε_t est un aléa. Nelson et Plosser ont d'ailleurs noté, en 1982 qu'un grand nombre de séries macroéconomiques étaient caractérisées par la présence d'une racine unitaire (c'est à dire une écriture de la forme $C_t = C_{t-1} + X_t$). Et c'est finalement en 1987 que Campbell a proposé un modèle *VAR* sur la consommation C et le revenu R , puis un modèle *VAR* intégrant dans chaque équation un modèle à correction d'erreur.

Une autre piste qui a été explorée à la même époque est celle des modèles non-linéaires. Cette voie a été ouverte dès 1982 par Engle, qui introduisit de la dynamique dans la volatilité, à l'aide des modèles *ARCH*. Ces modèles ont été très utilisés en finance, mais aussi pour des modèles d'inflation.

Parmi des améliorations apportées dans les années 90, on peut noter les modèles avec cycles, avec rupture de tendance, à changement de régime...etc. La théorie des modèles à changement de régime repose sur l'idée que derrière les variables observées existent des variables cachées, non observées.

Pour résumer l'histoire des applications économiques des séries temporelles, on peut retenir le schéma suivant

- **années 20** : macroéconomie descriptive : description des cycles (*courts = Slutsky, longs = Kondratieff*)
 - **années 50** : début de la théorie des séries temporelles, avec comme objectif principal, la prévision
 - **années 60** : application en macroéconomie, avec des modèles structurels : une vingtaine de variables, et 200 observations (*maximum*)
 - **années 70** : théorie de Box et Jenkins, s'appuyant sur un logiciel (*modèle linéaire*) : on considère les variables une à une, sur 200 observations (*début, à la même époque, de la théorie des panels en microéconomie : 3000 individus suivis sur 3 ou 4 périodes*)
 - **années 80** : en macroéconomie, modèles multivariés (*causalité, cointégration, codépendance*). Début de l'utilisation des modèles de séries temporelles sur données financières : beaucoup de variables, 2000 observations. Début des modèles à temps continu.
 - **années 90** : données hautes fréquences sur les marchés financiers (*de 4000 à plus de 2000000 d'observations*).
- Des compléments peuvent se trouver dans l'article de Chris Chatfield (1997) intitulé 'Forecasting in the 1990s'.

1.2.6 Remarque sur les processus de comptage ou à valeurs dans un espace d'états finis

A retenir 1 Les modèles que nous allons étudier dans ce cours sont basés sur l'étude de processus (X_t) où les variables observées sur supposées à valeurs réelles : $X_1, X_2, \dots, X_t, \dots \in \mathbb{R}$. On observera ainsi des prix, des variations de prix, des taux, des montants...etc. Des nombres de voyageurs seront, à la rigueur, considérés comme une variable réelle, mais deux cas seront exclus de notre étude, a priori :

- (i) les processus de comptage (ex : nombre d'accident pour un conducteur l'année t)
- (ii) les processus à valeurs dans un espace d'état fini

Dans le premier cas, la méthode la plus usuelle pour étudier les processus de comptage est de considérer des modèles de la forme suivante : soit (Y_t) le processus défini par

$$Y_t = U_t + \varepsilon_t = \sum_{i=1}^{Y_{t-1}} U_{i,t} + \varepsilon_t \text{ où } \begin{cases} (U_{i,t}) \text{ i.i.d.} & U_{i,t} \sim \mathcal{B}(p) \\ (\varepsilon_t) \text{ i.i.d.} & \varepsilon_t \sim \mathcal{P}(\lambda). \end{cases}$$

Les processus (U_t) et (ε_t) sont indépendants, conditionnellement à $Y_{t-1}, Y_{t-2}, \dots, U_{t-1}, U_{t-2}, \dots$. De plus, pour tout t , U_t suit une loi binomiale $\mathcal{B}(Y_{t-1}, p)$. On a alors

$$Y_t \sim \mathcal{P}\left(\frac{\lambda}{1-p}\right), \text{ et donc } \mathbb{E}(Y_t) = \frac{\lambda}{1-p}.$$

De plus, la fonction d'autocorrélation est donnée par

$$\gamma(h) = \text{cov}(Y_t, Y_{t-h}) = \frac{p^h \lambda}{1-p}.$$

Ce type de processus est appelé 'Poisson $AR(1)$ ' ou $INAR(1)$, selon la terminologie de Al-Osh et Alzaïd (1987)⁵. Pour information, les paramètres p et λ dans ce genre de modèle sont généralement estimés par maximum de vraisemblance. McKenzie (1988)⁶ a également montré qu'il est possible de faire de la prévision à l'aide de ce genre de modèles, puisque la loi de Y_{T+h} conditionnellement au passé observé jusqu'à la date T vérifie

$$\mathbb{P}(Y_{T+h} = x | X_T = x_T) = \sum_{s=0}^{\min\{x, x_T\}} \binom{x_T}{s} \pi^s (1-\pi)^{x_T-s} \frac{1}{(x-s)!} \left(\lambda \frac{1-\pi}{1-p} \right) \exp\left(-\lambda \frac{1-\pi}{1-p}\right), \text{ où } \pi = p^h,$$

dont la moyenne et la variance sont

$$\mathbb{E}(Y_{T+h} | X_T = x_T) = p^h x_T + \lambda \frac{1-p^h}{1-p} \text{ et } V(Y_{T+h} | X_T = x_T) = p^h (1-p^h) x_T + \lambda \frac{1-p^h}{1-p}.$$

Dans le cas où la loi de Poisson n'est pas la plus adaptée, McKenzie (1986)⁷ a proposé d'autres modèles pour avoir, marginalement, une loi géométrique, ou une loi binomiale négative.

Dans le second cas, le plus simple est de se ramener à la théorie des chaînes de Markov.

1.2.7 Remarque sur les données hautes fréquences

A retenir 2 Les modèles que nous allons étudier dans ce cours sont basés sur l'étude de processus (X_t) , observés à des dates régulières : $X_1, X_2, \dots, X_t, \dots$. Il peut s'agir, par exemple, de la version discrète d'un processus en temps continu : on observe $X_{t_1}, X_{t_2}, \dots, X_{t_n}, \dots$ où les dates t_i sont telles que $t_i - t_{i-1}$ soit constante pour tout i .

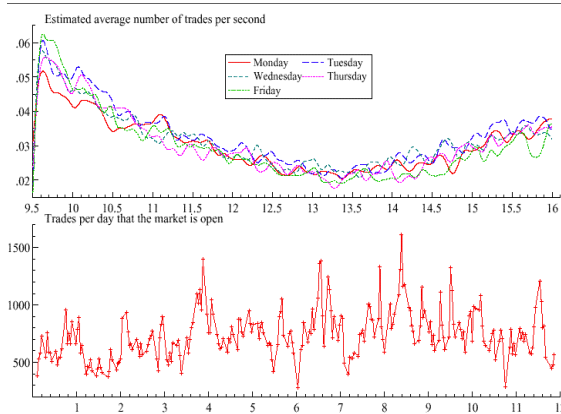
Dans le cas des données hautes fréquences, l'analyse est relativement différente, puisque l'aspect temporel doit être pris en compte. Par exemple, pour étudier la liquidité des marchés financiers, on considère les triplets de variables suivants : (T_i, V_i, P_i) , où T_i est la date de la i ème transaction, V_i le volume échangé lors de la transaction, et P_i le prix de cette transaction. Cette étude permet de changer l'échelle des temps : on ne considère plus le temps calendaire

⁵AL-OSH, M.A. & ALZAID, A. (1987). First-order interger-valued autoregressive ($INAR(1)$) process. *Journal of Time Series Analysis*. **8** 261-275.

⁶McKENZIE, E. (1988). Some $ARMA$ models for dependent sequences of Poisson counts. *Advances in Applied Probability*. **20** 822-835.

⁷McKENZIE, E. (1986). Autoregressive moving-average processes with negative-binomial and geometric marginal distribution. *Advances in Applied Probability*. **18** 679-705.

mais le temps des transactions. Et comme le montre le graphique ci-dessous, ces deux temps peuvent être relativement différents



Obs	Time	PRICE	VOL
868	10:00:04	1378.50	500
869	10:00:06	1379.75	1000
870	10:00:07	1379.50	800
871	10:00:08	1379.50	1250
872	10:00:13	1378.25	300
873	10:00:15	1379.00	1500
874	10:00:18	1379.75	500
875	10:00:19	1378.25	750
876	10:00:22	1380.25	1250
877	10:00:23	1379.50	1500
878	10:00:27	1379.50	1750
879	10:00:28	1380.25	250
880	10:00:29	1380.50	500
811	10:00:32	1375.75	500

En effet, il y a trois fois plus de transactions sur un marché tôt le matin qu'à l'heure du déjeuner : le temps entre deux transactions sera alors, en moyenne, trois fois plus long à 13 heures 30 qu'à 9 heures 45. Cette différence qui existe entre les heures de la journée peut se retrouver entre les jours de la semaine (*il y a ainsi, en moyenne, 10% de transaction en plus le mardi, comparé au vendredi ou au lundi*), ou au mois (*il y a, par jour ouvré, près de deux fois plus de transactions en septembre qu'en février*).

La notion de base pour étudier ce genre de données est la théorie des modèles de durées. On considère (T_i) , la suite des dates de transaction, et τ_i la date écoulée entre la i ème et la $i-1$ ème transaction : $\tau_i = T_i - T_{i-1}$.

Toutefois, dans ce cours, nous ne traiterons pas de ces aspects, mais nous considérerons plutôt des agrégations, ou des observations ponctuelles : P_t sera le prix observé à la date t (*par exemple tous les jours, ou toutes les heures*) et V_t le volume total échangé pendant la période (*en une journée, ou une heure*). Toutefois, il est à noter que même dans ce cas, où les volumes de données sont très importants, l'étude peut s'avérer plus complexe que dans le cas où l'on considère des séries économiques observées à 200 dates, en particulier à cause de la présence de multiples cycles (*un cycle d'une journée sera observée sur des données horaires par exemple, puis des cycles mensuels, ou trimestriels (publication de comptes), ou encore annuels...*).

1.3 Théorie des processus à temps discret

Deux types de processus sont utilisés dans la théorie des séries stationnaires

- (i) les processus stationnaires
- (ii) les processus markoviens

1.3.1 Stationnarité des processus

La **stationnarité** joue un rôle central dans la théorie des processus, car elle remplace (*de façon naturelle*) l'hypothèse d'observation i.i.d. en statistique. Deux notions sont généralement considérées. La première notion de stationnarité peut se définir de façon forte par une stabilité en loi du processus : quel que soit n , t_1, \dots, t_n et h , on a l'égalité entre les lois jointes

$$\mathcal{L}(Y_{t_1}, \dots, Y_{t_n}) = \mathcal{L}(Y_{t_1+h}, \dots, Y_{t_n+h})$$

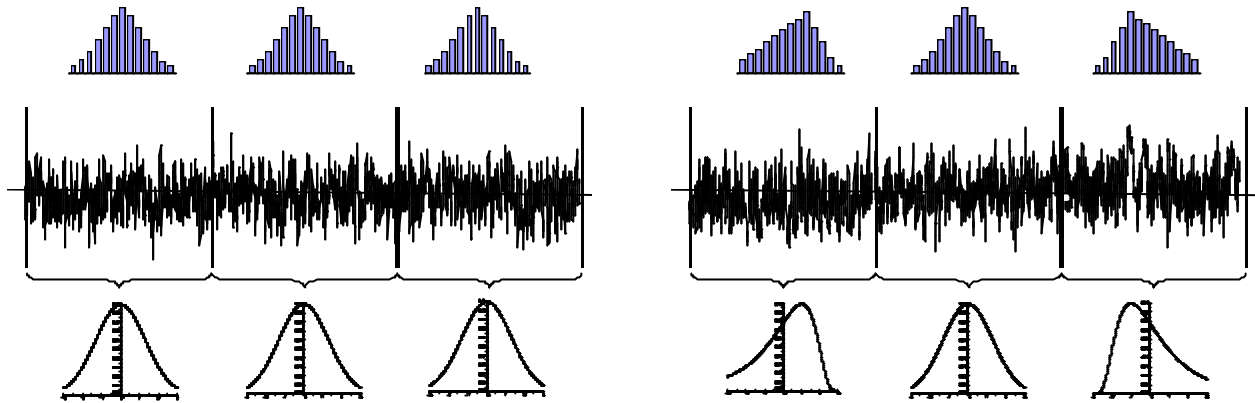
Cette définition toutefois peut être affaiblie : le processus est dit **stationnaire au second ordre** si

- la moyenne du processus est constante : $\mathbb{E}(Y_t) = m$ pour tout $t \in \mathbb{Z}$
- les autocovariances ne dépendent que de la différence entre les observations : $cov(X_t, X_s) = \gamma(|t-s|)$

Cette dernière propriété implique en particulier que la variance de Y_t est constante : $V(Y_t) = \sigma^2$.

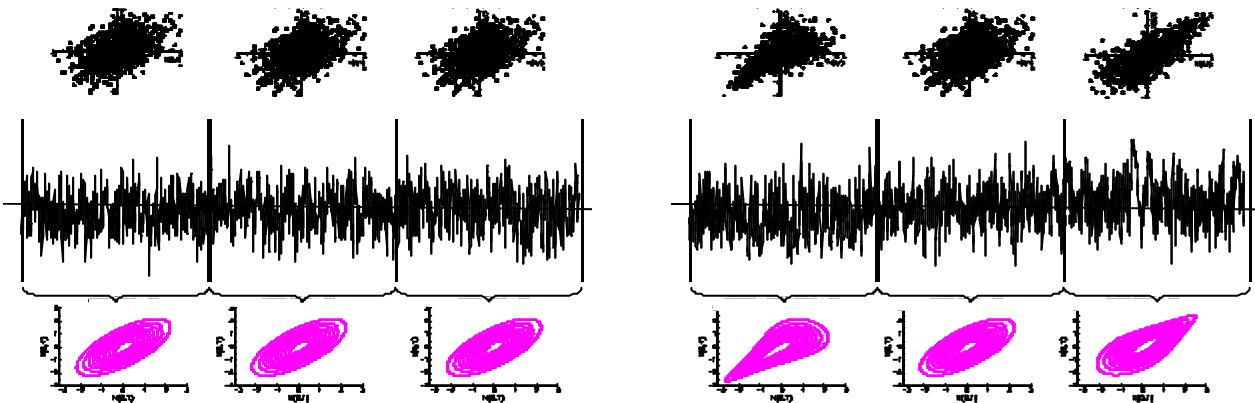
Remarque 2 Si l'on considère les lois marginales (à t fixé) du processus, la stationnarité (forte) signifie une stabilité de la loi marginale : la loi de Y_t et la loi de Y_s sont identiques pour $t \neq s$. La stationnarité du second ordre correspond uniquement à une stabilité des deux premiers moments : $\mathbb{E}(Y_t) = \mathbb{E}(Y_s)$ et $V(Y_t) = V(Y_s)$ pour $t \neq s$. Dans ce cas, rien n'empêche d'avoir des skewness et des kurtosis variables en fonction du temps. Le graphique ci-dessous à gauche

correspond à une stabilité au sens fort, et à droite, à une stationnarité au second ordre,



En particulier pour la stationnarité au sens fort, les moments d'ordre k , pour tout $k \in \mathbb{N}$, sont constants.

Remarque 3 Si l'on considère la dépendance temporelle, la stationnarité du second ordre suppose uniquement une stabilité de la corrélation (moment d'ordre 2) : $cov(X_t, X_{t+h}) = cov(X_s, X_{s+h})$. La stationnarité au sens fort est beaucoup plus forte que cette condition sur le moment d'ordre 2, puisqu'elle suppose une stabilité de toutes les lois jointes⁸ : en particulier, cette condition implique l'égalité en loi des couples (X_t, X_{t+h}) et (X_s, X_{s+h}) . Dans l'exemple ci-dessous, nous avons considéré deux processus dont les lois marginales reste constantes ($X_t \sim \mathcal{N}(0, 1)$ pour tout t), avec une stationnarité au sens fort à gauche (en particulier la loi (X_t, X_{t+h}) est égale à la loi de (X_s, X_{s+h})), et une stationnarité au second ordre à droite (en particulier, on a uniquement égalité des covariances $cov(X_t, X_{t+h}) = cov(X_s, X_{s+h})$)



La notion de stationnarité au second ordre, qui sera utilisée dans la première partie de ce cours, suppose uniquement une stabilité des deux premiers moments :

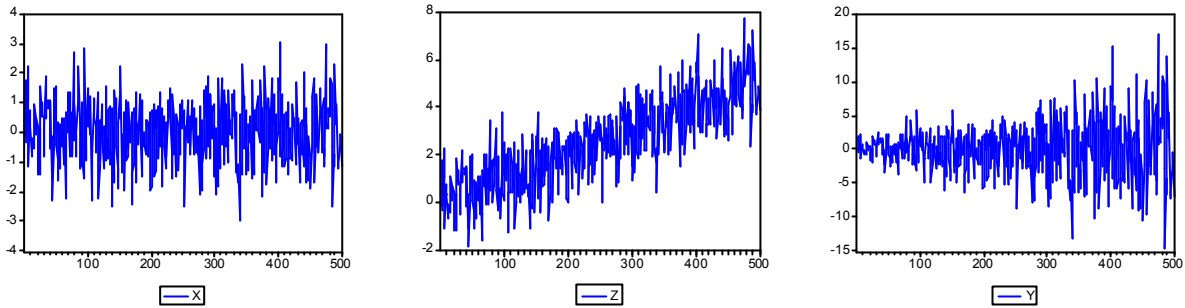
- la stationnarité au second ordre n'empêche pas une variation des moments d'ordres plus élevés (*asymétrie de la loi ou épaisseur des queue fonctions du temps*),
- la stabilité de la structure de dépendance entre X_t et X_{t+h} se résume à une stabilité du coefficient de corrélation (*ou de covariance*).

⁸Rappel : soient X_1 et X_2 de même loi, Y_1 et Y_2 de même loi, tels que $cov(X_1, Y_1) = cov(X_2, Y_2)$, alors on n'a pas égalité des lois jointes : $\mathcal{L}(X_1, Y_1) \neq \mathcal{L}(X_2, Y_2)$. En particulier, si X et Y suivent des lois normales $\mathcal{N}(\mu_X, \sigma_X^2)$ et $\mathcal{N}(\mu_Y, \sigma_Y^2)$ avec $corr(X, Y) = \rho$, alors on n'a pas nécessaire

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N \left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix} \right)$$

Un vecteur gaussien n'est pas uniquement un vecteur dont les lois marginales sont uniformes (*cf cours de probabilité*).

Remarque 4 Dans la pratique, on retrouve parfois des courbes aux allures sensiblement différentes,



La série (X_t) à gauche est (ou tout du moins semble) stationnaire, la série (Z_t) au centre est dite **non stationnaire en moyenne**, et la série (Y_t) à droite est dite **non stationnaire en variance**⁹.

L'exemple le plus simple de processus stationnaire est le bruit blanc. Toutefois, de la même façon qu'il est possible de définir deux notions de stationnarité, il existe deux sorte de bruit blanc. Le processus (ε_t) est un **bruit blanc faible** s'il existe σ^2 telle que

$$\begin{cases} \mathbb{E}(\varepsilon_t) = 0 \text{ pour tout } t \\ V(\varepsilon_t) = \mathbb{E}(\varepsilon_t^2) = \sigma^2 \text{ pour tout } t \\ \text{cov}(\varepsilon_t, \varepsilon_{t-h}) = \mathbb{E}(\varepsilon_t \varepsilon_{t-h}) = 0 \text{ pour tout } t, \text{ et pour tout } h \neq 0. \end{cases}$$

Aucune hypothèse d'indépendance n'est faite dans cette définition. Les variables aux différentes dates sont uniquement non corrélées (ce qui fera une différence importante, comme nous le verrons dans la partie sur les modèles ARCH). Cette hypothèse d'indépendance permet toutefois de définir un **bruit blanc fort**, i.e.

$$\begin{cases} \mathbb{E}(\varepsilon_t) = 0 \text{ et } V(\varepsilon_t) = \mathbb{E}(\varepsilon_t^2) = \sigma^2 \text{ (finie) pour tout } t \\ \mathcal{L}(\varepsilon_t) = \mathcal{L}(\varepsilon_{t-h}) \text{ pour tout } t, h \\ \varepsilon_t \text{ et } \varepsilon_{t-h} \text{ sont indépendantes pour tout } t, \text{ et pour tout } h \neq 0. \end{cases}$$

1.3.2 Propriété de Markov en temps discret

La théorie sur les chaînes de Markov (*en temps discret*) est également un élément important.

Cette propriété correspond à l'idée que l'on souhaite résumer l'information contenue dans les variables passées du processus par un nombre fini de variables (*les variables d'état*). Dans le cas le plus simple, on souhaite que les variables d'état soient des valeurs retardées du processus : toute l'information est contenue dans les k valeurs les plus récentes

$$\mathcal{L}(X_t | X_{t-1}, X_{t-2}, X_{t-3}, \dots) = \mathcal{L}(X_t | X_{t-1}, \dots, X_{t-k}),$$

qui peut se réécrire, à l'ordre 1,

$$(X_t | X_{t-1}, X_{t-2}, X_{t-3}, \dots) \stackrel{d}{=} (X_t | X_{t-1}).$$

Il est possible de montrer que cette relation est équivalente à

$$X_t = g(X_{t-1}, \varepsilon_t), \text{ où } (\varepsilon_t) \text{ est un bruit blanc.}$$

Toutefois, cette théorie, visant à chercher une fonction f telle que $X_t = f(X_{t-1}, \varepsilon_t)$ peut être difficile à implémenter. En économétrie, on cherche une relation du type $Y = g(X_1, \dots, X_n, \varepsilon)$, permettant d'expliquer une variable Y à l'aide de variables exogènes X_1, \dots, X_n . Cette fonction g étant a priori difficile à exhiber, la méthode la plus simple est de considérer le cas linéaire. De la même façon, la théorie des modèles *ARIMA* vise à expliquer X_t en fonction de son passé (*et éventuellement d'un bruit*), de manière linéaire.

⁹Dans ce cas particulier, il est possible d'utiliser la transformation dite de Box-Cox afin de réduire la variabilité de la série. On pose alors

$$Y_t = \begin{cases} (X_t^\lambda - 1) / \lambda & \text{si } \lambda \neq 0 \\ \log X_t & \text{sinon} \end{cases}$$

où le paramètre λ est estimé à l'aide du maximum de vraisemblance.

Remarque 5 Nous ne nous intéresserons, dans ce cours, que dans le cas où l'espace d'état est \mathbb{R} , c'est à dire que nous ne traiterons pas le cas des "chaînes de Markov" (où X_t prend ces valeurs dans un espace d'état fini ou dénombrable)

	t discret	t continu
espace d'état $\{i_1, \dots, i_n, \dots\}$	Cours sur les Chaînes de Markov	
espace d'état \mathbb{R}	Cours de Séries Temporelles	Cours de Calcul Stochastique

Les chaînes de Markov correspondent par exemple au cas où X_t est à valeurs dans un ensemble fini ($\{i_1, \dots, i_n, \dots\}$) ou dénombrable (\mathbb{N}) : par exemple les variables dichotomiques, où X_t vaut soit 0, soit 1.. Le calcul stochastique correspond au mouvement brownien, et aux processus de diffusion obtenus à partir du mouvement brownien. Le cas où le temps est continu et où les variables sont à valeurs dans \mathbb{N} (par exemple) correspond aux processus de comptage, aux processus de Poisson, à la théorie des files d'attente...etc.

1.4 Objectifs de l'études des séries temporelles

1.4.1 Description et modélisation

Le but est ici de déterminer les différentes composantes d'une série (X_t), en particulier, obtenir la série corrigée des variations saisonnières (*méthodes de désaisonnalisation*). Pour les séries stationnaires, on peut aussi chercher à modéliser la série à l'aide d'un modèle *ARMA*, par exemple dans le but de faire de la prévision.

1.4.2 Prévision

Sur la base d'observation X_1, \dots, X_T le but est de faire une prévision, à la date T , de la réalisation en $T + h$, notée $\hat{X}_T(h)$. Une première méthode est le lissage exponentiel, basé sur une formule de récurrence de la forme $\hat{X}_T(1) = \alpha X_t + (1 - \alpha) \hat{X}_{T-1}(h)$, où α , compris entre 0 et 1, est généralement choisi de façon à minimiser la somme des carrés des erreurs de prévision.

Dans le cas des modèles *ARMA*, de nombreuses relations existent afin de faire de la prévision, avec un intervalle de confiance. Nous verrons comment ces intervalles de confiance sont modifiés si une modélisation *ARCH* est retenue, ou du type mémoire longue.

Exemple 7 Quelques exemples de prévisions,

En 1977, Ken Olson, président du conseil d'administration, PDG et fondateur de la société Digital Equipment affirmait qu'« il n'y a aucune raison de vouloir posséder un ordinateur à la maison ».

Thomas Watson, président d'IBM, prédisait en 1943 : « Je crois que le marché mondial pourrait peut-être accueillir cinq ordinateurs. »

Une note de service de la Western Union qui affirmait, en 1876 : « Le téléphone a bien trop de défauts et de lacunes pour que nous le considérions sérieusement comme un moyen de communication. Cet appareil n'a pour ainsi dire aucune valeur à nos yeux. »

En 1895, Lord William Thomson Kelvin, président de la Société Royale des Sciences affirmait : « Il est impossible d'imaginer des machines volantes plus lourdes que l'air. »

« Le concept est intéressant et bien formulé, mais pour espérer avoir une note meilleure qu'un "C", encore faudrait-il que l'idée soit réalisable! » déclara un professeur de management de l'université de Yale en réponse à la proposition de Fred Smith de créer un service fiable de livraison de nuit (Smith fonda ensuite Federal Express Corp.)

Quelques jours avant le début de la crise, en 1929, Irving Fisher, Professeur d'économie à l'Université de Yale déclarait « Le marché de la Bourse semble avoir atteint un haut plateau permanent. »

« L'abdomen, la poitrine et le cerveau sont à jamais interdits à l'intrusion de la connaissance et de la chirurgie humaine. » selon Sir John Eric Ericksen, chirurgien Anglais, médecin personnel de la Reine Victoria, 1873.

1.4.3 Filtrage

Le lissage consiste à transformer une série de façon à détecter (pour éliminer ou au contraire conserver) certaines caractéristiques (*composante saisonnière, points abérants...*). Cette méthode permet également de détecter des ruptures au sein d'une série.

1.5 Conseils bibliographiques

Les principaux ouvrages servant de référence à ce cours sont les suivants,

- DROESBEKE, J.J., FICHET, B. & TASSI, P. (1995). '*Séries chronologiques - théorie et pratique des modèles ARIMA*', Economica
- GOURIEROUX, C. & MONFORT, A. (1995) '*Séries temporelles et modèles dynamiques*', Economica

Des compléments d'informations sur différents points abordés peuvent être trouvés également dans

- BOURBONNAIS, R. & TERRAZA, M. (1998). '*Analyse des séries temporelles en économie*', PUF
- BOX, G. & JENKINS, G.. (1970). '*Time Series analysis : forecasting and control*', Holden-Day [519.5 BOX]
- BROCKWELL, P.J. (1987) '*Time series : theory and methods*' Springer-Verlag
- COUTROT, B & DROESBEKE, J.J. (1995) '*Les Méthodes de prévision*' Presses Universitaires de France (Que sais-je ? 2157)
- DACUNHA-CASTELLE, D. & DUFLO, M. (1985). '*Probabilités et Statistiques - Tome 2 : Problèmes à temps mobile*' Masson
- HAMILTON, J. (1994). '*Time series analysis*', Princeton University Press [519.5 HAM]
- HARVEY, A.C. (1993) '*Time Series Models*' Cambridge: MIT Press [519.5 HAR]
- HYLLEBERG S. (1992), '*Modeling Seasonality*' Oxford University Press [330.115 MOD]
- LUTKEPOHL, H. (1991). '*Introduction to multiple time series analysis*' Springer-Verlag
- MELARD, G. (1990) '*Méthodes de prévision à court terme*'. Ellipses
- NERLOVE M, GREYER D.M, CARVALHO J.L. (1995). '*Analysis of Economic Time Series*' Academic Press.
- PINDYCK, R.S & RUBINFELD, L.D. (1984) '*Econometric models and economic forecasts*' McGraw-Hill [330.115 PI]

Des compléments et des documents au format pdf sont téléchargeables sur le site internet,

<http://www.crest.fr/pageperso/lfa/charpent/charpent.htm>

avec la version pdf de ce polycopié, des liens vers des notes de cours disponibles sur internet, et un certain nombre de bases de données qui peuvent être utilisées en guise d'exercices.

« *La nature semblait avoir sagement pourvu à ce que les sottises des hommes fussent passagères, et les livres les immortalisent.* » (Montesquieu, Les Lettres Persanes). Malgré les nombreuses relectures, il est possible qu'un certain nombre de coquilles, voire d'erreurs persistent. Merci de m'en tenir informé...

"Contrariwise," continued Tweedledee,
"if it was so, it might be; and if it were so,
it would be; but as it isn't, it ain't. That's logic."
Lewis Carroll.

2 Propriétés des processus univariés en temps discret

La pratique de l'analyse des séries temporelles vise à modéliser une (ou plusieurs, comme nous le développerons en fin de cours) série d'observations x_1, \dots, x_n par un processus aléatoire à temps discret, c'est à dire une suite (X_n) de variables aléatoires définies sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$, tel que l'on puisse penser que la série observée soit une réalisation du processus. En d'autres termes, x_1, \dots, x_n doit être obtenu comme tirage aléatoire de X_1, \dots, X_n suivant la probabilité \mathbb{P} , c'est à dire que se réalise un événement ω tel que $x_i = X_i(\omega)$ pour $i = 1, \dots, n$. Le but est alors, étant donnée une trajectoire x_1, \dots, x_n de reconstruire la dynamique du modèle sous-jacent, c'est à dire de comprendre la liaison entre X_i et son passé $X_{i-1}, X_{i-2}, \dots, X_1$.

2.1 Rappels sur les martingales à temps discret

Un processus (à temps discret) sur un espace $(\Omega, \mathcal{A}, \mathbb{P})$ est une suite de variables aléatoires $(X_t(\omega), t \in \mathbb{N})$, à valeurs dans un espace mesuré (E, \mathcal{E}) (dans le cas qui nous intéresse, $E = \mathbb{R}$). On peut considérer le processus comme la variable aléatoire $X(t, \omega)$, défini sur l'espace produit $\Omega \times \mathbb{N}$ muni de la tribu produit.

Définition 1 Une **filtration** $\{\mathcal{F}_t, t \in \mathbb{N}\}$ est la donnée d'une suite croissante (au sens de l'inclusion) de sous-tribus de \mathcal{A} . On posera $\mathcal{F}_\infty = \sup\{\mathcal{F}_t, t \in \mathbb{N}\}$: il s'agit de la plus petite tribu qui contienne toutes les \mathcal{F}_t .

La filtration la plus usuelle est obtenue de la façon suivante : on observe une suite (X_t) de variables aléatoires, et on considère $\mathcal{F}_t = \sigma(X_0, \dots, X_t)$, qui est la plus petite tribu qui rende mesurable les variables (X_0, \dots, X_n) . On appellera filtration naturelle cette filtration, et on la notera (\mathcal{F}_t^X) .

On dira que (X_t) est adaptée à la filtration (\mathcal{F}_t) si pour tout t , X_t est \mathcal{F}_t -mesurable. La filtration naturelle est la plus petite filtration par rapport à laquelle le processus soit adapté. On dira que le processus (X_t) est prévisible si pour tout $t \geq 1$, X_t est \mathcal{F}_{t-1}^X -mesurable.

Définition 2 Le processus $\{X_t, t \in \mathbb{N}\}$ muni de la filtration $\{\mathcal{F}_t, t \in \mathbb{N}\}$ tel que pour tout t , X_t soit intégrable. On dira que (X_t) est une **martingale** si et seulement si, pour tout t , $\mathbb{E}(X_{t+1}|\mathcal{F}_t) = X_t$ presque sûrement.

Remarque 6 Si pour tout t , $\mathbb{E}(X_{t+1}|\mathcal{F}_t) \geq X_t$ presque sûrement, on dira que (X_t) est une sous-martingale, et si pour tout t , $\mathbb{E}(X_{t+1}|\mathcal{F}_t) \leq X_t$ presque sûrement, on dira que (X_t) est une sur-martingale.

Si (X_t) est une (\mathcal{F}_t) -martingale, alors pour tout $h \geq 0$, $\mathbb{E}(X_{t+h}|\mathcal{F}_t) = X_t$. De plus, si la martingale est de carré intégrable, les accroissements sont orthogonaux : si $\Delta X_t = X_t - X_{t-1}$, pour $s \neq t$, $\mathbb{E}(\Delta X_t \Delta X_s) = 0$. Une des conséquences est que, pour tout $h \geq 0$

$$\mathbb{E}\left([X_{t+h} - X_t]^2\right) = \sum_{i=1}^h \mathbb{E}(\Delta X_{t+i}^2).$$

2.2 Rappels sur les Chaînes de Markov

Définition 3 Le processus $\{X_t, t \in \mathbb{N}\}$ est une **chaîne de Markov d'ordre 1** si et seulement si, pour tout t ,

$$\mathcal{L}(X_t | X_{t-1}, X_{t-2}, X_{t-3}, \dots) = \mathcal{L}(X_t | X_{t-1}).$$

Autrement dit, compte tenu de la trajectoire $(X_{T-1} = x_{T-1}, X_{T-2} = x_{T-2}, \dots)$ d'un processus (X_t) , la loi de X_T à l'instant T est entièrement déterminée par le fait que la valeur en $T-1$ soit x_{T-1} .

Théorème 1 Le processus $\{X_t, t \in \mathbb{N}\}$ est une **chaîne de Markov d'ordre 1** si et seulement s'il existe une fonction $g(\cdot)$ mesurable et un processus ε_t tel que $X_t = g(X_{t-1}, \varepsilon_t)$ - avec (ε_t) une suite de variables aléatoires, indépendantes et de même loi.

Lorsque l'application g ne dépend pas de t , la chaîne de Markov est dite **homogène**.

Exemple 8 Les processus AR(1) : $X_t = \alpha + \beta X_{t-1} + \varepsilon_t$, où (ε_t) est un bruit blanc, sont markoviens.

Exemple 9 En particulier, les processus de la forme $X_t = X_{t-1} + \varepsilon_t$ correspondent à une **marche aléatoire** :

- si $X_0 \in \mathbb{Z}$ et $\mathbb{P}(\varepsilon_t = -1) = \mathbb{P}(\varepsilon_t = +1) = 1/2$, on obtient la marche aléatoire symétrique sur \mathbb{Z} (jeu du pile ou face),

- si ε_t suit une loi normale centrée, on obtient une discrétisation du mouvement brownien.

2.3 Notions de 'processus stationnaire' et de 'processus non-stationnaire'

Définition 4 Un processus (X_t) est **stationnaire au second ordre** si

(i) pour tout t , $\mathbb{E}(X_t^2) < +\infty$,

(ii) pour tout t , $\mathbb{E}(X_t) = \mu$, constante indépendante de t ,

(iii) pour tout t et pour tout h , $\text{cov}(X_t, X_{t+h}) = \mathbb{E}([X_t - \mu][X_{t+h} - \mu]) = \gamma(h)$, indépendante de t .

Définition 5 La fonction $\gamma(\cdot)$ sera appelée **fonction d'autocovariance**

On peut montrer aisément que $\gamma(\cdot)$ est une fonction paire, au sens où $\gamma(h) = \gamma(-h)$ pour tout h , et que la variance $V(X_t)$ est constante, indépendante de t , $V(X_t) = \gamma(0)$.

Propriété 1 Si $(X_t, t \in \mathbb{Z})$ est un processus stationnaire, et si $(a_i, i \in \mathbb{Z})$ est une suite de réels absolument convergente, i.e. $\sum_{i \in \mathbb{Z}} |a_i| < +\infty$, alors, le processus (Y_t) défini par

$$Y_t = \sum_{i \in \mathbb{Z}} a_i X_{t-i}, \text{ pour tout } t \in \mathbb{Z},$$

est un processus stationnaire.

Corollaire 2 En particulier, si $(a_i, i \in \mathbb{Z})$ est une suite de réels finie, la suite Y_t est stationnaire. Par exemple, si $a_0 = a_1 = 1/2$, et $a_i = 0$ pour $i \notin \{0, 1\}$:

$$Y_t = \frac{1}{2} (X_t + X_{t-1}),$$

est stationnaire dès lors que (X_t) est stationnaire. De même pour $Y_t = X_t - X_{t-1}$.

Définition 6 Un processus (X_t) est **stationnaire au sens fort** si pour tous t_1, \dots, t_n et h on a l'égalité en loi

$$(X_{t_1}, \dots, X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, \dots, X_{t_n+h}).$$

Remarque 7 Cette notion revient à dire que la loi temporelle est invariante en temps. Cette stationnarité est beaucoup plus forte que la stationnarité du second ordre, puisqu'on ne recherche pas la stabilité de la loi, mais seulement la stabilité des deux premiers moments.

Définition 7 On appelle **bruit blanc** (parfois appelé bruit blanc faible) un processus (ε_t) stationnaire dont les autocovariances sont toutes nulles : $\gamma(h) = 0$ pour $h \neq 0$.

Remarque 8 Nous avons vu dans la partie précédente que (X_t) est une martingale si et seulement si, pour tout t , $\mathbb{E}(X_{t+1} | X_t, X_{t-1}, \dots) = X_t$ pour tout t , ou, de façon équivalente, cela signifie que $X_{t+1} = X_t + \varepsilon_t$ avec $\mathbb{E}(\varepsilon_{t+1} | \varepsilon_t, \varepsilon_{t-1}, \dots) = 0$ pour tout t . Cette notion est plus contraignante que celle de "marche aléatoire" : en effet, la propriété de martingale implique l'indépendance des accroissements (ε_t) alors que la définition de la marche aléatoire n'implique que la nullité des corrélations des accroissements.

Définition 8 Un processus stationnaire (X_t) sera dit **ergodique** si pour tout $p \in \mathbb{N}^*$, et pour tout fonction borélienne de \mathbb{R}^p à valeurs dans \mathbb{R} , on a

$$\frac{1}{N} \sum_{i=1}^N f(X_{i+1}, X_{i+2}, \dots, X_{i+p}) \rightarrow \mathbb{E}(f(X_1, X_2, \dots, X_p)), \text{ quand } N \rightarrow \infty,$$

qui peut être vu simplement comme une généralisation de la loi de grand nombre.

La notion de stationnarité (*faible, ou au second ordre*) se définit par une invariance des moments d'ordre 1 et 2 au cours du temps. Par opposition, on dira qu'une série est **non-stationnaire** si elle n'est pas stationnaire. On peut noter que la classe des processus non-stationnaire est alors relativement vaste, et surtout hétérogène : il existe différentes sources de non-stationnarité, et à chaque origine de non-stationnarité est associée une méthode propre de stationnarisation. Nelson et Plosser ont retenu, en 1982, deux classes de processus non-stationnaires : les processus *TS* (*trend stationary*) et les processus *DS* (*difference stationary*). Les premiers correspondent à une non-stationnarité de type déterministe, alors que les seconds correspondent à une non-stationnarité de type stochastique.

Définition 9 (X_t) est un processus **non-stationnaire TS** s'il peut s'écrire sous la forme $X_t = f(t) + Z_t$ où $f(t)$ est une fonction (déterministe) du temps, et (Z_t) est un processus stationnaire.

L'exemple le plus simple est celui de la tendance linéaire bruitée : $X_t = \alpha + \beta t + \varepsilon_t$. Ce processus est en effet non-stationnaire puisque son espérance vaut $\alpha + \beta t$ à la date t , et donc, dépend de t . Une des propriétés importantes de ce type de processus est qu'il n'y a pas persistance des chocs : l'influence d'un choc subit à un instant τ aura tendance à s'estomper au cours du temps, et la variable rejoint alors sa dynamique de long-terme, déterminée par $f(t)$.

Définition 10 (X_t) est un processus **non-stationnaire DS** - ou intégré d'ordre d , noté $I(d)$ - si le processus obtenu après d différenciation est stationnaire : $Z_t = \Delta^d X_t = (1 - L)^d X_t$ est stationnaire

Comme nous le verrons par la suite, le fait qu'il faille différencier d fois, c'est à dire multiplier par $(1 - L)^d$, polynôme de l'opérateur retard L , revient à chercher la présence de **racines unité** : si le processus $\Phi(L) X_t$ est stationnaire, si 1 est une racine du polynôme Φ , alors (X_t) sera non-stationnaire. C'est pour cela que la plupart des tests de non-stationnarité sont des tests de détection de racine unité.

2.4 Fonction d'autocovariance et densité spectrale

2.4.1 Autocovariance et autocorrélation

Définition 11 Pour une série stationnaire (X_t) , on définit la **fonction d'autocovariance**, pour tout t , par

$$h \mapsto \gamma_X(h) = \text{cov}(X_t, X_{t-h}) = \mathbb{E}(X_t X_{t-h}) - \mathbb{E}(X_t) \mathbb{E}(X_{t-h}).$$

Définition 12 Pour une série stationnaire (X_t) , on définit la **fonction d'autocorrélation**, pour tout t , par

$$h \mapsto \rho_X(h) = \text{corr}(X_t, X_{t-h}) = \frac{\text{cov}(X_t, X_{t-h})}{\sqrt{V(X_t)}\sqrt{V(X_{t-h})}} = \frac{\gamma_X(h)}{\gamma_X(0)}.$$

Cette fonction $\rho_X(\cdot)$ est à valeurs dans $[-1, +1]$, et $\rho_X(0) = 1$.

Définition 13 Un processus (ε_t) sera appelé **bruit blanc** (*faible*) s'il est stationnaire, centré et non-autocorrélé :

$$\mathbb{E}(\varepsilon_t) = 0, \quad V(\varepsilon_t) = \sigma^2 \quad \text{et} \quad \rho_\varepsilon(h) = 0 \quad \text{pour} \quad h \neq 0.$$

On parlera de bruit blanc fort s'il est indépendant et identiquement distribué (*i.i.d.*) : la notion d'indépendance est plus forte que la nullité des autocorrélations, et le fait que le processus soit identiquement distribué est plus fort que la stabilité des deux premiers moments.

Exemple 10 Processus *MA*(1) : $X_t = \varepsilon_t + \theta\varepsilon_{t-1}$ où (ε_t) est un bruit blanc centré de variance σ^2 ,

$$\begin{cases} \gamma(0) = [1 + \theta^2] \sigma^2 \\ \gamma(1) = \theta\sigma^2 \\ \gamma(h) = 0 \text{ si } |h| \geq 2 \end{cases}, \quad \text{soit } \rho(1) = \frac{\theta}{1 + \theta^2} \text{ et } \rho(h) = 0 \text{ pour } |h| \geq 2.$$

2.4.2 Densité spectrale

L'idée ici est que les coefficients d'autocovariance d'une série stationnaire correspondent aux coefficients de Fourier d'une mesure positive, appelée mesure spectrale du processus. Il est possible de montrer que cette mesure spectrale admet une densité, dite spectrale, par rapport à la mesure de Lebesgue sur $[-\pi, \pi]$, que nous noterons f_X . Dans le cas où la série des autocovariance est absolument convergente, la densité spectrale est alors définie comme la transformée de Fourier des coefficients d'autocovariance (*dans le cas où la somme des $|\gamma_X(h)|$ tend vers l'infini, la somme est à prendre au sens de L^2*) : comme l'ont montré Cramér, Kolmogorov, ou encore Wiener, on les résultats suivants,

(i) la suite des fonctions d'autocovariance $\gamma_X(h)$ d'un processus stationnaire peut être écrit sous la forme

$$\gamma_X(h) = \int_{-\pi}^{+\pi} \exp(i\omega h) dF_X(\omega),$$

où $F_X(\omega)/\gamma_X(0)$ est une fonction de répartition,

(ii) tout processus stationnaire peut se mettre sous la forme $X_t = \int_{-\pi}^{+\pi} \exp(i\omega t) dz(\omega)$ où $z(\omega)$ est une fonction aléatoire, complexe, à accroissements non corrélés. Cette représentation est appelée **représentation de Cramér**.

Définition 14 Soit (X_t) un processus stationnaire de fonction d'autocovariance $\gamma_X(\cdot)$, la **densité spectrale** de (X_t) s'écrit

$$f_X(\omega) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_X(h) \exp(i\omega h).$$

Propriété 2 Réciproquement, si $f_X(\cdot)$ est la densité spectrale de (X_t) alors

$$\gamma_X(h) = \int_{-\pi}^{+\pi} f_X(\omega) \exp(i\omega h) d\omega.$$

Exemple 11 Un bruit blanc (ε_t) est caractérisé par

$$\begin{cases} \gamma_\varepsilon(0) = V(\varepsilon_t) = \sigma^2 \\ \gamma_\varepsilon(h) = 0, \text{ pour } h \neq 0, \end{cases}$$

Alors sa densité spectrale est donnée par

$$f_\varepsilon(\omega) = \frac{\sigma^2}{2\pi} (= \text{constante}).$$

Propriété 3 Si la densité spectrale d'une série (Z_t) est constante, alors (Z_t) est un bruit blanc.

Preuve. En effet

$$\gamma_Z(h) = \int_{-\pi}^{+\pi} f_Z(\omega) \exp(i\omega h) d\omega = K \underbrace{\int_{-\pi}^{+\pi} \exp(i\omega h) d\omega}_{=0 \text{ sauf si } h=0}$$

■

Cette nullité de la fonction d'autocorrélation est donc une caractéristique du bruit blanc.

Propriété 4 Si X_t est une moyenne mobile,

$$X_t = \sum_{k \in \mathbb{Z}} a_k \varepsilon_{t-k}, \text{ où } (\varepsilon_t) \text{ est un bruit blanc } BB(0, \sigma^2),$$

avec $\sum |a_j| < +\infty$. Si on considère $Y_t = \sum_{j \in \mathbb{Z}} \beta_j X_{t-j}$ alors, on a la relation suivante

$$f_Y(\omega) = f_X(\omega) \cdot \left| \sum_{j \in \mathbb{Z}} \beta_j e^{i\omega j} \right|^2.$$

Exemple 12 $Y_t = X_t - \phi X_{t-1}$ où $|\phi| < 1$, alors $f_Y(\omega) = f_X(\omega) |1 + \phi e^{i\omega}|^2$.

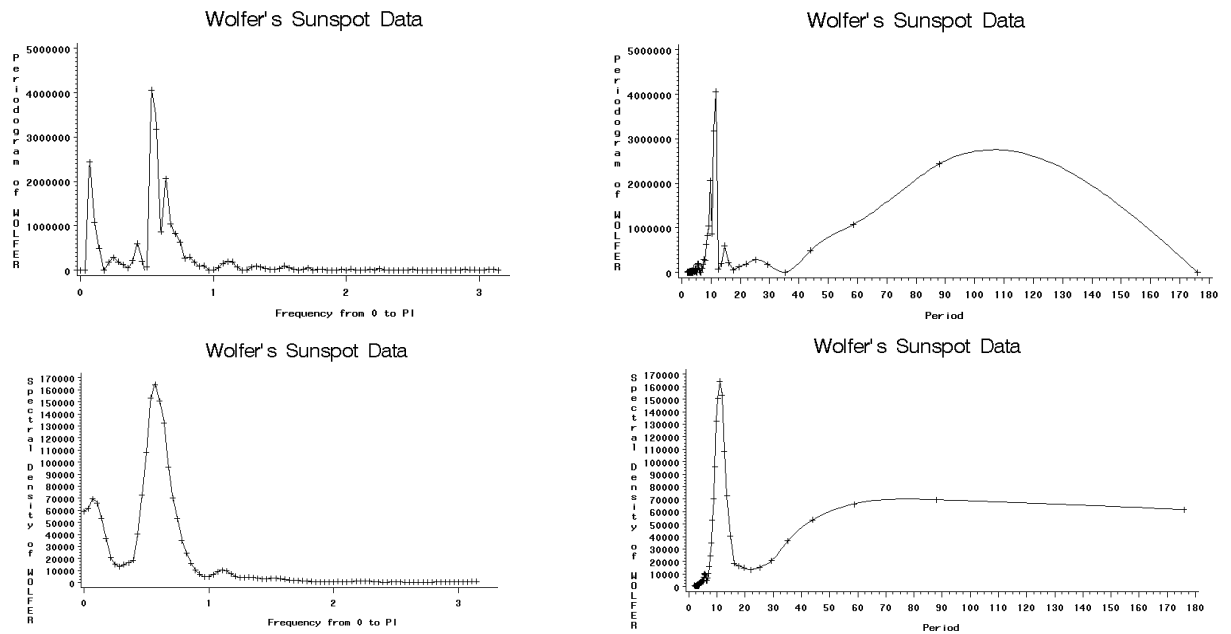
Remarque 9 La densité spectrale d'un processus peut être estimée sous SAS, à l'aide de la procédure **spectra**. Nous allons reprendre ici l'exemple de taches solaires observée de 1770 à 1869, tel que l'avait fait Wolfer

La procédure suivante permet d'obtenir le périodogramme de cette série

```

title "Wolfer/s Sunspot Data";
proc spectra data = sunspot out = b p s adjmean whitetest;
var wolfer;
weights 1 2 3 4 3 2 1;
run;

```



avec, respectivement, en haut, le périodogramme (P) en fonction de la fréquence (à gauche), et de la période (à droite), et en bas, la densité spectrale estimée (S) en fonction de la fréquence (à gauche), et de la période (à droite). Sur ces données, on observe un pic correspondant à une périodicité de 11 ans. Le graphique ci-dessous¹⁰ correspond au périodogramme obtenu sur différentes périodes d'observation,

2.4.3 Estimation de la fonction d'autocorrélation

Considérons un ensemble d'observations X_1, \dots, X_T .

La **moyenne empirique** est donnée par

$$\bar{X}_T = \frac{1}{T} \sum_{t=1}^T X_t.$$

¹⁰tiré de The analysis of economic time series, de Davis (1963) page 318.

La **fonction d'autocovariance empirique** est donnée par

$$\hat{\gamma}_T(h) = \frac{1}{T-h} \sum_{t=1}^{T-h} (X_t - \bar{X}_T)(X_{t-h} - \bar{X}_T),$$

et la **fonction d'autocorrélation empirique** est donnée par

$$\hat{\rho}_T(h) = \frac{\hat{\gamma}_T(h)}{\hat{\gamma}_T(0)}.$$

Si ces estimateurs sont biaisés (à *distance finie*), ils sont malgré tout asymptotiquement sans biais.

Propriété 5 *Les moments empiriques convergent vers les moments théoriques : $\bar{X}_T \rightarrow m$, $\hat{\gamma}_T(h) \rightarrow \gamma(h)$ et $\hat{\rho}_T(h) \rightarrow \rho(h)$ quand $T \rightarrow \infty$.*

En fait, comme nous le verrons par la suite, nous avons même normalité asymptotique des moments empiriques.

Remarque 10 *Bien que ces fonctions soient définies pour tout h tel que $-T < h < T$, la fonction d'autocovariance empirique fournit un estimateur très pauvre de $\gamma(h)$ pour des valeurs h proches de n . A titre indicatif, Box et Jenkins recommandent de n'utiliser ces quantités que si $T > 50$ et $h \leq T/4$. "In practice, to obtain usefull estimate of the autocorrelation function, we need at least 50 observations, and the estimated autocorrelations r_k could be calculated for $k = 1, \dots, K$ where K was not larger than, say, $T/4$ ".*

Afin, par exemple, de faire de la selection de modèles, il est important de pouvoir dire si les autocovariances empiriques sont significativement non nulles. Il est alors possible d'utiliser le résultat suivant

Propriété 6 *Si (X_t) est un processus linéaire, au sens où il satisfait $X_t = \sum_{j \in \mathbb{Z}} \phi_j \varepsilon_{t-j}$ où (ε_t) est une suite de variables i.i.d. centrées, telle que $\mathbb{E}(\varepsilon_t^4) = \eta \mathbb{E}(\varepsilon_t^2)^2 < +\infty$, où les ϕ_j définissent une série absolument convergente, et où η est une constante positive, alors, on a la formule dite de Bartlett,*

$$\lim_{T \rightarrow \infty} \text{Cov}(\hat{\gamma}_T(h), \hat{\gamma}_T(k)) = \eta \gamma(h) \gamma(k) + \sum_{i=-\infty}^{+\infty} \gamma(i) \gamma(i+k-h) + \gamma(i+k) \gamma(i-h).$$

Preuve. Brockwell et Davis (1991) page 226. ■

Ce théorème n'est, en théorie, valable que pour un bruit blanc fort. On peut également montrer que ces autocorrélation vérifient une propriété encore plus forte,

Propriété 7 *Si (X_t) est un processus linéaire, au sens où il satisfait $X_t = \sum_{j \in \mathbb{Z}} \phi_j \varepsilon_{t-j}$ où (ε_t) est une suite de variables i.i.d. centrées, telle que $\mathbb{E}(\varepsilon_t^4) = \eta \mathbb{E}(\varepsilon_t^2)^2 < +\infty$, et $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, et où les ϕ_j définissent une série absolument convergente, et où η est une constante positive, alors, on a, pour tout $p \geq 0$,*

$$\sqrt{n} \begin{pmatrix} \hat{\gamma}_T(0) \\ \vdots \\ \hat{\gamma}_T(p) \end{pmatrix} \rightarrow \mathcal{N} \left(\begin{pmatrix} \gamma(0) \\ \vdots \\ \gamma(p) \end{pmatrix}, V \right),$$

où V est la matrice de variance-covariance définie par

$$V = \left[\eta \gamma(h) \gamma(k) + \sum_{i=-\infty}^{+\infty} \gamma(i) \gamma(i+k-h) + \gamma(i+k) \gamma(i-h) \right]_{h,k=0,\dots,p}.$$

Preuve. Brockwell et Davis (1991) page 227. ■

2.4.4 Estimation de la densité spectrale

Le périodogramme est observations est défini comme le module au carré de la transformé de Fourier discrète des observations, i.e.

$$I_T(x) = \frac{1}{2\pi T} \left| \sum_{t=1}^T X_t \exp(itx) \right|^2 = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \hat{\gamma}_T(h) \exp(i\omega x).$$

Le plus souvent, on estime le périodogramme aux fréquences de Fourier, i.e. $x_k = 2k\pi/T$ pour $k = 1, \dots, T$, noté $I_{T,k}$. Sous des hypothèses de régularité de la densité spectrale, le périodogramme est un estimateur asymptotiquement sans biais de la densité spectrale. Mais il n'est pas consistant (on ne peut estimer que les T premier $\gamma(h)$ intervenant dans la définition du périodogramme à partir de T observations).

Exemple 13 Dans le cas d'un processus i.i.d. gaussien, évalué aux fréquences de Fourier de $]0, \pi[$ forme une suite de variables indépendantes, et identiquement distribuées, suivant une loi du χ^2 , centré, à deux degrés de liberté.

Exemple 14 Pour les processus dit à mémoire longue, la densité spectrale s'exprime sous la forme

$$f(x) = |1 - \exp(ix)|^{-2d} f_*(x),$$

où f_* est une fonction positive. Les valeurs du périodogramme sont asymptotiquement biaisées, et asymptotiquement corrélées. Le fait que cette fonction ait un pôle (ici en 0) est d'ailleurs une caractérisation de la mémoire longue.

Cette densité spectrale permet d'obtenir un grand nombre de résultat. Par exemple, il est possible d'estimer directement la variance du processus d'innovation¹¹, en utilisant la formule dite de Kolmogorov,

$$\sigma^2 = 2\pi \exp\left(\frac{1}{2\pi} \int_0^{2\pi} \log f_X(x) dx\right).$$

Un estimateur de cette variance est alors

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T \log I_{T,k}.$$

Dans le cas des processus à mémoire longue, la densité spectrale est de la forme $f_X(x) \sim Cx^{-2d}$. Un estimateur non paramétrique de d peut être obtenu en régressant localement le log-périodogramme dans un voisinage de la fréquence nulle. On appelle alors estimateur GPH

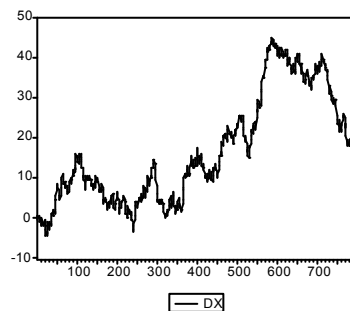
$$\hat{d} = \left(\sum_{k=1}^{m_T} L_{T,k}^2 \right)^{-1} \sum_{k=0}^{m_T} L_{T,n} \cdot \log L_{T,k} \text{ où } L_{T,k} = -2 \log |x_k| + \frac{2}{m_T} \sum_{j=1}^{m_T} \log I_{T,j},$$

et où m_T est une suite d'entiers positifs telle que $m_T \rightarrow 0 = \infty$ et $m_T/T \rightarrow 0$ quand $T \rightarrow \infty$.

2.5 Lien entre processus en temps continu et en temps discret

Définition 15 Un **mouvement brownien** W_t est un processus stochastique, défini pour $t \in \mathbb{R}^+$, tel que $W_0 = 0$ et tel que, quelles que soient les dates $t_1 < t_2 < \dots < t_k$, les variations du processus $W_{t_2} - W_{t_1}, W_{t_3} - W_{t_2}, \dots, W_{t_k} - W_{t_{k-1}}$ sont indépendantes, avec $\mathbb{E}(W_{t_i} - W_{t_j}) = 0$ et $V(W_{t_i} - W_{t_j}) = \sigma^2(t_i - t_j)$. De plus, les variations du processus entre deux dates t_i et t_j (telles que $t_i < t_j$) sont normalement distribuées $W_{t_i} - W_{t_j} \sim \mathcal{N}(0, \sigma^2(t_i - t_j))$.

Dans le cas où $\sigma^2 = 1$, on parlera de mouvement brownien standard. De plus, W_t est continu en t , sans être dérivable : bien que le processus soit continu, les variations ne sont pas bornées. Pour visualiser un mouvement brownien il suffit de considérer une marche aléatoire continue : on considère une marche aléatoire discrète ($X_t = X_{t-1} + \varepsilon_t$ où $\varepsilon_t \sim \mathcal{N}(0, 1)$), pour laquelle on diminue les intervalles temporels entre deux dates consécutives,



Propriété 8 Soit X_1, X_2, \dots, X_T un échantillon i.i.d., centré, de variance σ^2 . Soit $[\cdot]$ la partie entière au sens où $[x] \leq x < [x] + 1$ et $[x] \in \mathbb{Z}$, alors pour tout $0 < r < 1$,

$$\frac{1}{\sqrt{\sqrt{[rT]}}} \sum_{t=1}^{[rT]} X_t \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

¹¹Cette notion sera définie en détails par la suite.

Ce résultat est parfois appelé "Théorème Centrale Limite Fonctionnel". Notons $\overline{X}_T^{(r)}$ la variable construite à partir des $[rT]$ premières observations par

$$\overline{X}_T^{(r)} = \frac{1}{T} \sum_{t=1}^{[rT]} X_t,$$

du résultat précédent, il en découle que

$$\frac{\sqrt{T}\overline{X}_T^{(r)}}{\sigma} \xrightarrow{\mathcal{L}} \mathcal{N}(0, r) \text{ ou encore } \frac{\sqrt{T}(\overline{X}_T^{(r_2)} - \overline{X}_T^{(r_1)})}{\sigma} \xrightarrow{\mathcal{L}} \mathcal{N}(0, r_2 - r_1),$$

pour $r_1 < r_2$. Ceci permet de montrer que la suite des $\sqrt{T}\overline{X}_T^{(\cdot)}/\sigma$ est asymptotiquement distribuée comme un mouvement brownien, au sens où

$$\frac{\sqrt{T}\overline{X}_T^{(\cdot)}}{\sigma} \xrightarrow{\mathcal{L}} W.$$

Ce type de résultat est alors très utile pour obtenir des résultats analytiques sur les processus intégrés.

Considérons par exemple, une marche aléatoire définie par $X_t = X_{t-1} + \varepsilon_t$ où ε_t est un bruit blanc de variance σ^2 , soit $X_t = \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t$ pour tout t , avec la convention $X_0 = 0$. Notons $\overline{X}_T^{(r)}$ la variable construite à partir des $[rT]$ premières observations par

$$\overline{X}_T^{(r)} = \frac{1}{T} \sum_{t=1}^{[rT]} X_t = \frac{1}{T} (\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_i), \text{ où } \frac{i-1}{T} \leq r < \frac{i}{T},$$

on a alors

$$\sqrt{T} \int_0^1 \overline{X}_T^{(r)} dr = T^{-3/2} \sum_{t=1}^T x_{t-1},$$

et d'après le théorème central limite fonctionnel, $\sqrt{T}\overline{X}_T^{(\cdot)} \xrightarrow{\mathcal{L}} \sigma W$. On obtient alors le résultat suivant

$$T^{-3/2} \sum_{t=1}^T X_{t-1} \xrightarrow{T \rightarrow \infty} \sigma \int_0^1 W_s ds.$$

De façon analogue, on peut montrer que

$$T^{-2} \sum_{t=1}^T (X_{t-1})^2 \xrightarrow{T \rightarrow \infty} \sigma^2 \int_0^1 (W_s)^2 ds.$$

Ces résultats seront utilisés en particulier pour la détermination des propriétés asymptotiques des estimateurs obtenus à partir de séries intégrées.

La construction de l'**intégrale stochastique** s'obtient d'ailleurs comme passage à la limite sur des processus à temps discret¹². Considérons un découpage en T subdivisions de l'intervalle de temps $[0, 1]$: soit $s_t = t/T$ pour $t = 0, 1, \dots, T$. Considérons ici (X_{s_t}) , noté (Y_t) , un processus défini pour $t = 0, 1, \dots, T$. On appellera variation quadratique de la série chronologique (Y) la série chronologique définie par

$$\langle Y \rangle_t = \sum_{j=1}^t [Y_j - Y_{j-1}]^2 \text{ pour } t = 0, 1, \dots, T$$

La variation quadratique du mouvement brownien standard (W_t) est obtenu comme passage à la limite

$$\langle W \rangle_t = \lim_{T \rightarrow \infty} \sum_{j=1}^t [W_j - W_{j-1}]^2 = t$$

¹²De façon plus simple, l'intégrale d'une fonction aléatoire par rapport à une mesure déterministe de définie déjà comme une limite : soit A_t un processus en temps continu, et considérons un découpage en T subdivisions de l'intervalle de temps $[0, 1]$: soit $\tau_s = s/T$ pour $s = 0, 1, \dots, T$. Considérons ici X_{τ_s} , noté Y_s , le processus défini pour $s = 0, 1, \dots, T$, par $Y_s = A_t \mathbb{I}(Tt \leq s < T(t+1))$, alors

$$\int_0^t A_s ds = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{j=1}^s Y_s$$

De la même façon, l'intégrale stochastique se définit en temps discret par

$$Z_t = \int_0^t X_s dW_s = \sum_{i=1}^t \mathcal{X}_i [W_j - W_{j-1}]$$

où (X_s) est un processus discret : $X_s = \mathcal{X}_i$ pour $(i-1)/T \leq s < i/T$, puis par passage à la limite, s'étend aux processus en temps continu.

Aussi, tous les modèles financiers en temps continu ont un analogue en temps discret. Mais si les modèles en temps continu sont autant utilisés, c'est principalement parce que le calcul stochastique et la formule d'Ito permet d'analyser les problèmes de façon élégante et relativement rapide.

Un processus suivant l'équation stochastique

$$dY_t = f(t, Y_t) dt + g(t, Y_t) dW_t \text{ ou } Y_t = Y_0 + \int_0^t f(s, Y_s) ds + \int_0^t g(s, Y_s) dW_s$$

peut être assimilé à un processus en temps discret vérifiant l'équation (*approximation d'Euler*)

$$Y_{t+1} - Y_t = f(t, Y_t) + g(t, Y_t) [W_{t+1} - W_t] = f(t, Y_t) + g(t, Y_t) \varepsilon_t$$

où (ε_t) est un bruit blanc gaussien, de variance 1.

Remarque 11 Réciproquement, en reprenant un exemple de Nelson (1990), un modèle à temps discret de type *GARCH* (1,1) – *M* (multivarié), défini par

$$\begin{cases} Y_t = Y_{t-1} + f(\sigma_t^2) + \sigma_t \varepsilon_t \\ \sigma_{t+1}^2 = \omega + \sigma_t^2 (\beta + \alpha \varepsilon_t) \end{cases}$$

où (ε_t) est un bruit blanc gaussien, est l'analogue en temps discret de l'équation de diffusion

$$\begin{cases} dY_t = f(\sigma_t^2) dt + \sigma_t dW_t^1 \\ d\sigma_t^2 = (\omega - \theta \sigma_t^2) dt + \alpha \sigma_t^2 dW_t^2 \end{cases}$$

où (W_t^1) et (W_t^2) sont deux mouvements browniens centrés, réduits et indépendants.

La décomposition tendance-cycle

L'analyse des séries temporelles (*économiques par exemple*) repose le plus souvent sur une décomposition tendance-cycle de la série. Toutefois, cette décomposition, si elle est très utilisée en pratique, ne repose pas sur une construction théorique unique. Il est alors souvent admis que la tendance correspond à l'équilibre de long terme, alors que le cycle correspond à la dynamique de court terme. Mais cette distinction ne suffit pas pour identifier clairement les deux composantes, et des hypothèses supplémentaires sont alors nécessaires.

Deux approches sont alors généralement utilisées : la première consiste à utiliser une théorie économique (*c'est à dire un modèle structurel dont les composantes auront des interprétations économiques*), alors que la seconde tend à utiliser des outils statistiques "neutres". Nous allons nous intéresser ici à cette seconde approche. Néanmoins, nous pouvons dès à présent noter que cette neutralité est difficile à mettre en oeuvre : il existe une infinité de façon de construire la *tendance moyenne*, par exemple. Il existe alors de nombreuses méthodes pour évaluer la croissance tendancielle.

Nous allons nous concentrer ici sur des décompositions additives de la forme $(X_t) = (T_t) + (C_t)$. Un modèle multiplicatif peut en effet se ramener à un modèle additif en passant au logarithme. Les méthodes traditionnelles reposent sur deux techniques : l'estimation d'une tendance déterministe et le lissage. Des méthodes plus récentes se basent sur la notion de tendance stochastique, avec en particulier la méthode de Beveridge et Nelson, et les modèles à composantes inobservables.

Parmi les méthodes de lissage, l'approche la plus simple consiste à utiliser des moyennes mobiles on utilise alors une moyenne (*pondérée*) de la série (X_t) dont la durée correspond au cycle, qui conserve la tendance et élimine le cycle. La moyenne symétrique arithmétique est l'exemple le plus simple : on considère alors la série (Y_t) définie par

$$Y_t = M(X_t) = \frac{1}{2m+1}(X_{t-m} + X_{t-m+1} + \dots + Y_{t-1} + Y_t + Y_{t+1} + \dots + Y_{t+m}) \quad (3)$$

Ce type de filtre, comme nous le verrons par la suite, conserve les tendances linéaires, et filtre (*ou annule*) les séries périodiques de période $2m+1$. Toutefois, deux problèmes apparaissent dans l'utilisation des filtres moyennes-mobiles

- les points extrêmes de la série ne peuvent être traités de la même façon que les autres points (*effet de bord*)
- les séries lissées sont souvent autocorrélées, non pas à cause de la structure de la série initiale, mais il s'agit d'une conséquence du processus de lissage (*effet Slutsky-Yule*).

D'autres méthodes de lissage existent, par exemple en utilisant la méthode *PAT* (*phase average trend*) ou le filtre de Hodrick-Prescott (1980).

L'estimation d'un trend déterministe repose sur l'utilisation de fonctions simples, par exemple linéaires,

$$X_t = T_t + C_t = a + bt + C_t \quad (4)$$

Ces modèles apparaissent parfois dans la littérature sous le terme *TS* (*trend stationary*), et le cycle (*supposé stationnaire*) apparaît alors comme l'écart à la tendance. Cette tendance est alors estimée par régression. Cette méthode sera celle développée dans la première partie, même si elle a été fortement critiquée : la croissance à long terme est alors fixée de façon mécanique. Des modèles avec rupture de tendance ont ainsi été introduits.

Il convient toutefois de noter que cette décomposition tendance-cycle ne sont pas adaptées pour les séries non-stationnaires, et il convient d'intégrer une composante stochastique dans la tendance. Le modèle de Beveridge et Nelson propose d'exprimer les composantes à l'aide d'une représentation *ARIMA* de la série (*traité dans l'exercice* (5)). Les modèles à composantes inobservables repose sur l'utilisation de modèles espace-état (*introduits dans le paragraphe* (7.4.3) *sur le filtre de Kalman*). Ces deux méthodes sont présentées dans l'article de Doz, Rabault et Sobczack *Décomposition tendance-cycle : estimations par des méthodes statistiques univariées* (1995).

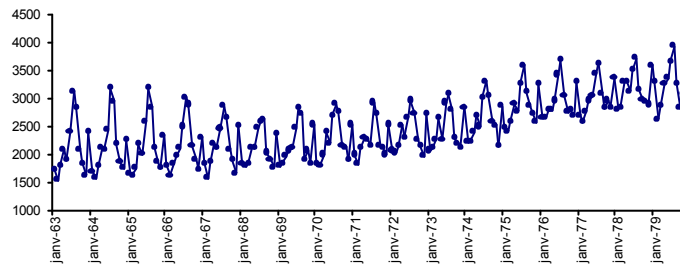
3 Désaisonnalisation par regression linéaire

3.1 Présentation des données

Nous considérons ici une série chronologique, mensuelle, comportant une forte saisonnalité : trafic voyageur de la SNCF (*source : Gouriéroux & Monfort (1995)*)

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1963	1750	1560	1820	2090	1910	2410	3140	2850	2090	1850	1630	2420
1964	1710	1600	1800	2120	2100	2460	3200	2960	2190	1870	1770	2270
1965	1670	1640	1770	2190	2020	2610	3190	2860	2140	1870	1760	2360
1966	1810	1640	1860	1990	2110	2500	3030	2900	2160	1940	1750	2330
1967	1850	1590	1880	2210	2110	2480	2880	2670	2100	1920	1670	2520
1968	1834	1792	1860	2138	2115	2485	2581	2639	2038	1936	1784	2391
1969	1798	1850	1981	2085	2120	2491	2834	2725	1932	2085	1856	2553
1970	1854	1823	2005	2418	2219	2722	2912	2771	2153	2136	1910	2537
1971	2008	1835	2120	2304	2264	2175	2928	2738	2178	2137	2009	2546
1972	2084	2034	2152	2522	2318	2684	2971	2759	2267	2152	1978	2723
1973	2081	2112	2279	2661	2281	2929	3089	2803	2296	2210	2135	2862
1974	2223	2248	2421	2710	2505	3021	3327	3044	2607	2525	2160	2876
1975	2481	2428	2596	2923	2795	3287	3598	3118	2875	2754	2588	3266
1976	2667	2668	2804	2806	2976	3430	3705	3053	2764	2802	2707	3307
1977	2706	2586	2796	2978	3053	3463	3649	3095	2839	2966	2863	3375
1978	2820	2857	3306	3333	3141	3512	3744	3179	2984	2950	2896	3611
1979	3313	2644	2872	3267	3391	3682	3937	3284	2849	3085	3043	3541
1980	2848	2913	3248	3250	3375	3640	3771	3259	3206	3269	3181	4008

Trafic voyageur SNCF - mensuel



3.2 Le modèle linéaire

La série X_t est la somme de 2 composantes déterministes : une **tendance** Z_t , d'une **saisonnalité** S_t et d'une **composante aléatoire** ε_t

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

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

$$\begin{cases} Z_t = Z_t^1 \beta_1 + Z_t^2 \beta_2 + \dots + Z_t^m \beta_m \\ S_t = S_t^1 \gamma_1 + S_t^2 \gamma_2 + \dots + S_t^n \gamma_n. \end{cases}$$

Le but est d'estimer les β_1, \dots, β_m et $\gamma_1, \dots, \gamma_n$ à partir des T observations.

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

3.2.1 Hypothèses sur les erreurs

On supposera l'hypothèse (H1) vérifiée, à savoir que les erreurs sont **centrées** : $\mathbb{E}(\varepsilon_t) = 0$, de **même variance** $V(\varepsilon_t) = \sigma^2$ et **non-corrélées** $cov(\varepsilon_t, \varepsilon_{t-h}) = 0$ pour tout $h > 0$.

3.2.2 Composante saisonnière du modèles

La forme de S_t dépend du type de données, et de la forme de la saisonnalité. On considèrera ici des fonctions S_t^i indicatrices,

$$S_t^i = \begin{cases} 0 & \text{si } t = \text{mois } i \\ 1 & \text{si } t \neq \text{mois } i \end{cases} \quad \text{ou } S_t^i = \begin{cases} 0 & \text{si } t = 0 \text{ [modulo } i] \\ 1 & \text{si } t \neq 0 \text{ [modulo } i] \end{cases}.$$

Exemple 15 Pour des données trimestrielles, on a $S_t = S_t^1\gamma_1 + S_t^2\gamma_2 + S_t^3\gamma_3 + S_t^4\gamma_4$ où S_t^j est la fonction indicatrice du trimestre j .

3.2.3 Composante tendancielle

Cette composante a généralement une forme simple, reflétant la croissance moyenne.

Exemple 16 Pour une tendance linéaire, $Z_t = \beta_1 + \beta_2 t$ on pose $Z_t^1 = 1$ et $Z_t^2 = t$.

Plusieurs types de composantes tendancielle existent :

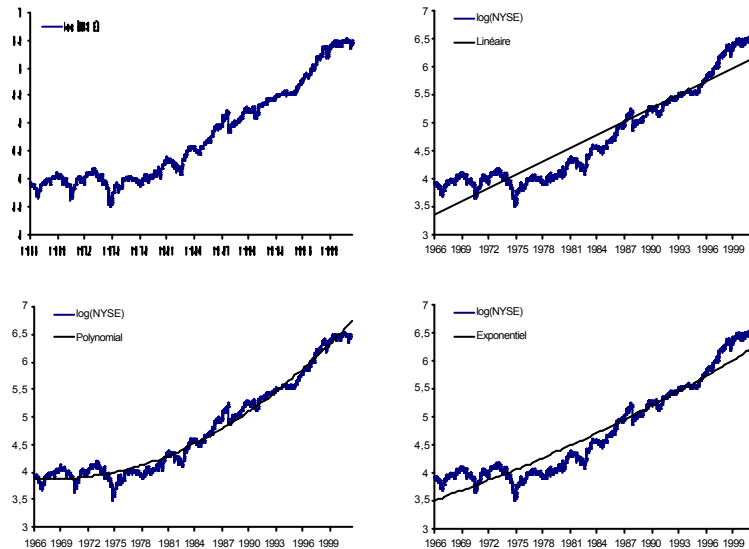
- (i) linéaire : $Z_t = \beta_0 + \beta_1 t$,
- (ii) exponentielle : $Z_t = \alpha\beta^t$, ou $Z_t = \alpha(1+r)^t$ ou encore $Z_t = \alpha \exp(rt)$,
- (iii) quadratique $Z_t = \beta_0 + \beta_1 t + \beta_2 t^2$,
- (iv) de Gompertz $Z_t = \exp(\alpha\beta^t + \gamma)$,
- (v) logistique $Z_t = [\alpha\beta^t - \gamma]^{-1}$.

Le cas (i) se traite par régression simple (cf partie suivante), le cas (ii) se ramène au cas (i) par transformation logarithmique, et le cas (iii) se traite par régression multiple. Il est également possible d'utiliser des modèles avec des ruptures :

$$Z_t = \begin{cases} \alpha_0 + \alpha_1 t & \text{pour } t \leq t_0 \\ \beta_0 + \beta_1 t & \text{pour } t > t_0. \end{cases}$$

Cette tendance est une des composante les plus compliquée à modéliser car il n'existe pas vraiment de méthode

Exemple 17 Considérons comme variable le logarithme de l'indice du New York Stock Exchange représenté ci-dessous, en haut à gauche, sur laquelle nous avons tenté trois ajustements différents : linéaires (en haut à droite), quadratique (en bas à gauche) et exponentiel (en bas à droite) :



La vraie difficulté est que la modélisation doit pouvoir avoir un "sens" : une tendance linéaire indique une croissance linéaire, alors qu'une tendance exponentielle indique une augmentation constante (rapport X_t/X_{t-1} constant). En revanche, une tendance quadratique peut être plus difficile à justifier puisque la plupart des modèles structurels sont généralement additifs (linéaires) ou multiplicatifs (linéaire en prenant le logarithme). Les tendances linéaires avec également très utilisée, puisqu'elle sont souvent plus adaptée qu'une tendance linéaire "simple", et surtout, la rupture a une interprétation structurelle.

3.2.4 Modèle trimestriel de Buys-Ballot (1847)

La désaisonnalisation par régression linéaire, dans le cas où la tendance est supposée linéaire, et les données sont trimestrielles, équivaut à tester le modèle linéaire

$$X_t = \underbrace{a + \beta t}_{Z_t} + \underbrace{\gamma_1 S_t^1 + \gamma_2 S_t^2 + \gamma_3 S_t^3 + \gamma_4 S_t^4}_{S_t} + \varepsilon_t,$$

où Z_t est la tendance (*linéaire*) et où S_t est la composante saisonnière. Supposons que les données commencent au 1er trimestre. Le modèle s'écrit alors, pour l'exemple du trafic SNCF

$$\begin{pmatrix} 5130 \\ 6410 \\ 8080 \\ 5900 \\ 5110 \\ 6680 \\ 8350 \\ 5910 \\ 5080 \\ \vdots \\ X_t \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \beta \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ \vdots \\ t \end{pmatrix} + \gamma_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ \vdots \\ S_t^1 \end{pmatrix} + \gamma_2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ S_t^2 \end{pmatrix} + \gamma_3 \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ \vdots \\ S_t^3 \end{pmatrix} + \gamma_4 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ \vdots \\ S_t^4 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \vdots \\ \varepsilon_t \end{pmatrix},$$

qui peut se réécrire, de façon matricielle,

$$\begin{pmatrix} 5130 \\ 6410 \\ 8080 \\ 5900 \\ 5110 \\ 6680 \\ 8350 \\ 5910 \\ 5080 \\ \vdots \\ X_t \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 & 0 \\ 1 & 3 & 0 & 0 & 1 & 0 \\ 1 & 4 & 0 & 0 & 0 & 1 \\ 1 & 5 & 1 & 0 & 0 & 0 \\ 1 & 6 & 0 & 1 & 0 & 0 \\ 1 & 7 & 0 & 0 & 1 & 0 \\ 1 & 8 & 0 & 0 & 0 & 1 \\ 1 & 9 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t & S_t^1 & S_t^2 & S_t^3 & S_t^4 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \vdots \\ \varepsilon_t \end{pmatrix} \text{ soit } X = Y\theta + \varepsilon$$

L'écriture de l'estimateur des moindres carrés ordinaires s'écrit $\hat{\theta} = (Y'Y)^{-1}Y'X$. Toutefois, cette écriture n'est possible que si $Y'Y$ est inversible, ce qui n'est pas le cas ici car la première colonne (*correspondant à la constante*) est égale à la somme des 4 dernières (*les composantes trimestrielles*). Deux méthodes sont alors possibles pour faire malgré tout l'identification du modèle.

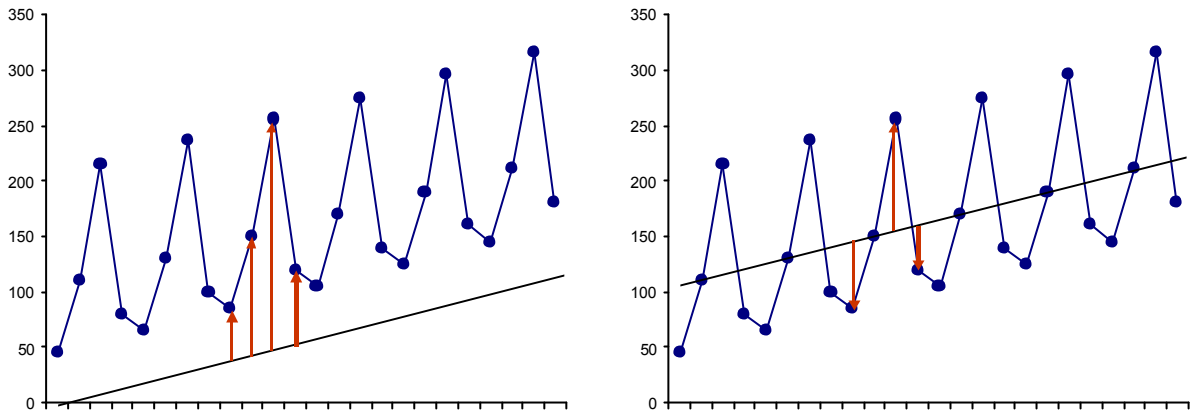
- ne pas tenir compte de la constante, et identifier le modèle

$$X_t = \beta t + \delta_1 S_t^1 + \delta_2 S_t^2 + \delta_3 S_t^3 + \delta_4 S_t^4 + \varepsilon_t, \quad (5)$$

rajouter une contrainte, et identifier le modèle

$$\begin{cases} X_t = \alpha + \beta t + \gamma_1 S_t^1 + \gamma_2 S_t^2 + \gamma_3 S_t^3 + \gamma_4 S_t^4 + \varepsilon_t \\ \text{sous contrainte } \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 = 0, \end{cases} \quad (6)$$

Cette dernière contrainte est arbitraire, mais correspond à une interprétation bien précise. Considérons la série représentée ci-dessous, avec une saisonnalité d'ordre 4 (*données trimestrielles*)



Le modèle (5) correspond au graphique ci-dessus à gauche : le modèle s'écrit

$$\begin{pmatrix} 45 \\ 110 \\ 215 \\ 80 \\ 65 \\ 130 \\ 235 \\ 100 \\ 85 \\ \vdots \\ X_t \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 & 0 \\ 3 & 0 & 0 & 1 & 0 \\ 4 & 0 & 0 & 0 & 1 \\ 5 & 1 & 0 & 0 & 0 \\ 6 & 0 & 1 & 0 & 0 \\ 7 & 0 & 0 & 1 & 0 \\ 8 & 0 & 0 & 0 & 1 \\ 9 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t & S_t^1 & S_t^2 & S_t^3 & S_t^4 \end{pmatrix} \begin{pmatrix} 5 \\ 40 \\ 100 \\ 200 \\ 60 \end{pmatrix},$$

alors que le modèle (6) correspond au graphique ci-dessus à droite et le modèle s'écrit alors

$$\begin{pmatrix} 45 \\ 110 \\ 215 \\ 80 \\ 65 \\ 130 \\ 235 \\ 100 \\ 85 \\ \vdots \\ X_t \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 & 0 \\ 1 & 3 & 0 & 0 & 1 & 0 \\ 1 & 4 & 0 & 0 & 0 & 1 \\ 1 & 5 & 1 & 0 & 0 & 0 \\ 1 & 6 & 0 & 1 & 0 & 0 \\ 1 & 7 & 0 & 0 & 1 & 0 \\ 1 & 8 & 0 & 0 & 0 & 1 \\ 1 & 9 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t & S_t^1 & S_t^2 & S_t^3 & S_t^4 \end{pmatrix} \begin{pmatrix} 100 \\ 5 \\ -60 \\ 0 \\ 100 \\ -40 \end{pmatrix}.$$

Dans le premier modèle, la tendance ne correspond pas tout à fait à la notion 'intuitive' de la saisonnalité

En fait, il est possible de voir aisément qu'un simple jeu de translations permet de passer d'un modèle à l'autre.

3.3 Estimateur des moindres carrés ordinaires (*mco*)

3.3.1 Solutions générales

On considère un modèle de la forme

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

La méthode des *mco* consiste à choisir les β_i et γ_j de façon à minimiser le carré des erreurs

$$\begin{aligned} (\widehat{\beta}_i, \widehat{\gamma}_j) &= \arg \min \left\{ \sum_{t=1}^T \varepsilon_t^2 \right\} \\ &= \arg \min \left\{ \sum_{t=1}^T \left[X_t - \sum_{i=1}^m Z_t^i \beta_i + \sum_{j=1}^n S_t^j \gamma_j \right]^2 \right\}. \end{aligned}$$

Notations : $\beta = (\beta_1, \dots, \beta_m)'$, $\gamma = (\gamma_1, \dots, \gamma_n)'$,

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

Le modèle s'écrit

$$X = Z\beta + S\gamma + \varepsilon = [Z|S] \begin{bmatrix} \beta \\ \gamma \end{bmatrix} + \varepsilon = Yb + \varepsilon,$$

et $\widehat{b} = (\widehat{\beta}, \widehat{\gamma})'$ vérifie alors l'équation

$$Y'Y\widehat{b} = Y'X \text{ soit } [Z \ S] \begin{bmatrix} Z' \\ S' \end{bmatrix} \widehat{b} = \begin{bmatrix} Z' \\ S' \end{bmatrix} X,$$

et donc

$$\begin{bmatrix} \widehat{\beta} \\ \widehat{\gamma} \end{bmatrix} = \begin{bmatrix} Z'Z & Z'S \\ S'Z & S'S \end{bmatrix}^{-1} \begin{bmatrix} Z'X \\ S'X \end{bmatrix},$$

ce qui donne les coefficients

$$\begin{cases} \widehat{\beta} = [Z'Z - Z'S(S'S)^{-1}S'Z]^{-1} [Z'X - Z'S(S'S)^{-1}S'X] \\ \widehat{\gamma} = [S'S - S'Z(Z'Z)^{-1}Z'S]^{-1} [S'X - S'Z(Z'Z)^{-1}Z'X]. \end{cases}$$

Remarque 12 *S'il n'y a pas d'effet saisonnier, $X = Z\beta + \varepsilon$, et on retrouve le modèle linéaire usuel, avec pour estimateur mco $\widehat{\beta} = [Z'Z]^{-1} Z'X$.*

3.3.2 Cas particulier : le modèle trimestriel de Buys-Ballot

Pour le modèle

$$X_t = \beta_1 + \beta_2 t + S_t^1 \gamma_1 + S_t^2 \gamma_2 + S_t^3 \gamma_3 + S_t^4 \gamma_4 + \varepsilon_t,$$

il est possible d'expliciter les différents coefficients. L'équation

$$\begin{cases} \min_{\beta, \gamma} \sum_{t=1}^T [X_t - \beta_1 - \beta_2 t - \sum_{j=1}^4 S_t^j \gamma_j]^2 \\ \text{sous contrainte (*) } \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 = 0, \end{cases}$$

peut se réécrire

$$\begin{cases} \min_{\beta, \gamma} \sum_{t=1}^T \left[X_t - \beta_2 t - \sum_{j=1}^4 S_t^j \delta_j \right]^2 & \text{où } \begin{cases} \beta_1 = [\delta_1 + \delta_2 + \delta_3 + \delta_4] / 4 \\ \gamma_j = \delta_j - \beta_1, \end{cases} \end{cases}$$

En notant N le nombre d'années entières ($N = T/4$), on pose

\widetilde{x}_n : moyenne des X_t relatives à l'année n

\overline{x}_j : moyenne des X_t relatives au trimestre j

\overline{x} : moyenne de toutes les observations X_t

On a alors les estimateurs suivant

$$\widehat{\beta}_2 = 3 \frac{\sum_{n=1}^N n \widetilde{x}_n - \frac{N(N+1)}{2} \overline{x}}{N(N^2 - 1)} \quad (7)$$

$$\widehat{\delta}_j = \overline{x}_j - [j + 2(N-1)] \widehat{\beta}_2 \text{ pour } j = 1, 2, 3, 4 \quad (8)$$

d'où finalement

$$\begin{cases} \widehat{\beta}_1 = [\widehat{\delta}_1 + \widehat{\delta}_2 + \widehat{\delta}_3 + \widehat{\delta}_4] / 4 \\ \widehat{\gamma}_j = \widehat{\delta}_j - \widehat{\beta}_1 \end{cases} \quad (9)$$

3.3.3 Généralisation des formules de Buys-Ballot (*tendance linéaire*)

Les relations obtenues dans le cas précédant peuvent en fait être généralisées dans le cas d'une périodicité m , et en notant (*de la même façon que précédemment*) N le nombre d'années entières. Le modèle s'écrit alors

$$X_t = \beta_1 + \beta_2 t + S_t^1 \gamma_1 + S_t^2 \gamma_2 + S_t^3 \gamma_3 + \dots + S_t^m \gamma_m + \varepsilon_t.$$

L'équation

$$\begin{cases} \min_{\beta, \gamma} \sum_{t=1}^T \left[X_t - \beta_1 - \beta_2 t - \sum_{j=1}^m S_t^j \gamma_j \right]^2 \\ \text{sous contrainte (*) } \gamma_1 + \gamma_2 + \gamma_3 + \dots + \gamma_m = 0, \end{cases}$$

admet alors pour solution, en notant

$$\begin{aligned} \hat{\beta}_2 &= \frac{12}{m} \frac{\sum_{n=1}^N n \tilde{x}_n - \frac{N(N+1)}{2} \bar{x}}{N(N^2 - 1)} \\ \hat{\beta}_1 &= \bar{x} - \hat{\beta}_2 \frac{Nm + 1}{2} \\ \hat{\gamma}_j &= \tilde{x}_j - \bar{x} - \hat{\beta}_2 \left[j - \frac{m+1}{2} \right] \end{aligned}$$

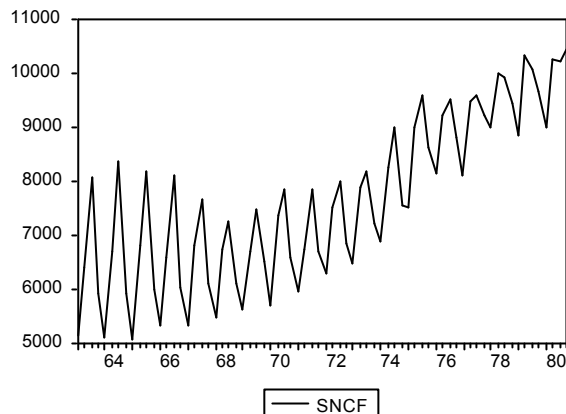
3.4 Application au trafic voyageur

3.4.1 Série agrégée par trimestre

Considérons la série du trafic SNCF agrégée par trimestre, représentée ci-dessous, avec en ligne les années, et en colonne les trimestres,

$n \setminus j$	1	2	3	4	\tilde{x}_n
1	5130	6410	8080	5900	6380
2	5110	6680	8350	5910	6513
3	5080	6820	8190	5990	6520
4	5310	6600	8090	6020	6505
5	5320	6800	7650	6110	6470
6	5486	6738	7258	6111	6398
7	5629	6696	7491	6494	6578
8	5682	7359	7836	6583	6865
9	5963	6743	7844	6692	6811
10	6270	7524	7997	6853	7161
11	6472	7871	8188	7207	7435
12	6892	8236	8978	7561	7917
13	7505	9005	9591	8608	8677
14	8139	9212	9522	8816	8922
15	8088	9494	9583	9204	9092
16	8983	9986	9907	9457	9583
17	8829	10340	10070	9669	9727
18	9009	10265	10236	10458	9992
\bar{x}_j	6605	7932	8603	7425	7641

représentée ci-dessous,



Considérons alors un modèle de la forme suivante, avec une saisonnalité en 4 composantes (les données étant trimestrielles : chaque composante correspondant à un trimestre), et une tendance supposée linéaire ($Z_t = \beta_1 + \beta_2 t$),

$$X_t = \beta_1 + \beta_2 t + S_t^1 \gamma_1 + S_t^2 \gamma_2 + S_t^3 \gamma_3 + S_t^4 \gamma_4 + \varepsilon_t,$$

Compte tenu de la sur-identification de ce modèle, on rajoute la contrainte que la somme des γ_j soit nulle (c'est à dire que la composante saisonnière soit centrée : $\mathbb{E}(S_t) = 0$). On peut alors faire l'estimation de la façon suivante :

- (i) on estime le modèle (5), c'est à dire sans contrainte, et sans constante β_1
- (ii) et on se ramène au modèle (6) en utilisant les relations.

Pour l'étape (i) deux méthodes analogues sont possibles : soit en utilisant les expressions des estimateurs, soit en effectuant la régression sous EViews

• Calcul direct des estimateurs

Les calculs ont été fait ici sous MSEXcel, et sont présentés ci-dessous :

	$n \setminus j$	$T1$	$T2$	$T3$	$T4$	\tilde{x}_n	$n \times \tilde{x}_n$
1963	1	5 130	6 410	8 080	5 900	6 380,00	6 380,00
1964	2	5 110	6 680	8 350	5 910	6 512,50	13 025,00
1965	3	5 080	6 820	8 190	5 990	6 520,00	19 560,00
1966	4	5 310	6 600	8 090	6 020	6 505,00	26 020,00
1967	5	5 320	6 800	7 650	6 110	6 470,00	32 350,00
1968	6	5 486	6 738	7 258	6 111	6 398,25	38 389,50
1969	7	5 629	6 696	7 491	6 494	6 577,50	46 042,50
1970	8	5 682	7 359	7 836	6 583	6 865,00	54 920,00
1971	9	5 963	6 743	7 844	6 692	6 810,50	61 294,50
1972	10	6 270	7 524	7 997	6 853	7 161,00	71 610,00
1973	11	6 472	7 871	8 188	7 207	7 434,50	81 779,50
1974	12	6 892	8 236	8 978	7 561	7 916,75	95 001,00
1975	13	7 505	9 005	9 591	8 608	8 677,25	112 804,25
1976	14	8 139	9 212	9 522	8 816	8 922,25	124 911,50
1977	15	8 088	9 494	9 583	9 204	9 092,25	136 383,75
1978	16	8 983	9 986	9 907	9 457	9 583,25	153 332,00
1979	17	8 829	10 340	10 070	9 669	9 727,00	165 359,00
1980	18	9 009	10 265	10 236	10 458	9 992,00	179 856,00
	\bar{x}_j	6 605	7 932	8 603	7 425	7641,39	

Pour chacune des années et chacun des trimestre, il est possible de calculer des moyennes : aussi, la moyenne pour 1963 était de 6380, et de 7435 pour 1973, et de façon analogue, la moyenne pour le premier trimestre est de 6605, et de 8603 pour le troisième. La moyenne totale est alors de 7641, pour ces 72 observations. Aussi, $N = 18$ (on a 18 années d'observations), et la pente de la droite de la tendance est donnée par

$$\hat{\beta}_2 = \frac{3}{N(N^2 - 1)} \left[\sum_{n=1}^N n \tilde{x}_n - \frac{N(N+1)}{2} \bar{x} \right] = \frac{3}{18(18^2 - 1)} [1\ 419\ 019\ -1\ 306\ 678] \approx 57.97$$

en utilisant les moyennes par trimestre, et par année, données dans le tableau ci-dessus, et

$$\widehat{\delta}_j = \bar{x}_j - [j + 2(N - 1)]\widehat{\beta}_2 \text{ et donc } \begin{cases} \widehat{\delta}_1 = 6605 - 35 \times 57.97 \approx 4577 \\ \widehat{\delta}_2 = 7932 - 36 \times 57.97 \approx 5845 \\ \widehat{\delta}_3 = 8603 - 37 \times 57.97 \approx 6459 \\ \widehat{\delta}_4 = 7425 - 38 \times 57.97 \approx 5222 \end{cases}$$

d'où finalement

$$\begin{cases} \widehat{\beta}_1 = [\widehat{\delta}_1 + \widehat{\delta}_2 + \widehat{\delta}_3 + \widehat{\delta}_4] / 4 \approx 5526 \\ \widehat{\gamma}_j = \widehat{\delta}_j - \widehat{\beta}_1 \end{cases} \text{ soit } \begin{cases} \widehat{\delta}_1 = 4577 - 5526 \approx -949 \\ \widehat{\delta}_2 = 5845 - 5526 \approx +320 \\ \widehat{\delta}_3 = 6459 - 5526 \approx +933 \\ \widehat{\delta}_4 = 5222 - 5526 \approx -304 \end{cases}$$

Aussi, le modèle s'écrit

$$\widehat{X}_t = 5526 + 58t - 949S_t^1 + 320S_t^2 + 933S_t^3 - 304S_t^4.$$

• Estimation sous *EViews*

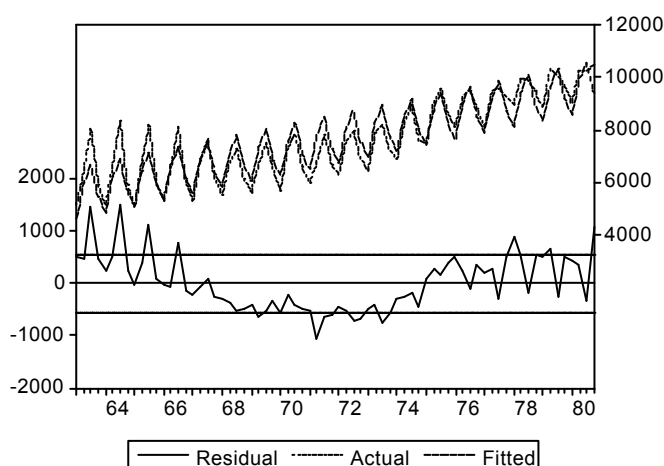
La régression se fait sur le modèle non-constraint, en ne prenant pas en compte la constante,

```
LS // Dependent Variable is SNCF
Sample: 1963:1 1980:4
Included observations: 72
```

Variable	Coefficient	Std. Error	T-Statistic	Prob.
TEMPS	57.96749	3.138733	18.46844	0.0000
TRIMESTRE1	4576.527	170.4094	26.85607	0.0000
TRIMESTRE2	5845.337	172.4495	33.89593	0.0000
TRIMESTRE3	6458.592	174.5222	37.00728	0.0000
TRIMESTRE4	5221.846	176.6264	29.56436	0.0000

R-squared	0.874440	Mean dependent var	7641.389
Adjusted R-squared	0.866943	S.D. dependent var	1515.211
S.E. of regression	552.7023	Akaike info criterion	12.69655
Sum squared resid	20467147	Schwartz criterion	12.85466
Log likelihood	-554.2395	F-statistic	116.6518
Durbin-Watson stat	0.807306	Prob(F-statistic)	0.000000

La lecture de la sortie donne effectivement les mêmes résultats numériques que les calculs présentés ci-dessus. Toutefois, il est aisé de voir que ce modèle est difficile à retenir, compte tenu du fait que les résidus ne semblent pas *i.i.d.*



L'erreur de modélisation (*les résidus*) est, en moyenne, beaucoup trop importante au début, ainsi qu'au milieu (*début des années 70*), mais l'erreur se faisant ici dans l'autre sens. Le caractère non-*i.i.d.* des résidus pouvait être devinée à la lecture des sorties de la régression, grâce au test de Durbin Watson, qui valide le caractère *AR*(1) des résidus.

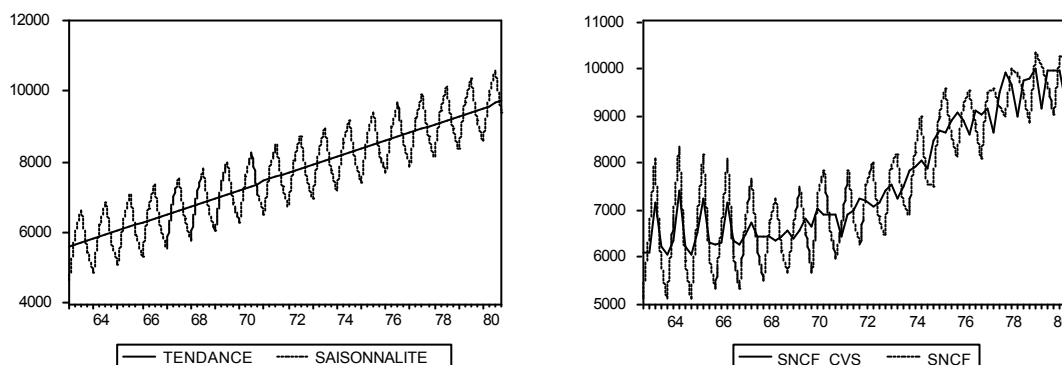
La **série ajustée** (*ci-dessous à gauche*) correspond à la série

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

avec (Z_t) en trait plein, et (\widehat{X}_t) en pointillés. Cette série pourra être prolongée afin de faire de la prévision. La **série corrigée des corrections saisonnières** (*CVS-ci-dessous à droite*) correspond à la série

$$\widehat{Y}_t = X_t - \widehat{S}_t = \sum_{i=1}^m Z_t^i \beta_i + \varepsilon_t.$$

Remarque 13 La composante saisonnière S_t correspond à $\sum_{j=1}^n S_t^j \gamma_j$, telle qu'elle apparaît dans le modèle contraint. Elle vérifie alors $\mathbb{E}(S_t) = 0$. Cette propriété n'est pas vérifiée dans le modèle sans constante.



3.4.2 Analyse sur données mensuelles

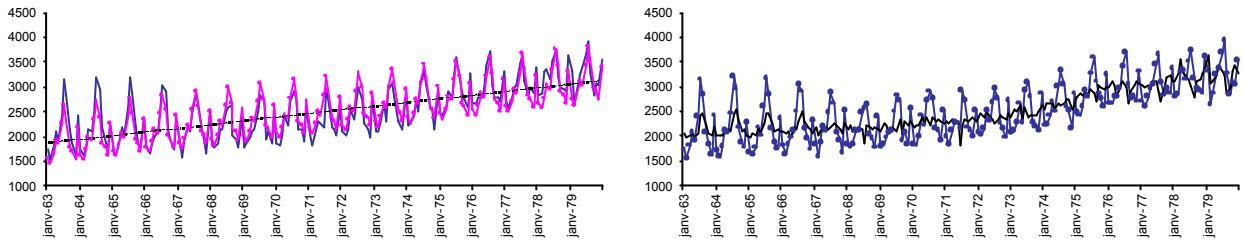
La méthode décrite ci-dessus donne les résultats suivants

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	\tilde{x}_n
1963	1750	1560	1820	2090	1910	2410	3140	2850	2090	1850	1630	2420	2127
1964	1710	1600	1800	2120	2100	2460	3200	2960	2190	1870	1770	2270	2171
1965	1670	1640	1770	2190	2020	2610	3190	2860	2140	1870	1760	2360	2173
1966	1810	1640	1860	1990	2110	2500	3030	2900	2160	1940	1750	2330	2168
1967	1850	1590	1880	2210	2110	2480	2880	2670	2100	1920	1670	2520	2157
1968	1834	1792	1860	2138	2115	2485	2581	2639	2038	1936	1784	2391	2133
1969	1798	1850	1981	2085	2120	2491	2834	2725	1932	2085	1856	2553	2192
1970	1854	1823	2005	2418	2219	2722	2912	2771	2153	2136	1910	2537	2288
1971	2008	1835	2120	2304	2264	2175	2928	2738	2178	2137	2009	2546	2270
1972	2084	2034	2152	2522	2318	2684	2971	2759	2267	2152	1978	2723	2387
1973	2081	2112	2279	2661	2281	2929	3089	2803	2296	2210	2135	2862	2478
1974	2223	2248	2421	2710	2505	3021	3327	3044	2607	2525	2160	2876	2639
1975	2481	2428	2596	2923	2795	3287	3598	3118	2875	2754	2588	3266	2892
1976	2667	2668	2804	2806	2976	3430	3705	3053	2764	2802	2707	3307	2974
1977	2706	2586	2796	2978	3053	3463	3649	3095	2839	2966	2863	3375	3031
1978	2820	2857	3306	3333	3141	3512	3744	3179	2984	2950	2896	3611	3194
1979	3313	2644	2872	3267	3391	3682	3937	3284	2849	3085	3043	3541	3242
1980	2848	2913	3248	3250	3375	3640	3771	3259	3206	3269	3181	4008	3331
\bar{x}_j	2195	2101	2309	2555	2489	2888	3249	2928	2426	2359	2205	2861	2547

qui donne les coefficients suivants

$\widehat{\beta}_2$	$\widehat{\delta}_1$	$\widehat{\delta}_2$	$\widehat{\delta}_3$	$\widehat{\delta}_4$	$\widehat{\delta}_5$	$\widehat{\delta}_6$	$\widehat{\delta}_7$	$\widehat{\delta}_8$	$\widehat{\delta}_9$	$\widehat{\delta}_{10}$	$\widehat{\delta}_{11}$	$\widehat{\delta}_{12}$
9.82	1038	943	1156	1380	1293	1667	1938	1517	1135	1123	975	1618

Ce qui donne la **série ajustée** (à gauche) et la **série corrigée des variations saisonnières** (à droite)



3.5 Propriétés statistiques des estimateurs

Sous l'hypothèse $\mathbb{E}(\varepsilon_t) = 0$ les estimateurs *mco* sont **sans biais** :

$$\mathbb{E}(\hat{\beta}_i) = \beta_i \text{ et } \mathbb{E}(\hat{\gamma}_j) = \gamma_j.$$

La variance des estimateurs peut être estimée par

$$\hat{V} \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = s^2 \begin{bmatrix} Z'Z & Z'S^{-1} \\ S'Z & S'S \end{bmatrix}, \text{ où } s^2 = \frac{1}{T-n-m} \sum_{t=1}^T \hat{\varepsilon}_t^2,$$

ce qui permet d'obtenir des intervalles de confiance sur les estimateurs.

3.6 Prévision à un horizon h

3.6.1 Calcul de la prévision

Soit $h \geq 1$. On suppose que le modèle reste valide en $T+h$ c'est à dire que

$$X_{T+h} = \sum_{i=1}^m Z_{T+h}^i \beta_i + \sum_{j=1}^n S_{T+h}^j \gamma_j + \varepsilon_{T+h},$$

avec $\mathbb{E}(\varepsilon_{T+h}) = 0$, $V(\varepsilon_{T+h}) = \sigma^2$ et $\text{cov}(\varepsilon_t, \varepsilon_{T+h}) = 0$ pour $t = 1, \dots, T$. La variable X_{T+h} peut être approchée par

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

Cette prévision est la meilleur (au sens de l'erreur quadratique moyenne) prévision, linéaire en X_1, \dots, X_T et sans biais. Un intervalle de confiance de cette prévision est de la forme

$$\left[\hat{X}_T(h) - \phi_{1-\alpha/2} \sqrt{\hat{e}_h}; \hat{X}_T(h) + \phi_{1-\alpha/2} \sqrt{\hat{e}_h} \right],$$

où $\phi_{1-\alpha/2}$ est le quantile d'ordre α de la loi de Student à $T-m-n$ degrés de liberté, et où

$$\begin{aligned} \hat{e}_h &= \hat{\mathbb{E}} \left(\left[\hat{X}_T(h) - X_{T+h} \right]^2 \right) = \hat{V} \left(\sum_{i=1}^m Z_{T+h}^i \hat{\beta}_i + \sum_{j=1}^n S_{T+h}^j \hat{\gamma}_j - \varepsilon_{T+h} \right) \\ &= \begin{bmatrix} \hat{\beta}' & \hat{\gamma}' \end{bmatrix} \hat{V} \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} + s^2. \end{aligned}$$

3.6.2 Application au trafic SNCF

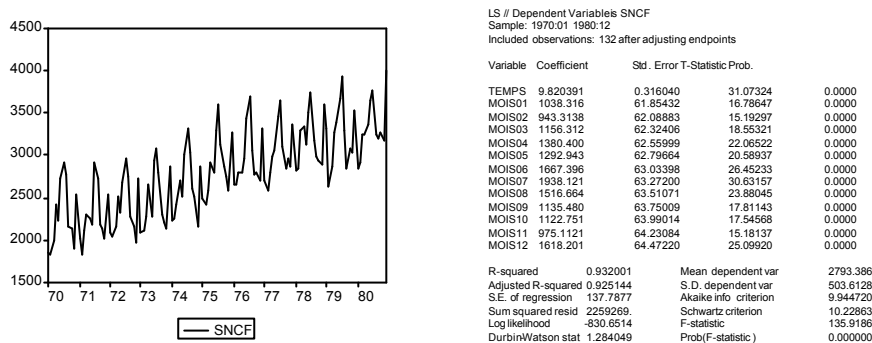
Dans l'exemple considéré précédemment, en données mensuelles, considérons désormais l'ensemble des données entre janvier 1970 et décembre 1980, et considérons le modèle suivant

$$X_t = \beta_2 t + S_t^1 \delta_1 + S_t^2 \delta_2 + S_t^3 \delta_3 + S_t^4 \delta_4 + S_t^5 \delta_5 + S_t^6 \delta_6 + S_t^7 \delta_7 + S_t^8 \delta_8 + S_t^9 \delta_9 + S_t^{10} \delta_{10} + S_t^{11} \delta_{11} + S_t^{12} \delta_{12} + \varepsilon_t$$

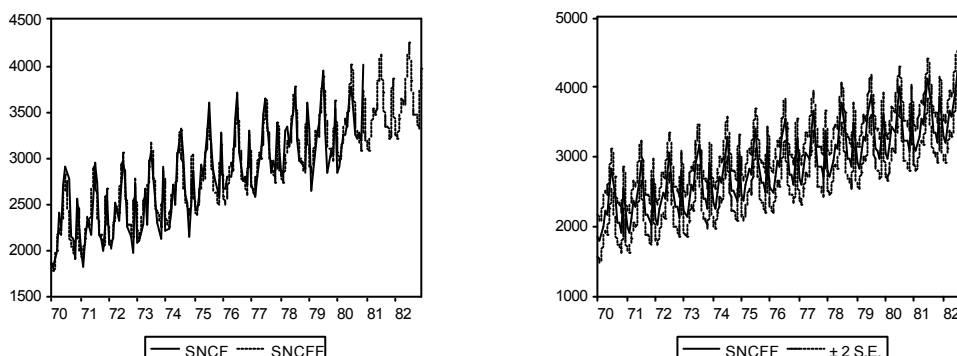
L'estimation par la méthode des moindres carrés donne l'estimation suivante

$\widehat{\beta}_2$	$\widehat{\delta}_1$	$\widehat{\delta}_2$	$\widehat{\delta}_3$	$\widehat{\delta}_4$	$\widehat{\delta}_5$	$\widehat{\delta}_6$	$\widehat{\delta}_7$	$\widehat{\delta}_8$	$\widehat{\delta}_9$	$\widehat{\delta}_{10}$	$\widehat{\delta}_{11}$	$\widehat{\delta}_{12}$
9.82	1038	943	1156	1380	1293	1667	1938	1517	1135	1123	975	1618

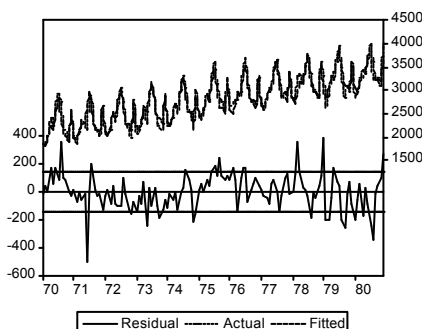
avec les estimations d'écart-types suivantes (*obtenues sous EViews*) :



Comme le montre la sortie ci-dessus à droite, tous les paramètres sont significatifs, le R^2 est relativement bon (93%), la statistique de Fisher F est suffisamment grande pour valider le modèle. La courbe de gauche ci-dessous correspond à la prévision du nombre de voyageurs pour 1982 et 1983, et l'intervalle de confiance de cette prévision est donné à droite,



A 95%, l'intervalle de confiance correspond à la prévision ± 145 (*soit prévision $\pm 5\%$*). Si cette prévision est aussi robuste, c'est aussi parce que l'on a restreint l'intervalle d'étude à 1970 – 1980, en enlevant les premières années. Les résidus ainsi obtenus sont représentés ci-dessous



4 Désaisonnalisation par la méthode des moyennes mobiles

On considère une série temporelle (X_t) admettant une décomposition

$$X_t = Z_t + S_t + \varepsilon_t \text{ pour } t = 1, \dots, T$$

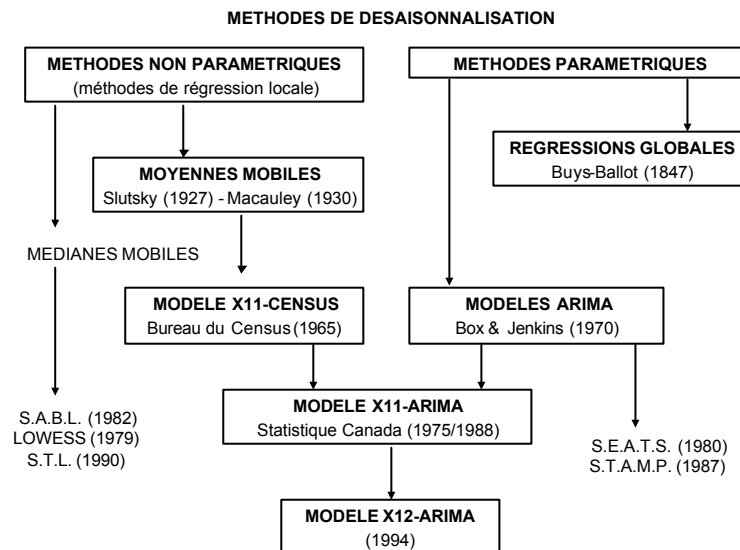
Le but est de trouver une transformation du processus X_t qui annule la composante saisonnière S_t : on cherche un "filtre" ϕ tel que $Y_t = \phi(X_t) = Z_t + \varepsilon_t$.

L'utilisation des **moyennes mobiles** est relativement ancienne puisqu'elle remonte à Poynting (1884) puis Hooker (1901) qui, les premiers, ont tenté d'oter (*et de distinguer*) la tendance et la composante cyclique pour des séries de prix en considérant des moyennes glissantes. En 1930, Macauley a introduit une méthode pour désaisonnaliser les séries au sein de la Réserve Fédérale américaine, basée sur l'utilisation de moyennes mobiles centrées d'ordre 12, pour obtenir une estimation de la tendance. Dans les années 50, le bureau du Census aux Etats Unis a commencé à développer des modèles basés sur l'utilisation de moyennes mobiles, modèles qui ont abouti à la méthode X11 en 1965.

Cette méthode a pu être développée grâce aux développements informatiques importants qui ont eu lieu à cette époque. Une des implications est que des décisions, faisant habituellement appel au jugement de l'utilisateur, ont pu être en grande partie automatisées. De plus, l'informatique a facilité l'utilisation de régressions visant à corriger les effets de jours ouvrables (*nombre de jours travaillés dans le mois*).

En 1975, suite au développement des modèles *ARIMA* (consécutif à la publication des travaux de Box et Jenkins dans les années 70), le modèle X11 a pu évoluer vers le modèle dit X11-ARIMA.

Le graphique ci-dessous résume l'évolution des méthodes de désaisonnalisation, montrant la différence entre modèles paramétriques (*régression linéaire - chapitre précédent*) et les modèles non-paramétriques,



La méthode de Buys-Ballot, basé sur une régression globale du modèle a été présenté dans la partie précédente. Nous allons présenter dans cette partie les méthodes basées sur des **régressions locales**. Les régressions locales consistent à ajuster des polynômes, en général par les moindres carrés, sur des intervalles glissants (*se décallant à chaque fois d'un point*). Au "centre" de cette intervalle, la donnée lissée est la valeur, à cette date, du polynôme ajusté. Ces régressions locales reviennent à appliquer des moyennes mobiles.

4.1 Généralités sur les moyennes mobiles

4.1.1 Notion d'opérateur retard L

Définition 16 On appellera **opérateur retard** L (=lag, ou B =backward) l'opérateur linéaire défini par

$$L : X_t \mapsto L(X_t) = LX_t = X_{t-1},$$

et **opérateur avance** F (=forward)

$$F : X_t \mapsto F(X_t) = FX_t = X_{t+1},$$

Remarque 14 $L \circ F = F \circ L = \mathbb{I}$ (opérateur identité) et on notera par la suite $F = L^{-1}$ et $L = F^{-1}$.

Polynômes d'opérateurs L

(i) Il est possible de composer les opérateurs : $L^2 = L \circ L$, et plus généralement,

$$L^p = \underbrace{L \circ L \circ \dots \circ L}_{p \text{ fois}} \quad \text{où } p \in \mathbb{N}$$

avec la convention $L^0 = \mathbb{I}$. On notera que $L^p(X_t) = X_{t-p}$.

(ii) Soit A le polynôme, $A(z) = a_0 + a_1z + a_2z^2 + \dots + a_pz^p$. On notera $A(L)$ l'opérateur

$$A(L) = a_0\mathbb{I} + a_1L + a_2L^2 + \dots + a_pL^p = \sum_{k=0}^p a_kL^k.$$

Soit (X_t) une série temporelle. La série (Y_t) définie par $Y_t = A(L)X_t$ vérifie

$$Y_t = A(L)X_t = \sum_{k=0}^p a_kX_{t-k}.$$

Par passage à la limite, on peut aussi définir des séries formelles

$$A(z) = \sum_{k=0}^{\infty} a_kz^k \quad \text{et} \quad A(L) = \sum_{k=0}^{\infty} a_kL^k.$$

Propriété 9 Pour toutes moyennes mobiles A et B , alors

$$\begin{cases} A(L) + B(L) = (A + B)(L) \\ \alpha \in \mathbb{R}, \alpha A(L) = (\alpha A)(L) \\ A(L) \circ B(L) = (AB)(L) = B(L) \circ A(L). \end{cases}$$

La moyenne mobile $C = AB = BA$ vérifie alors

$$\left(\sum_{k=0}^{\infty} a_kL^k \right) \circ \left(\sum_{k=0}^{\infty} b_kL^k \right) = \left(\sum_{i=0}^{\infty} c_iL^i \right) \quad \text{où } c_i = \sum_{k=0}^i a_kb_{i-k}.$$

4.1.2 Les moyennes mobiles

Définition 17 Une **moyenne mobile** est un opérateur linéaire, combinaison linéaire d'opérateurs retard

$$M = \sum_{i=-m_1}^{m_2} \theta_i L^{-i}, \quad \text{où } m_1, m_2 \in \mathbb{N},$$

qui peut s'écrire

$$M = L^{m_1} \sum_{i=0}^{m_1+m_2} \theta_{i-m_1} L^{-i} = L^{m_1} \sum_{i=0}^{m_1+m_2} \theta_{i-m_1} F^i = L^{m_1} \Theta(F), \quad (10)$$

où $\Theta(\cdot)$ est un polynôme appelé **polynôme caractéristique** de M , de **degré** $m_1 + m_2$, et $m_1 + m_2 + 1$ sera appelé **ordre** de M (correspondant au nombre (théorique) de terme de M).

Définition 18 Si $m_1 = m_2 = m$, la moyenne mobile sera dite **centrée**. De plus, si M est centrée, et que pour tout i , $\theta_i = \theta_{-i}$ alors la moyenne mobile est dite **symétrique**.

Exemple 18 La moyenne mobile $M_1(X_t) = (X_t + X_{t-1})/2$, soit $M_1 = (L + \mathbb{I})/2 = L[\mathbb{I} + F]/2$ est de degré 1, d'ordre 2 et n'est pas centrée (ni symétrique).

Exemple 19 La moyenne mobile $M_2(X_t) = (X_{t+1} + 2X_t + X_{t-1})/4$, soit $M_2 = (L^{-1} + 2\mathbb{I} + L)/4 = L[\mathbb{I} + 2F + F^2]/4$ est de degré 2, d'ordre 3, est centrée et symétrique.

On peut déjà noter, pour les moyennes centrées symétriques, sont nécessairement d'ordre impair (*pour être centrées*). Pour m impair, on considèrera les moyennes mobiles d'ordre $m = 2p + 1$ définie par

$$M_m(X_t) = \frac{1}{m} [X_{t-p} + X_{t-p+1} + \dots + X_{t-1} + X_t + X_{t+1} + \dots + X_{t+p-1} + X_{t+p}].$$

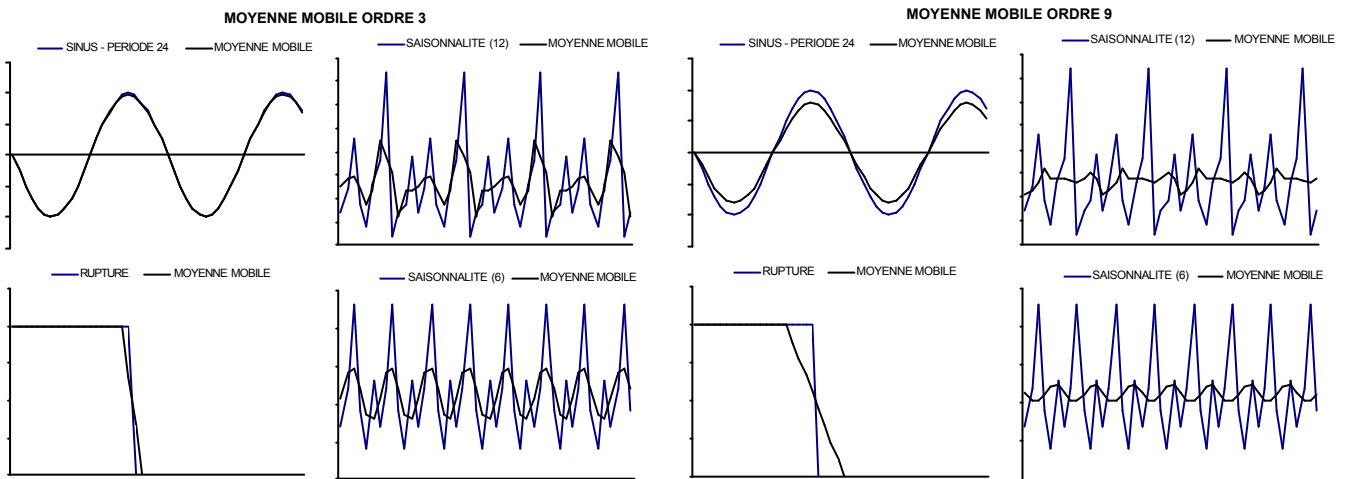
Exemple 20 *La moyenne mobile d'ordre 3* - Cette moyenne mobile a pour coefficients $1/3, 1/3, 1/3$,

$$M_3(X_t) = \frac{1}{3} [X_{t-1} + X_t + X_{t+1}].$$

Exemple 21 *La moyenne mobile d'ordre 9* - Cette moyenne mobile a pour coefficients $1/9, 1/9, \dots, 1/9$,

$$M_9(X_t) = \frac{1}{9} [X_{t-4} + X_{t-3} + \dots + X_t + \dots + X_{t+4}].$$

Les deux moyennes précédentes sont représentées sur les graphiques dessous,

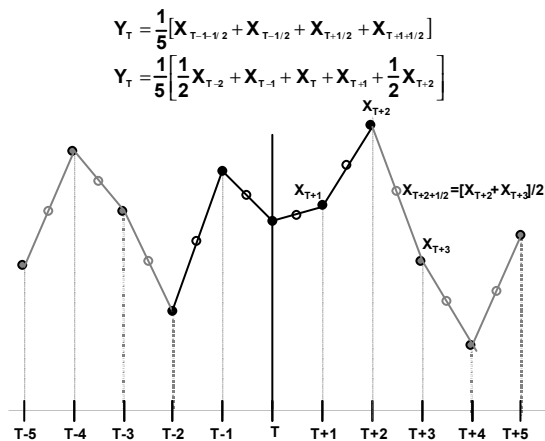


Toutefois, il est possible de construire des moyennes mobiles centrées et symétriques d'ordre pair, de façon artificielle. Pour cela, pour $m = 2p$ on considèrera les moyennes mobiles définies par

$$M_m(X_t) = \frac{1}{m} [X_{t-p+1/2} + \dots + X_{t-1/2} + X_{t+1/2} + \dots + X_{t+p-1/2}],$$

où $X_{t-1/2}$ est obtenue comme valeur intermédiaire entre X_{t-1} et X_t . Cette moyenne mobile peut donc se réécrire

$$\begin{aligned} M_m(X_t) &= \frac{1}{m} \left[\frac{1}{2} (X_{t-p} + X_{t-p+1}) + \dots + \frac{1}{2} (X_{t-1} + X_t) + \frac{1}{2} (X_t + X_{t+1}) + \dots + \frac{1}{2} (X_{t+p-1} + X_{t+p}) \right] \\ &= \frac{1}{m} \left[\frac{1}{2} X_{t-p} + X_{t-p+1} + \dots + X_{t-1} + X_t + X_{t+1} + \dots + X_{t+p-1} + \frac{1}{2} X_{t+p} \right]. \end{aligned}$$



Cette moyenne mobile d'ordre pair est en fait une moyenne mobile d'ordre impair, que l'on notera $M_{2 \times p}$, définie par

$$M_{2 \times p}(X_t) = \frac{1}{2m} [X_{t-p} + 2X_{t-p+1} + \dots + 2X_{t-1} + 2X_t + 2X_{t+1} + \dots + 2X_{t+p-1} + X_{t+p}].$$

Exemple 22 La moyenne mobile 2×4 - Cette moyenne mobile permet d'estimer des tendances dans le cas de données trimestrielles, elle est d'ordre 5 et de coefficients $1/8, 1/4, 1/4, 1/4, 1/8$

$$M_{2 \times 4}(X_t) = \frac{1}{8} [X_{t-2} + 2X_{t-1} + 2X_t + 2X_{t+1} + X_{t+2}].$$

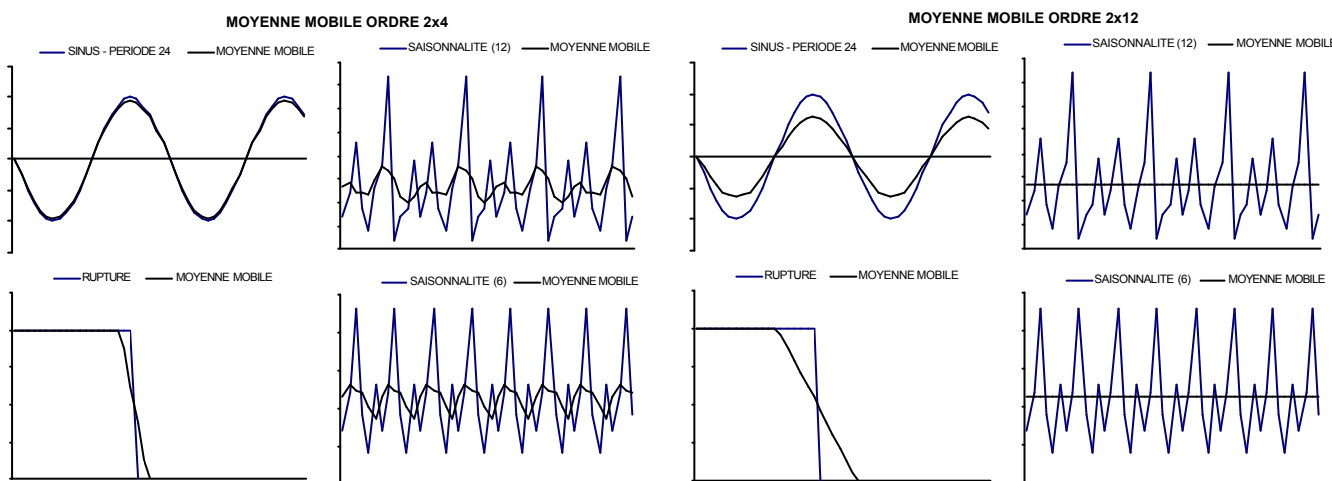
Comme nous le verrons par la suite, elle élimine les saisonnalités trimestrielles des séries trimestrielles, elle conserve les tendances linéaires, et elle réduit de 75% la variance d'un bruit blanc.

Exemple 23 La moyenne mobile 2×12 - Cette moyenne mobile permet d'estimer des tendances dans le cas de données mensuelles, elle est d'ordre 13 et de coefficients $1/24, 1/12, 1/12, \dots, 1/12, 1/24$

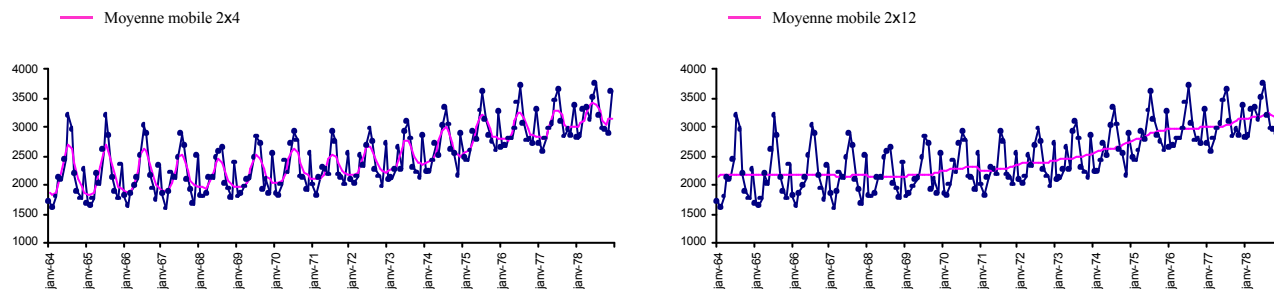
$$M_{2 \times 12}(X_t) = \frac{1}{24} [X_{t-6} + 2X_{t-5} + 2X_{t-4} + \dots + 2X_{t+5} + X_{t+6}].$$

Comme nous le verrons par la suite, elle élimine les saisonnalités annuelles des séries mensuelles, elle conserve les tendances linéaires, et elle réduit de plus de 90% la variance d'un bruit blanc.

Les deux moyennes précédentes sont représentées sur les graphiques dessous



Ces moyennes mobiles peuvent être appliquées à l'exemple du trafic SNCF de la partie précédente,



Exemple 24 La moyenne mobile 3×3 - Cette moyenne mobile est d'ordre 5 et de coefficients $1/9, 2/9, 3/9, 2/9, 1/9$

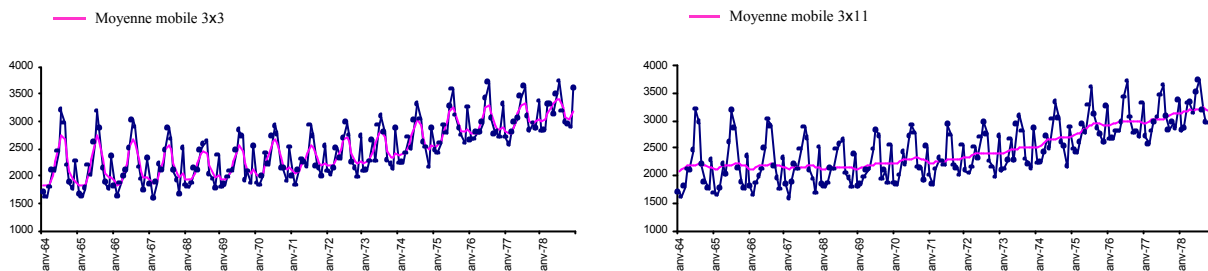
$$M_{3 \times 3}(X_t) = \frac{1}{9} [X_{t-2} + 2X_{t-1} + 3X_t + 2X_{t+1} + X_{t+2}].$$

Exemple 25 La moyenne mobile 3×9 - Cette moyenne mobile est d'ordre 11 et de coefficients $1/27, 2/27, 3/27, 3/27, \dots, 3/27, 2/27, 1/27$

$$M_{3 \times 9}(X_t) = \frac{1}{27} [X_{t-5} + 2X_{t-4} + 3X_{t-3} + 3X_{t-2} + 3X_{t-1} + \dots + 3X_{t+4} + 2X_{t+4} + X_{t+5}].$$

Ces deux moyennes mobiles conservent les droites, et réduisent respectivement de 75% et de 90% la variance d'un bruit blanc.

Ces moyennes mobiles peuvent être appliquées à l'exemple du trafic SNCF de la partie précédente,



Exemple 26 Les moyennes mobiles d'Henderson - Ces moyennes mobiles sont utilisées dans la méthode X11 pour extraire la tendance d'une estimation de la série corrigée des variations saisonnières (partie (4.4.2)). Ces moyennes reposent sur l'utilisation de l'opérateur $H = \sum [\Delta^3 \theta_i]^2$ où Δ est l'opérateur "différence première" ($\Delta X_t = X_t - X_{t-1}$). Cette quantité est nulle dans le cas où les θ_i se retrouvent sur une parabole : H mesure la distance entre la forme parabolique et la forme de la fonction définissant les θ_i . Henderson a cherché les moyennes mobiles centrées, d'ordre impair, conservant les polynômes de degré 2, et minimisant la fonction H :

$$\min_{\theta} \sum [\Delta^3 \theta_i]^2 \text{ sous contraintes } \sum_{i=-p}^{+p} \theta_i = 1, \sum_{i=-p}^{+p} i\theta_i = 0 \text{ et } \sum_{i=-p}^{+p} i^2\theta_i = 0,$$

ce qui donne les expressions explicites de θ_i , en posant $n = p + 2$

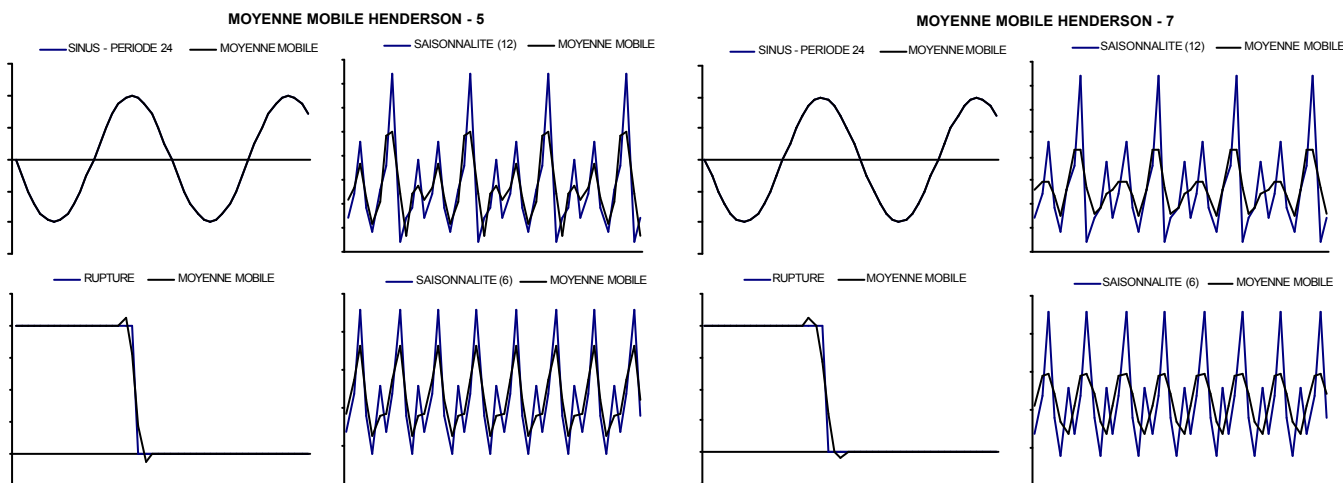
$$\theta_i = \frac{315 [(n-1)^2 - i^2] [n^2 - i^2] [(n+1)^2 - i^2] [3n^2 - 16 - 11i^2]}{8n(n^2 - 1)(4n^2 - 1)(4n^2 - 9)(4n^2 - 25)}.$$

Cette relation permet d'obtenir un certain nombre de moyennes mobiles

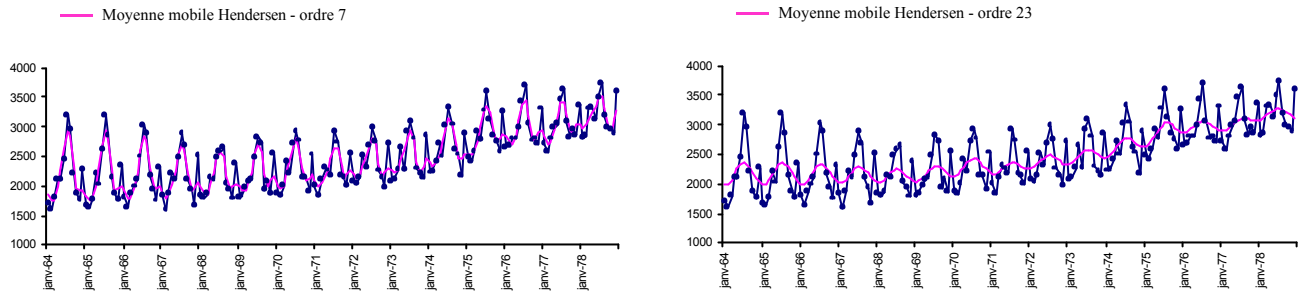
$$5 \text{ termes : } M_5(X_t) = \frac{1}{286} [-21X_{t-2} + 84X_{t-1} + 160X_t + 84X_{t+1} - 21X_{t+2}],$$

$$7 \text{ termes : } M_7(X_t) = \frac{1}{715} [-42X_{t-3} + 42X_{t-2} + 210X_{t-1} + 295X_t + 210X_{t+1} + 42X_{t+2} - 42X_{t+3}].$$

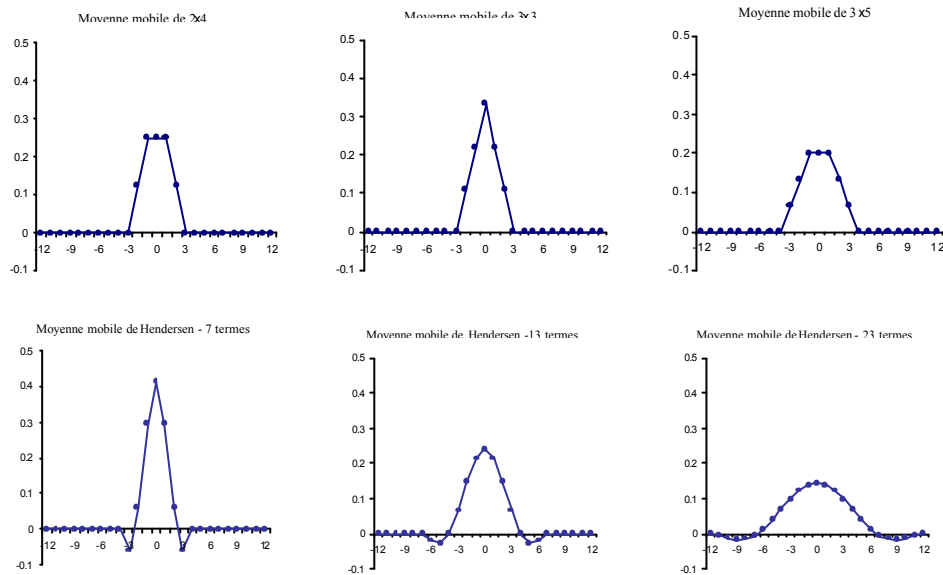
Les deux moyennes précédentes (5 et 7 termes) sont représentées sur les graphiques ci-dessous



Ces moyennes mobiles peuvent être appliquées à l'exemple du trafic SNCF de la partie précédente,



Les courbes ci-dessous montrent la forme de l'évolution des coefficients i



4.1.3 L'espace des opérateurs moyenne-mobile

Définition 19 Soient M_1 et M_2 deux moyennes mobiles. Le **produit** de M_1 et M_2 est obtenu par composition des moyennes mobiles

$$M_1 M_2 (X_t) = M_1 \circ M_2 (X_t).$$

Propriété 10 Ce produit est commutatif et associatif

$$M_1 M_2 = M_2 M_1 \text{ et } M_1 (M_2 M_3) = (M_1 M_2) M_3.$$

De plus, le produit est distributif par rapport à l'addition.

Propriété 11 L'ensemble des moyennes mobiles constitue un espace vectoriel.

Propriété 12 La composée de deux moyennes mobiles symétriques est symétrique.

4.2 Vecteurs propres associés à une moyenne mobile

Définition 20 Soit M une moyenne mobile. S'il existe λ et (X_t) non nul tels que $M(X_t) = \lambda X_t$, (X_t) sera **vecteur propre** associé à la **valeur propre** λ .

4.2.1 Les séries absorbées : $\lambda = 0$

Définition 21 Une suite (X_t) est dite **absorbée** par M si et seulement si $M(X_t) = 0$ pour tout t .

Exemple 27 Soit M la moyenne mobile définie par $M(X_t) = X_t + X_{t-1} + X_{t-2}$. La série chronologique définie récursivement par $Y_t = -[Y_{t-1} + Y_{t-2}]$ est absorbée par M .

Propriété 13 Les vecteurs propres associés à la valeur propre $\lambda = 0$ forment un espace vectoriel de dimension $m_1 + m_2$, dont une base est constituée des $Z_t^k = (\alpha_k r^t)$ pour $k = 0, 1, \dots, p-1$, où r est racine non nulle du polynôme Θ .

Exemple 28 Dans l'exemple (27), on peut chercher à construire une base de la forme $Z_t = r^t$, qui devra satisfaire

$$r^t + r^{t-1} + r^{t-2} = 0 \text{ pour tout } t$$

c'est à dire $r^2 + r + 1 = 0$. Aussi, r est une racine du polynôme caractéristique de M si et seulement si

$$r = \frac{-1 \pm i\sqrt{3}}{2} \text{ soient } r_1 = \exp\left(\frac{2i\pi}{3}\right) \text{ et } r_2 = \exp\left(-\frac{2i\pi}{3}\right)$$

Aussi, les suites absorbées sont nécessairement de la forme

$$X_t = \lambda r_1^t + \mu r_2^t, \text{ pour tout } t.$$

Or

$$\begin{cases} r_1^t = \cos(2t\pi/3) + i \sin(2t\pi/3) \\ r_2^t = \cos(2t\pi/3) - i \sin(2t\pi/3). \end{cases}$$

Et donc, l'espace vectoriel des suites absorbées par M admet pour base réelle

$$\mathcal{B} = \left\{ \cos\left(2t\frac{\pi}{3}\right), \sin\left(2t\frac{\pi}{3}\right) \right\},$$

ce qui correspond à des séries chronologiques de la forme

$$X_t = \lambda \cos\left(2t\frac{\pi}{3}\right) + \mu \sin\left(2t\frac{\pi}{3}\right) \text{ pour tout } t.$$

4.2.2 Absorption de la composante saisonnière

Théorème 3 Une moyenne mobile M absorbe la composante saisonnière de période T si et seulement si son polynôme caractéristique est divisible par $1 + z + \dots + z^{T-1}$.

Preuve. Par définition de la périodicité des composantes saisonnières, si T est la période, les S_t sont les suites solutions de

$$S_{t+1} + \dots + S_{t+T} = 0 \text{ pour tout } t \in \mathbb{Z}$$

L'espace vectoriel des solutions est engendré par les suites (r^t) où r satisfait

$$1 + r + r^2 + \dots + r^{T-1} = 0, \text{ c'est à dire } r = \exp\left(2ik\frac{\pi}{T}\right) \text{ où } k = 1, \dots, T-1$$

■

4.2.3 Les séries invariantes : $\lambda = 1$

Définition 22 Une suite (X_t) est dite **invariante** par M si et seulement si $M(X_t) = 0$ pour tout t

Une suite (X_t) est dite invariante par M si elle est absorbée par $(M - \mathbb{I})$.

Propriété 14 (i) Les suites constantes sont invariantes par M si et seulement si la somme de ses coefficients vaut 1,

(ii) Les polynômes de degré k sont invariantes par M si et seulement si 1 est racine d'ordre au moins $k+1$ de $\Phi = \Theta(z) - z^{m_1}$, où $M = L^{m_1}\Theta(F)$,

(iii) Si M est symétrique et conserve les constantes, alors M conserve les polynômes de degré 1.

Preuve. En reprenant (10), on peut écrire

$$M = \sum_{i=-m_1}^{m_2} \theta_i L^{-i} = L^{m_1} \Theta(F).$$

(i) Si $x_t = k \neq 0$ pour tout t , alors $Mx_t = x_t$ (*suite invariante*) se réécrit

$$\sum_{i=-m_1}^{m_2} \theta_i^{-i} k = k,$$

avec k non nul, donc la somme des coefficients vaut 1.

(iii) Soit X_t la suite telle que $X_t = t$. Alors

$$\begin{aligned} MX_t &= \theta_{-m_1}(t - m_1) + \dots + \theta_{m_1}(t + m_1) = t(\theta_{-m_1} + \dots + \theta_{m_1}) + m_1(-\theta_{-m_1} + \theta_{m_1}) \\ &\quad + (m_1 - 1)(-\theta_{-m_1+1} + \theta_{m_1-1}) + \dots + 1.(-\theta_{-1} + \theta_1) + 0.\theta_0, \end{aligned}$$

soit $MX_t = t.1 + 0 = t = X_t$ par symétrie ($\theta_{-k} = \theta_k$). ■

Les propriétés (i) et (iii) montrent dans quel cas la tendance de la série reste invariante : ces séries peuvent servir à enlever la composante saisonnière, pour récupérer la tendance linéaire.

4.2.4 Transformation de suites géométriques (r^t)

Propriété 15 Soit $M = L^{m_1} \Theta(F)$ une moyenne mobile de polynôme caractéristique Θ . Alors toute suite (r^t) est vecteur propre de M associé à la valeur propre $\lambda = r^{-m_1} \Theta(r)$.

Preuve. De (10), on peut écrire

$$M(r^t) = \sum_{i=-m_1}^{m_2} \theta_i r^{t+i} = r^{t-m_1} \sum_{i=-m_1}^{m_2} \theta_i r^{i+m_1} = r^t \times r^{-m_1} \Theta(r),$$

et donc $M(r^t) = r^{-m_1} \Theta(r) \times r^t$ pour tout t , ce qui correspond à la définition d'éléments propres. ■

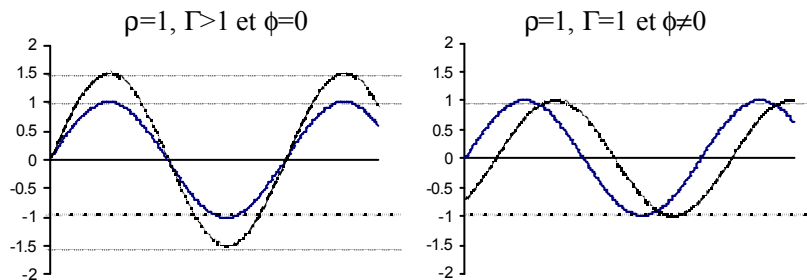
Suites géométriques réelles Si r est réel alors l'application d'une moyenne mobile à la suite géométrique (r^t) revient à faire une homothétie de rapport $r^{-m_1} \Theta(r)$.

Suites géométriques complexes Si $r = \rho e^{i\omega}$ alors appliquer M revient à multiplier r^t par le nombre complexe $r^{-m_1} \Theta(r) = \Gamma(\omega) e^{i\phi(\omega)}$,

$$\begin{cases} r^t = \rho^t e^{i\omega t} \\ M r^t = \rho^t e^{i\omega t} \Gamma(\omega) e^{i\phi(\omega)} = [\Gamma(\omega) \rho^t] e^{i[\omega t + \phi(\omega)]}, \end{cases}$$

ce qui correspond à un **effet d'échelle** (le module faisant intervenir le coefficient $\Gamma(\omega)$) comme dans le cas réel, mais aussi un **effet de phase** puisque l'argument se voit ajouter un élément $\phi(\omega)$.

Exemple 29 Une suite de la forme $\rho^t \sin \omega t$ sera transformée par M en $[\Gamma(\omega) \rho^t] \sin [\omega t + \phi(\omega)]$,



Propriété 16 Si M est symétrique, alors l'effet de phase sur la suite géométrique $(e^{i\omega t})$ est soit $\phi = 0$ soit $\phi = \pi$.

Preuve. Ce résultat se montre en explicitant $M(e^{i\omega t})$ et en notant que $\sum_{k=-m}^m \theta_{|j|} e^{i\omega k}$ est réel. Si ce réel est positif, alors $\phi = 0$ et s'il est négatif $\phi = \pi$. ■

Remarque 15 Ce résultat ne marche que pour la suite géométrique $(e^{i\omega t})$ et devient faux pour $\rho \neq 1$, même pour une moyenne mobile symétrique.

4.2.5 Moyenne mobile différence $\Delta_p = (\mathbb{I} - L)^p$

Considérons comme moyenne mobile l'opérateur différence $\Delta_p = (\mathbb{I} - L)^p$ pour $p > 0$. Cette moyenne mobile transforme un polynôme de degré $k \leq p$ en une constante. En effet, appliquer Δ_1 revient à abaisser le degré du polynôme de 1, car $\Delta_1(t^k) = t^k - (t-1)^k$, polynôme de degré $k-1$, et récursivement, appliquer $\Delta_p = \Delta_1^p$ revient à abaisser le degré du polynôme de p .

Une telle moyenne mobile permet d'éliminer une tendance qui serait un polynôme de bas degré. Nous allons étudier ici son effet sur une suite de la forme $(e^{i\omega t})$.

- Cas $p = 1$

$$(\mathbb{I} - L)(e^{i\omega t}) = e^{i\omega t} - e^{i\omega(t-1)} = e^{i\omega t} [1 - e^{-i\omega}] = e^{i\omega t} \cdot 2 \sin \frac{\omega}{2} \cdot e^{i[\pi-\omega]/2}$$

soit $\Gamma(\omega) = \Gamma_1(\omega) = 2 \sin(\omega/2)$ et $\phi(\omega) = \phi_1(\omega) = [\pi - \omega]/2$.

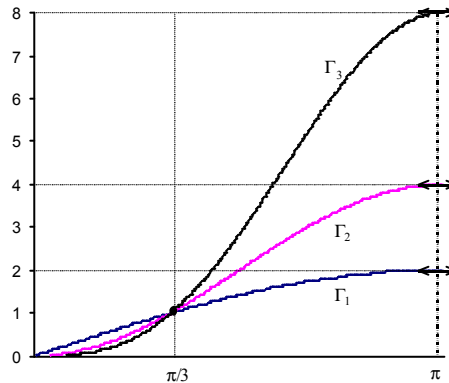
- Cas $p \geq 1$

$$\Delta_p = (\mathbb{I} - L)^p = \sum_{j=0}^p \binom{p}{j} (-1)^j L^j$$

donc

$$\Delta_p(e^{i\omega t}) = \sum_{j=0}^p \binom{p}{j} (-1)^j e^{-i\omega j} = e^{i\omega t} [1 - e^{-i\omega}]^p = e^{i\omega t} \cdot 2^p \left[\sin \frac{\omega}{2} \right]^p e^{ip[\pi-\omega]/2}$$

soit $\Gamma(\omega) = \Gamma_p(\omega) = [2 \sin(\omega/2)]^p$ et $\phi(\omega) = \phi_p(\omega) = p[\pi - \omega]/2$ où $\omega \in [0, \pi]$



Le graphique ci-dessus, représentant le facteur d'échelle Γ en fonction de ω montre que

$$\begin{cases} \text{si } \omega \leq \pi/3, \text{ le facteur d'échelle décroît avec } p \\ \text{si } \omega \geq \pi/3, \text{ le facteur d'échelle croît avec } p \end{cases}$$

Le déphasage est d'autant plus grand aux basses fréquences (ω petit) que p est plus élevé : pour un cycle trimestriel¹³ ($\omega = 2\pi/3$) l'amplitude est augmentée, alors que pour des cycles semi-annuels ($\omega = \pi/3$) l'amplitude est inchangée.

¹³En considérant des données mensuelles $T = 12$: $\omega = 2\pi/12$ correspond à 12 mois, $\omega = 2\pi/12 \times 2 = \pi/3$ correspond à 6 mois, $\omega = 2\pi/12 \times 4 = \pi/3$ correspond à 3 mois...etc.

Exemple 30 *Considérons la série de cycle $\pi/3$*

X_t	2	1	-1	-2	-1	1	2	1	-1	-2	-1	1
$\Delta_1 X_t$	1	-1	-2	-1	1	2	1	-1	-2	-1	1	2
$\Delta_2 X_t$	-1	-2	-1	1	2	1	-1	-2	-1	1	2	1
$\Delta_3 X_t$	-2	-1	1	2	1	-1	-2	-1	1	2	1	-1
$\Delta_4 X_t$	-1	1	2	1	-1	-2	-1	1	2	1	-1	-2
$\Delta_5 X_t$	1	2	1	-1	-2	-1	1	2	1	-1	-2	-1

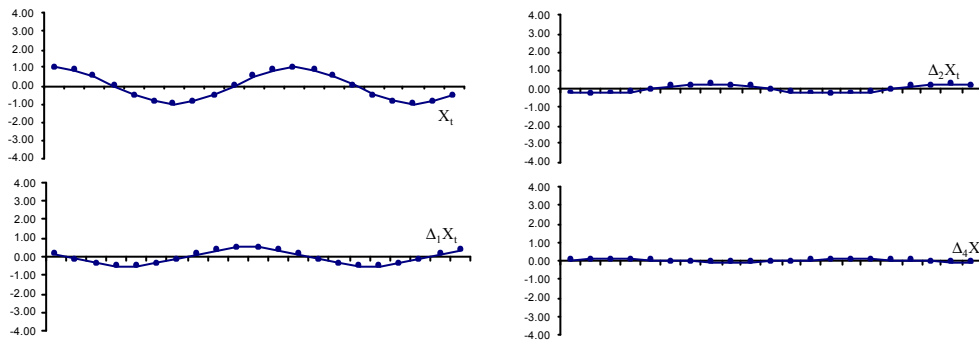
Comme on peut le noter, suite ne change pas d'amplitude, elle est juste déphasée.

Exemple 31 *Considérons la série de cycle $\pi/2$*

X_t	1	0	-1	0	1	0	-1	0	1	0	-1	0
$\Delta_1 X_t$	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1
$\Delta_2 X_t$	0	-2	0	2	0	-2	0	2	0	-2	0	2
$\Delta_3 X_t$	-2	-2	2	2	-2	-2	2	2	-2	-2	2	2
$\Delta_4 X_t$	-4	0	4	0	-4	0	4	0	-4	0	4	0
$\Delta_5 X_t$	-4	4	4	-4	-4	4	4	-4	-4	4	4	-4

On retrouve là aussi un déphasage, avec un coefficient d'amplitude qui augmente avec p . On peut ainsi noter que $\Delta_4 X_t$ correspond à 4 fois la série initiale X_t , avec un déphasage puisque les valeurs positives deviennent négatives, et inversement.

Exemple 32 *Considérons la série de cycle $\pi/6$, alors, comme le montre les graphiques ci-dessous (à gauche X_t et $\Delta_1 X_t$, et à droite $\Delta_2 X_t$ et $\Delta_4 X_t$), on retrouve là aussi un déphasage, avec un coefficient d'amplitude qui diminue avec p .*



4.2.6 Moyenne mobile différence saisonnière $\Delta_{p,s} = (\mathbb{I} - L^s)^p$

Cette moyenne mobile permet d'absorber les composantes saisonnières de période s .

$$\Delta_{p,s}(e^{i\omega t}) = e^{i\omega t} [1 - e^{-i\omega s p}],$$

alors

$$\begin{cases} \Gamma_{p,s}(\omega) = 2^p [\sin \omega s / 2]^p \\ \phi_{p,s}(\omega) = p [\pi - \omega s] / 2. \end{cases}$$

En considérant des données mensuelles, et une composante saisonnière annuelle (période 12) et $s = 12$, alors $\Gamma_{p,12}(2\pi/12) = 0$. On retrouve ainsi le fait que cette moyenne mobile $\Delta_{p,12}$ élimine une composante saisonnière de période 12. Toutefois, les saisonnalités de périodes 8 mois, ou 24 mois sont amplifiées.

4.2.7 Moyenne mobile impaire

Une moyenne mobile d'ordre impair peut s'écrire

$$M = \sum_{j=-q}^q \theta_j L^{-j} \text{ où } q \in \mathbb{N} \setminus \{0\} \text{ et } \theta_j = \frac{1}{2q+1}.$$

On peut montrer que

$$M(e^{i\omega t}) = e^{i\omega t} \frac{\sin[(q+1/2)\omega]}{(2q+1)\sin[\omega/2]},$$

(démonstration en exercice) d'où

$$\Gamma(\omega) = \left| \frac{\sin[(q+1/2)\omega]}{(2q+1)\sin[\omega/2]} \right|.$$

La moyenne mobile étant symétrique, le déphasage est 0 ou π (selon les valeurs de ω).

4.2.8 Moyenne mobile paire

Une moyenne mobile dite d'ordre impair peut s'écrire

$$M = \sum_{j=-q}^q \theta_j L^{-j} \text{ où } q \in \mathbb{N} \setminus \{0\} \text{ et } \theta_j = \begin{cases} 1/2q & \text{pour } |j| \neq q \\ 1/4q & \text{sinon.} \end{cases}$$

On peut montrer que

$$M(e^{i\omega t}) = e^{i\omega t} \frac{\sin[q\omega]}{2q} \cotan\left(\frac{\omega}{2}\right),$$

(démonstration en exercice) d'où

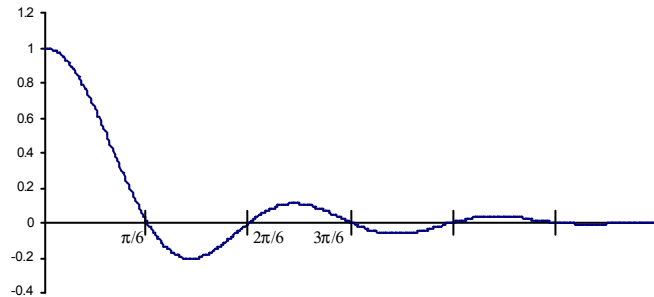
$$\Gamma(\omega) = \left| \frac{\sin[q\omega]}{2q} \cotan\left(\frac{\omega}{2}\right) \right|.$$

Pour des données mensuelles, par exemple, on prend $q = 6$, et

$$\Gamma(\omega) = \left| \frac{\sin[6\omega]}{12} \cotan\left(\frac{\omega}{2}\right) \right|,$$

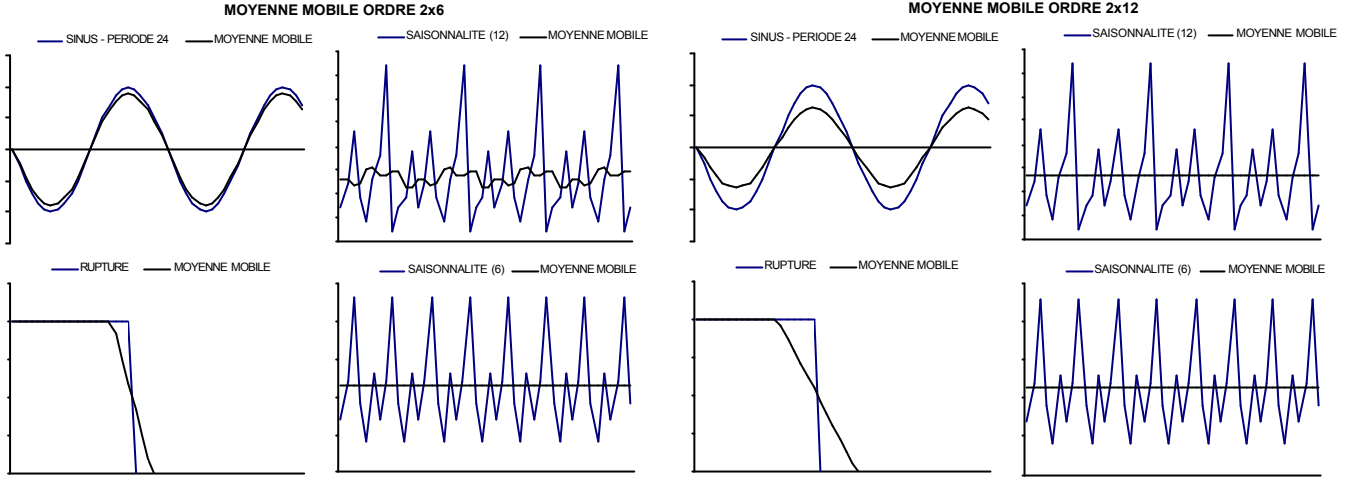
qui s'annule en $\pi/6, 2\pi/6, 3\pi/6, 4\pi/6, 5\pi/6, \dots$ correspondant à des périodes 12, 6, ...

Remarque 16 La moyenne mobile d'ordre pair avec $q = 6$ absorbe les composantes saisonnières périodiques de périodes 12 (et les harmoniques), tout en conservant les basses fréquences (correspondant à la tendance). Cette moyenne mobile est appelée $M_{2 \times 12}$ dans la partie (4.4.2) sur la méthode X11.



Exemple 33 Les graphiques ci-dessous représentent, à gauche, la moyenne mobile 2×6 et à droite, la moyenne

mobile 2×12



La première permet de faire disparaître la saisonnalité semi-annuelle (période 6 - en bas à droite), et la seconde, la saisonnalité annuelle (période 12 - en haut à droite). On peut également noter, pour la sinusoïde de période 24, que plus l'ordre de la moyenne mobile est important, plus l'amplitude diminue.

4.3 Notion de 'bruit blanc'

De la même façon que pour la notion de stabilité, il existe deux façon de définir le bruit blanc,

Définition 23 On appellera **bruit blanc 'faible'** toute suite $(\varepsilon_t, t \in \mathbb{Z})$ telle que $\mathbb{E}(\varepsilon_t) = 0$ et $V(\varepsilon_t) = \sigma^2$ pour tout $t \in \mathbb{Z}$ et tel que

$$\gamma(h) = \text{cov}(\varepsilon_t, \varepsilon_{t-h}) = 0 \text{ pour } h \neq 0.$$

Définition 24 On appellera **bruit blanc 'fort'** toute suite $(\varepsilon_t, t \in \mathbb{Z})$ telle que (ε_t) soit i.i.d.

Remarque 17 On notera par la suite $(\varepsilon_t) \sim BB(0, \sigma^2)$ pour bruit blanc faible.

4.3.1 Transformation d'un bruit blanc

Propriété 17 Soit M la moyenne mobile définie par (10) et $(\varepsilon_t) \sim BB(0, \sigma^2)$ pour $t \in \mathbb{Z}$. Le processus $X_t = M(\varepsilon_t)$ est stationnaire, centré ($\mathbb{E}(X_t) = 0$), tel que

$$\sum_h \gamma_X(h) z^h = \sigma^2 M(z) M\left(\frac{1}{z}\right) \text{ où } M(z) = \sum_{i=-m_1}^{m_2} \theta_i z^{-i} \text{ pour } z \neq 0.$$

Preuve. Le caractère centré découle de l'écriture

$$X_t = \sum_{j=-m_1}^{m_2} \theta_j \varepsilon_{t+j}, \text{ où } (\varepsilon_t) \text{ est centré.}$$

Il est possible de prolonger cette somme sur \mathbb{Z} en posant $\theta_j = 0$ pour $j < -m_1$ ou $j > m_2$. Aussi

$$X_t X_{t+h} = \sum_{j=-\infty}^{+\infty} \theta_j \varepsilon_{t+j} \sum_{k=-\infty}^{+\infty} \theta_k \varepsilon_{t+h+k} = \sigma^2 \sum_{j,k=-\infty}^{+\infty} \theta_j \theta_k \varepsilon_{t+j} \varepsilon_{t+h+k},$$

et donc

$$\mathbb{E}(X_t X_{t+h}) = \sigma^2 \sum_{k+h=j}^{+\infty} \theta_j \theta_k = \sigma^2 \sum_{j=-\infty}^{+\infty} \theta_j \theta_{j-h}.$$

Ainsi $\mathbb{E}(X_t X_{t+h}) = \text{cov}(\varepsilon_t, \varepsilon_{t+h})$ qui ne dépend que de h .

On peut alors écrire

$$\begin{aligned} \sum_h \gamma_X(h) z^h &= \sigma^2 \sum_{h=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} \theta_j \theta_{j-h} z^h = \sigma^2 \sum_{h=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} \theta_j \theta_{j-h} z^j z^{h-j} \\ &= \sigma^2 \sum_{j=-\infty}^{+\infty} \theta_j z^j \sum_{h=-\infty}^{+\infty} \theta_{j-h} z^{h-j} = \sigma^2 \sum_{j=-\infty}^{+\infty} \theta_j z^j \sum_{i=-\infty}^{+\infty} \theta_i \frac{1}{z^i}, \end{aligned}$$

en effectuant le changement d'indice $i = j - h$. ■

A retenir 3 *Le but des moyennes mobiles est*

- (i) *d'absorber les composantes saisonnières en laissant invariantes les tendances,*
- (ii) *de réduire la variance des perturbations.*

Définition 25 *L'indice de réduction de la moyenne mobile M est donné par*

$$\tau = \frac{\mathbb{E}(M X_t^2)}{\mathbb{E}(X_t^2)} = \sum_j \theta_j^2.$$

Exemple 34 *Dans le cas d'une moyenne mobile définie par $M(X_t) = [X_t + X_{t-1}]/2$, alors $\tau = 1/2$.*

4.4 Les procédures X11 et X12

4.4.1 Un algorithme simple de désaisonnalisation

Considérons une série mensuelle X_t composée d'une tendance Z_t , d'un cycle saisonnier S_t , et d'une perturbation ε_t , de façon additive $X_t = Z_t + S_t + \varepsilon_t$. L'algorithme suivant, en 4 étapes, permet de désaisonnaliser la série X_t

(1) *Estimation de la tendance par moyenne mobile $Z_t^{(1)} = M(X_t)$, où la moyenne mobile M est choisie de façon à reproduire au mieux la tendance, tout en éliminant la composante saisonnière, et en réduisant la perturbation au maximum*

(2) *Estimation de la composante saisonnière et de la perturbation $\Sigma_t = S_t + \varepsilon_t : \Sigma_t^{(1)} = X_t - Z_t^{(1)}$,*

(3) *Estimation de la composante saisonnière par moyenne mobile $S_t^{(1)} = M'(\Sigma_t^{(1)})$ et $\varepsilon_t^{(1)} = \Sigma_t^{(1)} - S_t^{(1)}$. Il s'agit ici de lisser les valeurs de la composante Σ_t de chaque mois pour extraire l'évolution du coefficient saisonnier du mois concerné. La moyenne mobile M' utilisée ici devra reproduire la composante saisonnière de chaque mois en réduisant au maximum la composante irrégulière. Une contrainte de normalisation des coefficients devra être imposée (*somme nulle*).*

(4) *Estimation de la série corrigée des variations saisonnières $\hat{X}_t = (Z_t^{(1)} + \varepsilon_t^{(1)}) = X_t - S_t^{(1)}$.*

La difficulté ici est donc de bien choisir les deux moyennes mobiles à utiliser M et M' .

4.4.2 L'algorithme de base de la méthode X11

Cette méthode propose deux moyennes mobiles dans le cas de données mensuelles. L'algorithme devient

(1) *Estimation de la tendance-par moyenne mobile 2×12*

$$Z_t^{(1)} = M_{2 \times 12}(X_t),$$

Cette moyenne mobile est paire, avec $q = 6$. Aussi, les 13 coefficients sont $1/24, 1/12, 1/12, \dots, 1/12, 1/24$. Cette moyenne mobile conserve les tendances linéaires, élimine les saisonnalités d'ordre 12 et minimise la variance de la perturbation.

(2) *Estimation de la composante saisonnière et de la perturbation $\Sigma_t = S_t + \varepsilon_t$*

$$\Sigma_t^{(1)} = X_t - Z_t^{(1)},$$

(3) *Estimation de la composante saisonnière par moyenne mobile 3×3 sur chaque mois*

$$S_t^{(1)} = M_{3 \times 3}(\Sigma_t^{(1)}) \text{ et } \varepsilon_t^{(1)} = \Sigma_t^{(1)} - S_t^{(1)}.$$

La moyenne mobile utilisée ici est une moyenne mobile sur 5 termes, dite 3×3 , dont les coefficients sont $1/9, 2/9, 3/9, 2/9, 1/9$ qui conserve les composantes linéaires. Les coefficients sont alors normalisés de telle sorte que leur somme, sur toute une période de 12 mois, soit nulle.

$$\widehat{S}_t^{(1)} = S_t^{(1)} - M_{2 \times 12} \left(S_t^{(1)} \right),$$

(4) *Estimation de la série corrigée des variations saisonnières*

$$\widehat{X}_t^{(1)} = X_t - \widehat{S}_t^{(1)}.$$

Cette première estimation de la série corrigée des variations saisonnières doit, par construction, contenir moins de saisonnalité.

(5) *Estimation de la tendance par moyenne mobile de Henderson sur 13 termes*

$$Z_t^{(2)} = M_{13} \left(\widehat{X}_t^{(1)} \right).$$

Si les moyennes mobiles d'Henderson n'ont pas de propriétés spéciales quant à l'élimination de la saisonnalité, mais elles lissent relativement bien, tout en conservant (*localement*) les polynômes d'ordre 2.

(6) *Estimation de la composante saisonnière et de la perturbation Σ_t*

$$\Sigma_t^{(2)} = X_t - Z_t^{(2)},$$

(7) *Estimation de la composante saisonnière par moyenne mobile 3×5 sur chaque mois*

$$S_t^{(2)} = M_{3 \times 5} \left(\Sigma_t^{(2)} \right) \text{ et } \varepsilon_t^{(2)} = \Sigma_t^{(2)} - S_t^{(2)}.$$

La moyenne mobile utilisée ici est une moyenne mobile sur 7 termes, dite 3×5 , dont les coefficients sont $1/15, 2/15, 3/15, 3/15, 2/15, 1/15$ qui conserve les composantes linéaires. Les coefficients sont alors normalisés de telle sorte que leur somme, sur toute une période de 12 mois, soit nulle.

$$\widehat{S}_t^{(2)} = S_t^{(2)} - M_{2 \times 12} \left(S_t^{(2)} \right),$$

(8) *Estimation de la série corrigée des variations saisonnières*

$$\widehat{X}_t^{(2)} = X_t - \widehat{S}_t^{(2)}.$$

Remarque 18 *Cette méthode permet de repérer les points abérants d'une série.*

Comme l'ont montré Gouriéroux et Monfort (1990), cet algorithme peut se résumer à l'application d'un unique moyenne mobile qui peut être explicitée matriciellement. Les 8 points de l'algorithme s'écrivent

- (1) $Z_t^{(1)} = M_{2 \times 12} (X_t)$
- (2) $\Sigma_t^{(1)} = X_t - Z_t^{(1)} = (\mathbb{I} - M_{2 \times 12}) (X_t)$
- (3) $S_t^{(1)} = M_{3 \times 3} \left(\Sigma_t^{(1)} \right) = M_{(3)} (\mathbb{I} - M_{2 \times 12}) (X_t)$
 $\widehat{S}_t^{(1)} = S_t^{(1)} - M_{2 \times 12} \left(S_t^{(1)} \right) = M_{(3)} (\mathbb{I} - M_{2 \times 12})^2 (X_t)$
- (4) $\widehat{X}_t^{(1)} = X_t - \widehat{S}_t^{(1)} = \left(\mathbb{I} - M_{(3)} (\mathbb{I} - M_{2 \times 12})^2 \right) (X_t)$
- (5) $Z_t^{(2)} = M_{13} \left(\widehat{X}_t^{(1)} \right) = M_{13} \left(\mathbb{I} - M_{(3)} (\mathbb{I} - M_{2 \times 12})^2 \right) (X_t)$
- (6) $\Sigma_t^{(2)} = X_t - Z_t^{(2)} = \left(\mathbb{I} - \left(\mathbb{I} - M_{(3)} (\mathbb{I} - M_{2 \times 12})^2 \right) \right) (X_t)$
- (7) $\widehat{S}_t^{(2)} = S_t^{(2)} - M_{2 \times 12} \left(S_t^{(2)} \right) = (\mathbb{I} - M_{2 \times 12}) M_{(5)} \left(\mathbb{I} - M_{(3)} (\mathbb{I} - M_{2 \times 12})^2 \right) (X_t)$
- (8) $\widehat{X}_t^{(2)} = X_t - \widehat{S}_t^{(2)} = \left(\mathbb{I} - (\mathbb{I} - M_{2 \times 12}) M_{(5)} \left(\mathbb{I} - M_{(3)} (\mathbb{I} - M_{2 \times 12})^2 \right) \right) (X_t)$

où $M_{(3)}$ est la moyenne mobile définie sur 49 mois, dont les coefficients sont

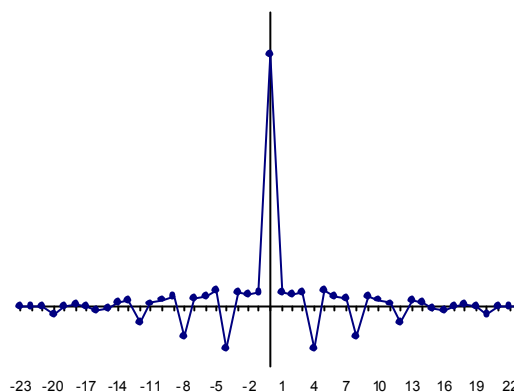
$$\{1/9, 0, 0, \dots, 0, 2/9, 0, \dots, 0, 3/9\}$$

et $M_{(5)}$ est la moyenne mobile définie sur 73 mois, dont les coefficients sont

$$\{1/27, 0, 0, \dots, 0, 2/27, 0, \dots, 0, 3/27, 0, 0, \dots, 0, 3/27\}$$

La moyenne mobile ainsi définie est d'ordre 169, c'est à dire qu'en toute rigueur, il faudrait 84 observations, soit 7 ans de part et d'autre pour pouvoir estimer ce filtre.

Remarque 19 *Un algorithme similaire existe pour des données trimestrielles, qui peut là aussi s'écrire sous la forme d'une seule moyenne mobile, portant sur 28 trimestres de part et d'autre (7 ans là aussi). Dans le cas trimestriel, l'allure des coefficients θ_i est la suivante*



4.4.3 Amélioration de la méthode : $X11 - ARIMA$ et $X12 - ARIMA$

La méthode $X12 - ARIMA$ est fondée sur le même principe que $X11$. Elle décompose la série selon un schéma additif (ou multiplicatif) : tendance + cycle saisonnier + aléa irrégulier. Les composantes sont obtenues à l'issue d'un processus itératif basé sur des lissages à l'aide de moyennes mobiles¹⁴.

Toutefois, la méthode $X11$ décrite dans la partie précédente pose des problèmes en début et en fin de série. Une des solutions est alors d'utiliser un modèle de type $ARIMA$ pour faire de la prévision de la série, et ainsi mieux s'ajuster sur les bornes de l'intervalle d'étude. La méthode la plus simple pour prolonger les séries ($X11$ tel que le fait SAS) est d'utiliser des moyennes mobiles asymétriques, en particulier les moyennes mobiles de Musgrave. Toutefois, les observations les plus récentes sont moins fiables que les données au centre.

L'extension $X11 - ARIMA$ (1975) puis $X12 - ARIMA$ a résolu en partie ces problèmes : la série initiale est modélisée par un processus $ARIMA$, puis prolongée au début et à la fin. L'erreur quadratique des estimation est ainsi minimisée.

De plus des améliorations ont été ajoutées, par exemple pour prendre en compte de la saisonnalité non-régulière. En particulier il est possible de prendre en compte l'effet Pâques, observé dans certaines série. Pâques est une fête ne tombant pas à la même date tous les ans. De nombreux modèles mathématiques ont été mis en place pour prévoir à l'avance la date des jours de Pâques¹⁵. Entre 1900 et 2100, le jour de Pâques tombera ainsi entre le 22 mars et le 25 avril. Or cette fête entraîne généralement un changement en niveau dans un certain nombre d'activité (à cause du Vendredi Saint, et du Lundi de Pâques férié), et ceci a des répercussions soit en mars, soit en avril (pour des données mensuelles), soit au premier, soit au second semestre (pour des données trimestrielles). Cet 'effet Pâques' a pu ainsi être mis en évidence sur des ventes en grande surface (chocolat, fleurs), mais aussi sur des nombres de mariages. Il aura aussi un impact si l'on considère des données mensuelles, obtenues comme agrégation de données journalières, puisqu'il faudra intégrer une pondération fonction du nombre de jours ouvrables.

4.4.4 Utilisation de la méthode $X11$ et comparaison avec les modèles $ARIMA$ saisonniers

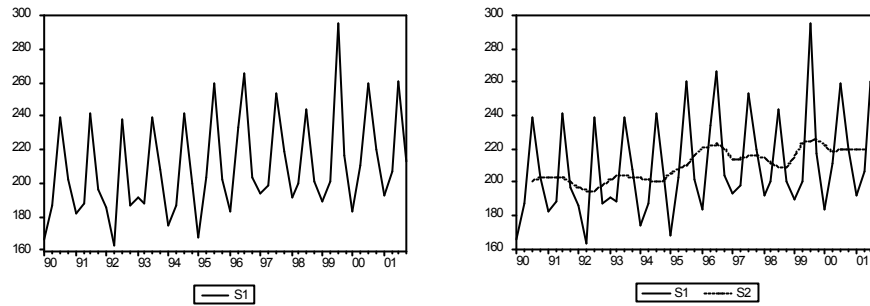
L'article de Cleveland et Tiao 'Decomposition of Seasonal Time Series : a model for the Census $X11$ program' (1976) étudie sur des cas simples l'application de la méthode $X11$.

¹⁴Dans le cas d'un schéma multiplicatif : une moyenne mobile d'ordre 12 (données mensuelles) fournit une première estimation de la tendance. La série initiale est alors divisée par cette estimation pour donner des rapports SI (saisonnalité/irrégulier). On applique alors une moyenne mobile 3×3 (par exemple) à chaque mois pour éliminer la composante irrégulière.

¹⁵Voir Gardner 'Mathematical Games' (1981), Scientific American (Février 81) pour plus d'information sur les différents algorithmes.

4.4.5 Exemple simple inspiré de la méthode X11

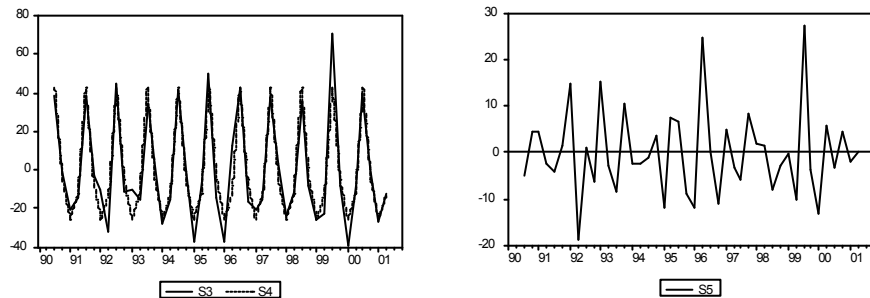
Considérons ici une série trimestrielle S_t^1 , observée sur 12 ans, et représentée ci-dessous à gauche,



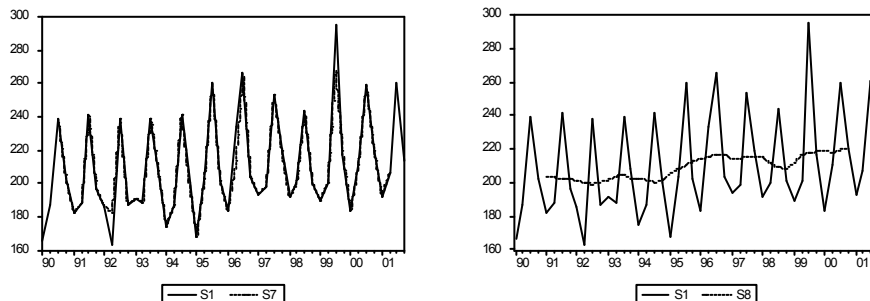
Appliquons à cette série un filtre moyenne mobile de façon à désaisonnaliser la série : $S_t^2 = M_{2 \times 2}(S_t^1)$, représentée ci-dessus à droite. La série ainsi obtenue est alors étudiée attentivement afin de repérer des points abérants. Pour cela, on considère la composante saisonnière, $S_t^3 = S_t^1 - S_t^2 = S_t^1 - M_{2 \times 2}(S_t^1)$. Afin de définir une courbe de référence, considérons la série de type Buys-Ballot associée, c'est à dire définie par la moyenne des composantes saisonnières. S_t^4 prend alors 4 valeurs, suivant la valeur de t [4] :

$$S_t^4 = \begin{cases} \left[S_{1991-01}^3 + S_{1992-01}^3 + \dots + S_{2001-01}^3 \right] / 11 & \text{si } t \text{ est au premier trimestre} \\ \left[S_{1991-02}^3 + S_{1992-02}^3 + \dots + S_{2001-02}^3 \right] / 11 & \text{si } t \text{ est au deuxième trimestre} \\ \left[S_{1990-03}^3 + S_{1991-03}^3 + \dots + S_{2000-03}^3 \right] / 11 & \text{si } t \text{ est au troisième trimestre} \\ \left[S_{1990-04}^3 + S_{1991-04}^3 + \dots + S_{2000-04}^3 \right] / 11 & \text{si } t \text{ est au quatrième trimestre} \end{cases}$$

Les deux composantes saisonnières ainsi définies sont représentées ci-dessous à gauche, ainsi que la différence $S_t^5 = S_t^4 - S_t^3$, à droite

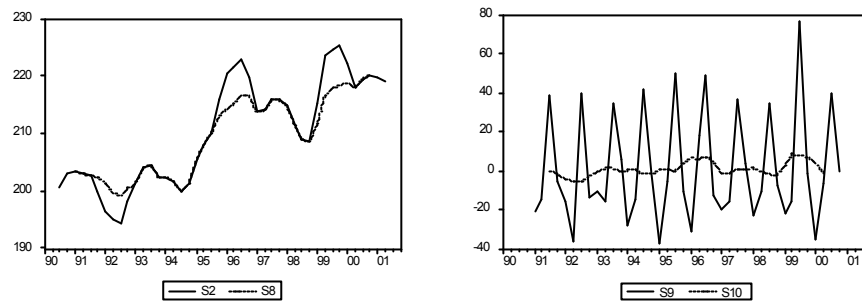


En se fixant un seuil a priori, on peut repérer trois points "abérants", au sens où $|S_t^5|$ est relativement grand. Ces points seront alors exclus de l'échantillon : on remplace $S_t^1 = S_t^2 + S_t^3$ par $S_t^7 = S_t^2 + S_t^4$ pour les trois valeurs abérantes repérées. La comparaison entre S_t^1 et S_t^7 est présentée ci-dessous à gauche. Cette série, corrigée des valeurs aberrantes, est alors lissée



Le graphique ci-dessous à gauche compare la série lissée obtenue sur la série brute (*en trait plein*), et sur la série corrigée des valeurs aberrantes (*en pointillés*). A partir de série lissée, S_t^8 , on peut en déduire la composante saisonnière, comme

différence avec la série initiale, $S_t^9 = S_t^1 - S_t^8$. Comme le suggère le graphique ci-dessous à droite, il est possible de lisser cette série afin de dégager la "vraie" composante saisonnière :



Comme le montre le graphique ci-dessus, à gauche, l'utilisation du lissage par moyenne mobile permet de repérer des points extrêmes et aberrants (*au sens statistiques, car ils peuvent avoir une explication exogène : météo, campagnes de publicité...etc.*), et en corrigeant la série initiale en excluant ces points, on peut mieux lisser la série.

4.5 Description du modèle *TRAMO/SEATS*

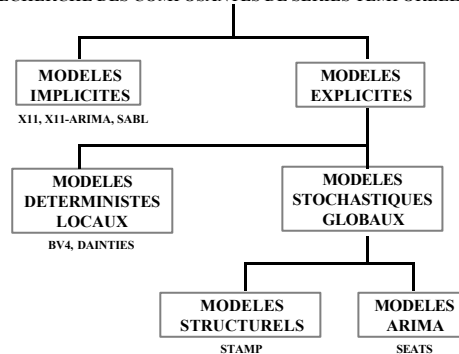
Le modèle *TRAMO/SEATS*¹⁶ est basé sur l'utilisation des modèles *ARIMA*, et a été développé par Maravall et Gomez, suite à l'algorithme de Burman en 1980. L'algorithme est alors le suivant :

- 1) un modèle *ARIMA* est ajusté automatiquement
- 2) les points aberrants sont identifiés : ces deux procédures sont faites par le module *TRAMO*
- 3) le module *SEATS* calcule la densité spectrale de la série linéarisée (*par le modèle ARIMA*)
- 4) la série est décomposée en un cycle et une tendance saisonnière : utilisation du filtre de Wiener-Kolmogorov¹⁷

Les hypothèses sont que la série linéarisée qui est traitée par *SEATS* peut s'écrire $Z_t = S_t + N_t$, où les deux composantes sont indépendantes : les deux séries suivent des modèles *ARIMA* (*inconnu*) dont les polynômes autorégressifs respectifs n'ont pas de racine commune.

Parmi les algorithmes récents de désaisonnalisation, on peut résumer la littérature suivant le schéma suivant

RECHERCHE DES COMPOSANTES DE SERIES TEMPORELLES



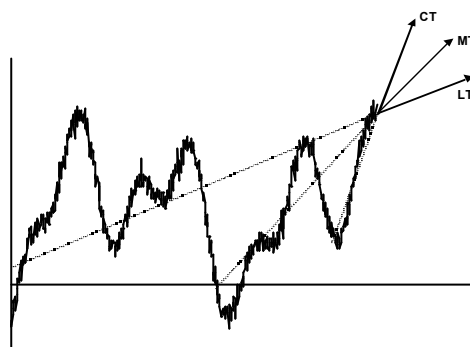
¹⁶*TRAMO* signifie *Time series Regression with ARIMA noise, Missing observations and Outliers*.

¹⁷Le filtre *WK* est détaillé dans '*Applied Time Series Analysis : Modelling, Forecasting, Unobserved Components Analysis and the Wiener-Kolmogorov Filter*' de C. Planas (1997)

” Les prévisions sont difficiles, surtout lorsqu’elles concernent l’avenir.”
 Jacques Chirac, Le Figaro - Février 1993

5 La prévision par lissage exponentiel

Les méthodes de lissages consistent à extrapoler une série en vue de faire de la prévision. Or comme on le voit sur l’exemple ci-dessous, une extrapolation simple (*linéaire en l’occurrence*) dépend fortement du type de résultats que l’on cherche à avoir : prévision à court, moyen, ou long terme



Ces trois méthodes diffèrent suivant le poids que l’on accorde aux observations passées.

5.1 Principe du lissage exponentiel simple

On dispose de N observations X_1, \dots, X_N . On souhaite prévoir, à la date $T = 1, \dots, N$, la valeur à un horizon 1, ou à un horizon quelconque h .

Définition 26 La prévision $\hat{X}_T(h)$ fournie par la méthode de lissage exponentiel simple, avec la constante de lissage β , $0 < \beta < 1$ est

$$\hat{X}_T(h) = (1 - \beta) \sum_{j=0}^{T-1} \beta^j X_{T-j}$$

On donne un poids d’autant moins important que les observations sont loins (*dans le passé*), avec une décroissance exponentielle :

- β proche de 1 : prise en compte de tout le passé
- β proche de 0 : prise en compte d’avantage des valeurs récentes (*plus sensible aux fluctuations*)

Remarque 20 Si β ne dépend pas de h , $\hat{X}_T(h)$ ne dépend pas de h , dont $\hat{X}_T(h) = \hat{X}_T$. Cette valeur \hat{X}_T est la prévision faite en T de la valeur en $T + 1$. Nous appellerons cette série \hat{X}_T (série lissée à la date t) ou F_{T+1} (valeur prédite pour la date $T + 1$).

Remarque 21 Pour certains logiciels permettant de faire du lissage exponentiel, la constante de lissage n’est pas β mais $\alpha = 1 - \beta$.

5.1.1 Méthode adaptative de mise à jour (ordre 1)

Propriété 18 *Méthode adaptative de mise à jour (ordre 1)*

$$\begin{aligned}\widehat{X}_T &= \widehat{X}_{T-1} + [1 - \beta] (X_T - \widehat{X}_{T-1}) \\ &= [1 - \beta] X_T + \beta \widehat{X}_{T-1} = \alpha X_T + [1 - \alpha] \widehat{X}_{T-1}\end{aligned}\quad (11)$$

Cette relation s'écrit également

$$F_{T+1} = \alpha X_T + (1 - \alpha) F_T$$

Propriété 19 \widehat{X}_T peut être vu comme une régression sur une constante, avec des pondérations exponentielles

Preuve. Le programme de minimisation

$$\min_c \left\{ \sum_{j=0}^{T-1} \beta^j (X_{T-j} - c)^2 \right\} \quad (12)$$

admet pour solution

$$\widehat{c} = \frac{1 - \beta}{1 - \beta^T} \sum_{j=0}^{T-1} \beta^j X_{T-j} \quad (13)$$

et pour T assez grand $\widehat{X}_T = \widehat{c}$. ■

5.1.2 Choix de la constante de lissage

Au delà des méthodes qualitative de rigidité ou de souplesse du modèle aux fluctuations conjoncturelles, il est possible d'utiliser des méthodes de type 'minimisation de la somme des carrés des erreurs de prévision' :

$$\widehat{\beta} = \arg \min \left\{ \sum_{t=1}^T \left[X_{t+1} - (1 - \beta) \sum_{j=0}^{t-1} \beta^j X_{t-j} \right]^2 \right\}$$

5.1.3 Lien entre robustesse de la prévision et choix de β

Il n'existe pas de relation a priori entre l'erreur de prévision et le paramètre β .

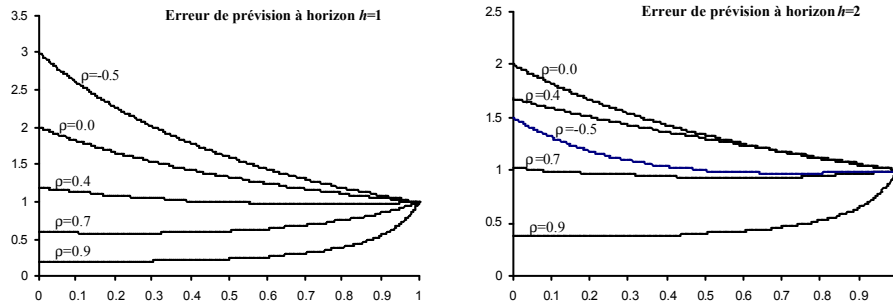
Exemple 35 Soit (X_t) un processus AR(1) de corrélation ρ , de variance 1, $X_t = \rho X_{t-1} + \varepsilon_t$. L'erreur de prévision à horizon h est

$$\Delta(\rho, \beta, h) = \mathbb{E} \left(\left[X_{T+h} - \widehat{X}_T(h) \right]^2 \right) \text{ avec } \widehat{X}_T(h) = (1 - \beta) \sum_{j=0}^{T-1} \beta^j X_{T-j}$$

On peut montrer que

$$\Delta(\rho, \beta, h) = \frac{2}{1 + \beta} + \frac{2(1 - \beta)(\beta\rho - \rho^h - \beta\rho^h)}{(1 + \beta)(1 - \beta\rho)}$$

L'erreur de prévision varie de la façon suivante en fonction de β



pour $\rho < 1/3$, l'erreur de prévision décroît avec β , et pour $\rho > 1/3$, l'erreur croît avec β .

5.1.4 Exemple d'application

Considérons ici une série de ventes d'une entreprise, sur 18 mois

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
C.A. 98	1293	1209	1205	1273	1220	1290	1243	1203	1390	1360	1353	1343
C.A. 99	1364	1330	1377	1332								

En septembre, une modification structurelle (*nouvelle unité de production*) a provoqué un saut des ventes. Les méthodes de lissage permettent d'effectuer de la prévision en intégrant ce genre de rupture. Une moyenne arithmétique (*non-pondérée, notée \bar{y}*) conduirait à sous-estimer les valeurs futures : il est alors naturel d'introduire des poids plus importants pour les valeurs récentes. La prévision à horizon 1 est alors

$$\bar{y}_T(1) = \frac{1}{T} \sum_{i=0}^{T-1} y_{T-i} \text{ et } \hat{y}_T(1) = \sum_{i=0}^{T-1} \alpha [1 - \alpha]^i y_{T-i} = \sum_{i=0}^{T-1} [1 - \beta]^i y_{T-i} \text{ pour } \alpha, \beta \in]0, 1[\text{ et } \beta = 1 - \alpha$$

cette dernière valeur étant obtenue par lissage exponentiel (*simple*). Nous noterons ici $\hat{y}_1, \dots, \hat{y}_T$ la série lissée, et F_1, \dots, F_T la série des valeurs prédites.

Mise en place de l'algorithme (α fixé) Pour la première valeur ($T = 0$), on considère comme valeur initiale une moyenne des premières valeurs observées. EVIEWS considère une moyenne sur les 8 premiers mois,

$$F_1 = \hat{y}_0 = \frac{1}{8} (1293 + \dots + 1203) = 1242$$

Pour construire la série lissée, à α fixé, on utilise la relation de mise à jour

$$\hat{y}_j = \alpha y_j + (1 - \alpha) \hat{y}_{j-1} \text{ ou } F_{j+1} = \alpha y_j + (1 - \alpha) F_j$$

Avec comme constante de lissage $\alpha = 0.3$, on obtient

$$F_2 = \hat{y}_1 = \alpha y_1 + (1 - \alpha) F_1 = 0.3 * 1293 + 0.7 * 1242 = 1257.3$$

puis

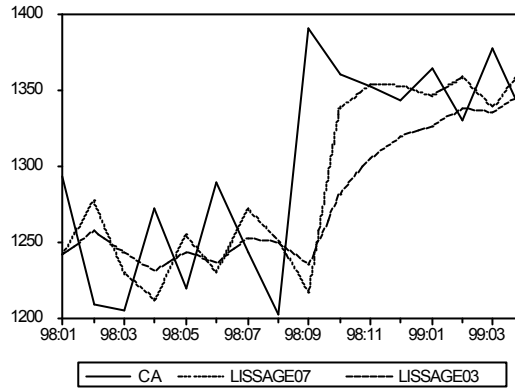
$$F_3 = \alpha y_2 + (1 - \alpha) F_2 = 0.3 * 1209 + 0.7 * 1257.3 = 1242.81$$

Comme on peut le voir, nous estimation pour la date 2 était de 1257.3. Or la vraie valeur était plus faible, à savoir 1209. Aussi, pour la date 3, la prévision sera une correction de ce 1257.3 en prenant en compte (*avec un poids correspondant à la valeur α*) l'erreur qui avait été faite : en l'occurrence, F_3 sera plus faible que F_2 (*la différence étant $\alpha [F_2 - y_2]$*)

	y_{j-1}	F_{j-1}	F_j
1	1293		1242.00
2	1209	1242.00	1257.30
3	1205	1257.30	1242.81
4	1273	1242.81	1231.47
5	1220	1231.47	1243.93

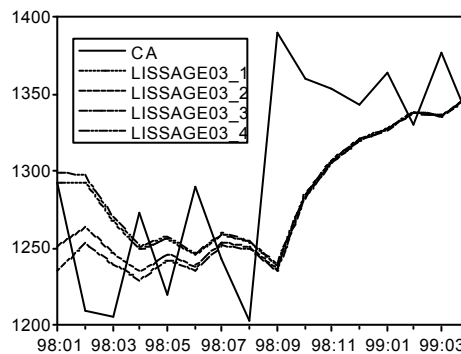
(*puisque $F_j = \alpha y_{j-1} + (1 - \alpha) F_{j-1}$*) d'où finalement la série lissée exponentiellement pour 1998

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
y_j	1293	1209	1205	1273	1220	1290	1243	1203	1390	1360	1353	1343
$\hat{y}_j (\alpha = 0.3)$	1242	1257	1243	1231	1244	1237	1253	1250	1236	1282	1305	1320
$\hat{y}_j (\alpha = 0.7)$	1242	1278	1230	1212	1255	1230	1272	1252	1218	1338	1353	1353



On peut noter que plus α est proche de 1, plus la courbe lissée colle aux données (\hat{y}_j est proche de y_j) : pour $\alpha = 1$, la prévision F_{j+1} sera la dernière valeur observée (y_j). Un coefficient de lissage plus faible (par exemple $\alpha = 0.3$) permet en revanche de bien lisser les aléas importants de la série. La série lissée s'adapte également au changement de niveau observé en septembre. Toutefois, cette adaptation se fait d'autant plus lentement que α est faible : les prévisions sont alors biaisées (*sous-estimation dans cet exemple*) pendant la période d'adaptation, comme on peut le voir sur le graphique ci-dessous : dès octobre 1998, la prévision faite avec un coefficient de 0.7 avait atteint un niveau "correct", alors que la prévision avec une pondération de 0.3 est plus lente (*ce qui rend la prévision moins sensible à un choc exogène : si cette variation en septembre n'avait été qu'un choc, une faible pondération aurait permis de ne pas le prendre en compte*).

On peut noter que le lissage dépend non seulement de α mais aussi de la valeur initiale choisie. Comme le montre le graphique de gauche, on observe une convergence (*asymptotique*), avec des valeurs très proches pour T proche de 16. En choisissant une valeur initiale proche des premières valeurs de la série ($\hat{y}_0 = y_1$ ou \hat{y}_0 moyenne des 2 ou 3 premières observations), on s'aperçoit que la courbe observée et les courbes lissées sont quasiment confondues au bout de 10 mois (*cette valeur dépendant de α : la convergence est d'autant plus rapide que α est grand*). Le graphique ci-dessous correspond à plusieurs initialisations : $F_1 = y_1$ (prévision parfaite - (1)), $F_1 = (y_1 + y_2) / 2$ (moyenne des deux premières valeurs - (2)), $F_1 = (y_1 + y_2 + y_3) / 2$ (moyenne des trois premières valeurs - (3)) et $F_1 = (y_1 + \dots + y_n) / n$ (moyenne de l'échantillon - (4)).



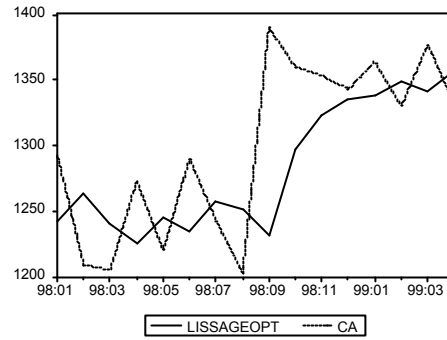
Remarque 22 Il convient de faire attention : dans la littérature, les courbes lissées sont soit F_t , soit \hat{y}_{t-1} . Certains auteurs décalent ainsi (dans les tableaux ou les graphiques) la courbe lissée.

A retenir 4 La formule itérative pour construire la série lissée de X_t pour $t = 1, \dots, N$ est la suivante

$$\begin{cases} F_0 = X_1 \text{ ou } [X_1 + \dots + X_p] / p \\ F_{t+1} = \alpha X_t + (1 - \alpha) F_t \text{ pour } 0 \leq t \leq N \\ F_t = F_{N+1} \text{ pour } t \geq N + 1 \end{cases}$$

Choix de la constante de lissage Ce choix peut relever de considérations empiriques : des fortes pondérations pour les valeurs récentes (α élevé) donne de meilleures prévisions à court terme qu'à long terme. Toutefois, une des méthodes les plus utilisées est la minimisation des moindres carrés des erreurs (*prévision/réalisation*) à un horizon $h = 1$.

L'algorithme (13) donne ici un paramètre $\alpha = 0.418$, qui correspond à une somme des erreurs de prévision de 48178, c'est à dire un écart type de l'erreur valant 54.874.



Remarque 23 *L'intervalle de confiance de la prévision est alors de la forme*

$$\widehat{X}_T(h) \pm 1.96\sigma_X C_h \text{ où } C_h^2 = 1 + \frac{1-\beta}{(1+\beta)^3} \left[(1+4\beta+5\beta^2) + 2h(1-\beta)(1+3\beta) + 2h^2(1-\beta)^2 \right]$$

5.2 Principe de lissage exponentiel double

Le lissage exponentiel *simple* est adapté à des séries pouvant être ajustée par une constante au voisinage de T . Le principe de lissage exponentiel *double* permet de faire un ajustement par une droite, à savoir approcher X_t par Y_t où

$$Y_t = A + (t - T) B$$

La prévision à horizon h s'écrit

$$F_{T+h} = \widehat{X}_T(h) = \widehat{A}(T) + h\widehat{B}(T)$$

De même que pour (12) le programme d'optimisation pour estimer A et B s'écrit

$$\min_{A,B} \left\{ \sum_{j=0}^{T-1} \beta^j (X_{T-j} - [A + (T-j) B])^2 \right\} \quad (14)$$

Théorème 4 *La solution de (14) est donnée par*

$$\widehat{A}(T) = 2S_1(T) - S_2(T) \text{ et } \widehat{B}(T) = \frac{1-\beta}{\beta} [S_1(T) - S_2(T)]$$

en posant

$$S_1(t) = (1-\beta) \sum_{k=0}^{t-1} \beta^k X_{t-k} = (1-\beta) X_t + \beta S_1(t-1) \text{ (série lissée)}$$

$$\begin{aligned} S_2(t) &= (1-\beta) \sum_{k=0}^{t-1} \beta^k S_1(t-k) = (1-\beta) S_1(t) + \beta S_2(t-1) \text{ (série lissée 2 fois)} \\ &= (1-\beta)^2 \sum_{k=0}^{t-1} \sum_{i=0}^{t-k-1} \beta^{i+k} X_{t-(k+i)} \end{aligned}$$

Preuve. Gouriéroux et Monfort (1995) pages 110-111 ■

5.2.1 Méthode adaptative de mise à jour (ordre 1)

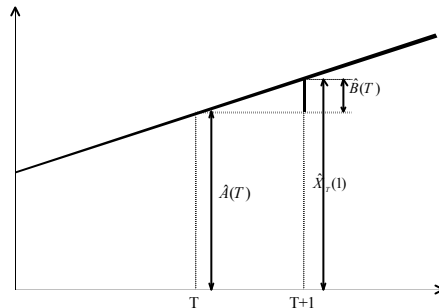
Pour obtenir la formule de mise à jour (à l'ordre 1) permettant de passer de T à $T+1$, on peut utiliser le résultat suivant

Propriété 20 Si à la date T , $F_{T+1} = \hat{X}_T(1) = \hat{A}(T) + \hat{B}(T)$, alors, en $T+1$

$$\begin{cases} \hat{A}(T+1) = (1 - \beta^2) [X_{T+1} - \hat{X}_T(1)] + \hat{A}(T) + \hat{B}(T) \\ \hat{B}(T+1) = \hat{B}(T) + (1 - \beta^2) [X_{T+1} - \hat{X}_T(1)] \end{cases} \quad (15)$$

Preuve. Gouriéroux et Monfort (1995) pages 112-113 ■

Dans le cas d'une prévision parfaite, i.e. $X_{T+1} = \hat{X}_T(1)$, on aurait $\hat{A}(T+1) = \hat{A}(T) + \hat{B}(T)$ et $\hat{B}(T+1) = \hat{B}(T)$



Dans ce cas, les droites de prévision en T et en $T+1$ sont les mêmes, et la pente, en particulier, est inchangée ($\hat{B}(T+1) = \hat{B}(T)$).

Remarque 24 L'intervalle de confiance de la prévision est alors de la forme

$$\hat{X}_T(h) \pm 1.96\sigma_X \sqrt{\frac{2\beta}{2\beta - 1}}$$

5.2.2 Application de la méthode de lissage exponentiel double

Considérons la série suivante, correspondant à un indice d'activité

	Trim 1	Trim 2	Trim 3	Trim 4
1982	—	9050	9380	9378
1983	9680	10100	10160	10469
1984	10738	10910	11058	11016
1985	10869	11034	11135	10845
1986	11108	11115	11424	10895
1987	11437	11352	11381	11401
1988	11507	11453	11561	—

Le lissage exponentiel double est très proche du lissage exponentiel simple, sauf que l'on fait un ajustement au voisinage de T non plus par une constante, mais par une droite. En fait, la série (correspondant à un indice) est une série "croissante" : l'ajustement par lissage exponentiel simple sous-estimerait les valeurs réalisées (graphique page 5.2.2). Le programme de minimisation s'écrit ici

$$\min_{A,B} \left\{ \sum_{j=0}^{T-1} \alpha^j (X_{t-j} - [A_T + B_T(T-j)])^2 \right\}$$

La prévision à horizon h est alors $\hat{y}_T(h) = A_T + B_T h$. Trois formulations sont possibles pour écrire la série lissée

Formulation classique - Les coefficients A_j et B_j sont donnés par

$$A_j = 2\hat{y}_j^1 - \hat{y}_j^2 \text{ et } B_j = \frac{\alpha}{1 - \alpha} [\hat{y}_j^1 - \hat{y}_j^2]$$

où les \hat{y}_j^1 et \hat{y}_j^2 sont obtenus récursivement par deux lissages consécutifs,

$$\begin{cases} \hat{y}_j^1 = \alpha y_j + (1 - \alpha) \hat{y}_{j-1}^1 : \text{lissage exponentiel simple de } y_j \\ \hat{y}_j^2 = \alpha \hat{y}_j^1 + (1 - \alpha) \hat{y}_{j-1}^2 : \text{lissage exponentiel simple de } \hat{y}_j^1 \end{cases}$$

Formules de lissage direct - l'erreur de lissage e_j est donnée par $e_j = y_j - \hat{y}_j = y_j - [A_{j-1} + B_{j-1}]$, et donc

$$\begin{cases} A_j = A_{j-1} + B_{j-1} + [1 - (1 - \alpha)^2] e_j \\ B_j = B_{j-1} + \alpha^2 e_j \end{cases}$$

ce qui donne une relation permettant d'obtenir récursivement les A_i et les B_i .

Formules de mise à jour - cette expression est en fait la même que la précédente, sauf que l'on remplace l'erreur de prévision par la dernière observation y_j ,

$$\begin{cases} A_j = \lambda y_j + (1 - \lambda) [A_{j-1} + B_{j-1}] \\ B_j = \mu [A_j - A_{j-1}] + (1 - \mu) B_{j-1} \end{cases} \quad \text{où } \lambda = 1 - (1 - \alpha)^2 \text{ et } \mu = \frac{\alpha}{2 - \alpha} \quad (16)$$

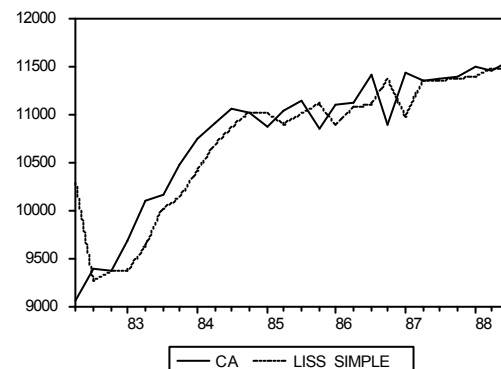
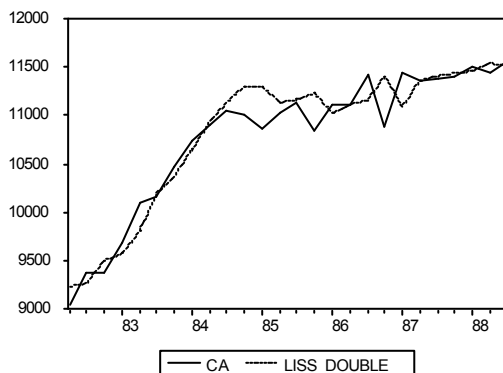
Remarque 25 A_j et B_j sont utilisés pour calculer \hat{y}_j , prévision à horizon 1 faite à la date j , soit F_{j+1} .

Encore une fois, l'initialisation de l'algorithme est important. Une méthode possible est de considérer comme valeur initiale pour A_1 la première valeur y_1 . La pente B_1 peut alors être choisie comme la pente moyenne entre la date 1 et une date t_0 , telle que $B_1 = [y_{t_0} - y_1] / t_0$.

Dans le cas qui nous intéresse on obtient la série lissée suivante, en prenant comme constante de lissage $\alpha = 0.384$, et comme valeurs initiales de $A_0 = y_1$ et B_0 la pente sur une période de 10 observations (*soient* 9050 et 177) - à l'aide de la relation (16), et $\lambda = 0.6205$ et $\mu = 0.2376$

	y_j	A_j	B_j	F_{j+1}
1982 - 2	9050	9050.00	165.30	9215.30
1982 - 3	9380	9112.73	140.93	9253.66
1982 - 4	9378	9332.05	159.56	9491.61
1983 - 1	9680	9421.11	142.81	9563.92

Série observée					Série lissée				
	Trim 1	Trim 2	Trim 3	Trim 4		Trim 1	Trim 2	Trim 3	Trim 4
1982	—	9050	9380	9378	1982	—	9215	9254	9492
1983	9680	10100	10160	10469	1983	9564	9796	10189	10372
1984	10738	10910	11058	11016	1984	10646	10932	11143	11303
1985	10869	11034	11135	10845	1985	11295	11138	11166	11234
1986	11108	11115	11424	10895	1986	11023	11118	11159	11405
1987	11437	11352	11381	11401	1987	11095	11364	11412	11443
1988	11507	11453	11561	—	1988	11461	11541	11524	—

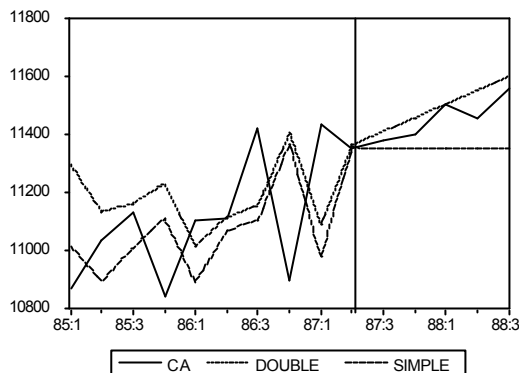


A titre de comparaison, nous avons ajouté à droite le lissage exponentiel simple optimal qui aurait été obtenu sur les mêmes données. Ce lissage simple est relativement mal adapté à ce type de données (*croissantes*) puisque nous allons continuellement sous-évaluer la vraie valeur en période de croissance forte.

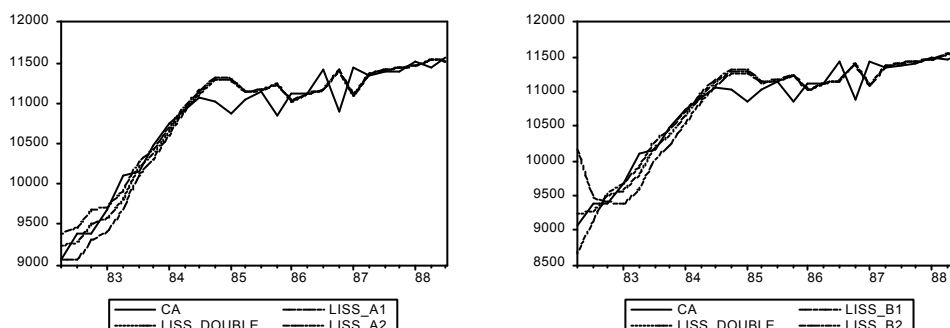
Supposons que la série ait été observée jusqu'au troisième trimestre de 1987. La série lissée jusqu'à cette date reste la même, et les prévisions pour les trimestres suivant aurait été obtenus en utilisant $A = A_{1987-3} = 11412$, $B = B_{1987-3} = 47.02$, et $\hat{y}_T(h) = A + Bh$

	réalisation	prévision (double)	prévision (simple)
1987-4	11401	11459	11352
1988-1	11507	11506	11352
1988-2	11453	11553	11352
1988-3	11561	11600	11352

ce qui est représenté sur les graphiques ci-dessous,



Les graphiques ci-dessous représentent l'impact des valeurs initiales, avec à gauche, un changement de A_0 et à droite un changement de B_0 . Comme on peut le remarquer, l'ajustement se fait plus rapidement que pour le lissage exponentiel simple



A retenir 5 La formule itérative pour construire la série lissée de X_t pour $t = 1, \dots, N$ est la suivante

$$\begin{cases} S_0^1 = X_1 \text{ ou } [X_1 + \dots + X_p]/p \\ S_0^2 = 0 \\ S_{t+1}^1 = \alpha X_t + (1 - \alpha) S_t^1 \text{ pour } 0 \leq t \leq N \\ S_{t+1}^2 = \alpha S_t^1 + (1 - \alpha) S_t^2 \text{ pour } 0 \leq t \leq N \\ A_{t+1} = 2S_{t+1}^1 - S_{t+1}^2 \text{ pour } 0 \leq t \leq N \\ B_{t+1} = \alpha [S_{t+1}^1 - S_{t+1}^2] / (1 - \alpha) \\ F_{t+1} = A_{t+1} + B_{t+1} \text{ pour } 0 \leq t \leq N \\ F_t = A_{N+1} + (t - N - 1) B_{N+1} \text{ pour } t \geq N + 1 \end{cases}$$

5.3 Lissage exponentiel multiple, ou généralisé

Cette généralisation a été proposée par Brown en 1962, permettant d'ajuster au voisinage de T une fonction plus complexe qu'une fonction affine. La résolution de ce problème repose sur la notion de vecteurs de fonctions à matrice de transition fixe.

Définition 27 Le vecteur $f(t) = [f_1(t), \dots, f_n(t)]'$, où $t \in \mathbb{Z}$ est dit à **matrice de transition fixe** s'il existe une matrice A régulière telle que

$$f(t) = Af(t-1) \text{ pour tout } t \in \mathbb{Z}$$

La méthode du lissage exponentiel généralisé consiste à ajuster au voisinage de T de la série X_t une fonction $\phi(t-T)$ de la forme

$$\phi(t) = \sum_{i=1}^n \alpha_i f_i(t) \text{ où } f_i(\cdot) \text{ est à matrice de transition fixe}$$

Cette classe de fonction $\phi(\cdot)$ comprend la plupart des fonctions usuelles.

(1) **Les fonctions constantes** - $\phi(t) = c$, obtenues avec $f(t) = 1$ et $A = 1$. Dans ce cas, on retrouve le principe de lissage exponentiel simple,

(2) **Les fonctions linéaires** - $\phi(t) = \alpha + \beta t$, obtenues avec $f(t) = [1, t]'$ de matrice de transition

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \text{ puisque } \begin{bmatrix} 1 \\ t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ t-1 \end{bmatrix}$$

Dans ce cas, on retrouve le principe de lissage exponentiel double,

(3) **Les fonctions polynômiales de degré p** - Cette famille est obtenue en prenant comme base une base de $\mathbb{R}_p(X)$ (espace des polynômes de degré inférieur ou égal à p). En particulier, on peut choisir la base

$$\mathcal{B}_p = \left\{ P_k(t) = \frac{1}{k!} t(t-1) \dots (t-k+1), k = 1, \dots, p+1 \right\}$$

obtenue à l'aide du triangle de Pascal, et définie par récurrence par

$$P_k(t) = P_{k-1}(t-1) + P_k(t-1) \text{ pour } k > 1$$

Le vecteur $f(t) = [P_1(t), \dots, P_{p+1}(t)]$ est alors de matrice de transition (fixe)

$$A = \begin{bmatrix} 1 & 0 & 0 & & 0 & 0 \\ 1 & 1 & 0 & & 0 & 0 \\ 0 & 1 & 1 & & 0 & 0 \\ & & & \ddots & \ddots & \\ 0 & 0 & 0 & \ddots & 1 & 0 \\ 0 & 0 & 0 & & 1 & 1 \end{bmatrix}$$

(4) **Les fonctions sinusoïdales** - Les fonctions $\phi(t) = \alpha \sin \omega t + \beta \cos \omega t$ sont obtenues en prenant $f(t) = [\sin \omega t, \cos \omega t]'$, et dans ce cas

$$A = \begin{bmatrix} \cos \omega & \sin \omega \\ -\sin \omega & \cos \omega \end{bmatrix}$$

(6) **Les fonctions exponentielles** - Les fonctions $\phi(t) = \alpha \exp(\beta t)$ sont obtenues en prenant $f(t) = \exp(\beta t)$ et dans ce cas $A = \exp(\alpha)$.

Cette méthode de lissage se met en place de la façon suivante.

De la même façon que (14), le programme d'optimisation s'écrit

$$\min_a \left\{ \sum_{j=0}^{T-1} \beta^j (X_{t-j} - f'(-j) a)^2 \right\} \quad (17)$$

où la notation f' désigne la transposée de f . Posons

$$x = \begin{bmatrix} X_T \\ \vdots \\ X_1 \end{bmatrix}, F = \begin{bmatrix} f_1(0) & f_n(0) \\ \vdots & \vdots \\ f_1(-T+1) & f_n(-T+1) \end{bmatrix} = \begin{bmatrix} f'(0) \\ \vdots \\ f'(-T+1) \end{bmatrix} \text{ et } = \text{diag}(1, 1/\beta, \dots, 1/\beta^{T-1})$$

Le programme (17) correspond à la régression (*linéaire*) de x sur les colonnes de F , associée à la matrice de covariance . On en déduit que la solution à (17) est unique, et est donnée par

$$\hat{a}(T) = (F' \quad {}^{-1}F)^{-1} F' \quad y = [M(T)]^{-1} Z(T)$$

où

$$M(T) = F' \quad {}^{-1}F = \sum_{j=0}^{T-1} \beta^j f(-j) f'(-j) \text{ et } Z(T) = F' \quad y = \sum_{j=0}^{T-1} \beta^j f(-j) X_{T-j}$$

La matrice $M(T)$ converge vers une matrice finie M quand $T \rightarrow +\infty$: on peut estimer $\hat{a}(T)$ en utilisant cette matrice limite,

$$\hat{a}(T) = M^{-1} Z(T) \text{ avec } M(T) = \sum_{j=0}^{\infty} \beta^j f(-j) f'(-j)$$

Et la prévision à horizon h faite à la date T est

$$\hat{X}_T(h) = f'(h) \hat{a}(T)$$

5.3.1 Méthode adaptative de mise à jour (ordre 1)

Pour cela, notons que

$$Z(T+1) = X_{T+1}f(0) + \beta A^{-1}Z(T)$$

et on peut alors écrire

$$\hat{a}(T+1) = X_{T+1}M^{-1}f(0) + \beta M^{-1}A^{-1}M\hat{a}(T)$$

que l'on peut encore noter

$$\hat{a}(T+1) = \lambda X_{T+1} + \mu \hat{a}(T) \quad \text{où} \quad \begin{cases} \lambda = M^{-1}f(0) \\ \mu = \beta M^{-1}A^{-1}M \end{cases}$$

où les matrices λ et μ sont indépendantes de T . Cette relation peut se mettre sous la forme suivante, proche de (11),

$$\hat{a}(T+1) = A'\hat{a}(T) + \lambda [X_{T+1} - \hat{X}_T(1)]$$

5.4 Les méthodes de Holt-Winters (1960)

5.4.1 Méthode non saisonnière

Cette méthode est une généralisation de la méthode de lissage exponentiel mais avec un point de vue différent de celui introduit dans le lissage exponentiel généralisé. De la même façon que pour le lissage exponentiel double, l'ajustement se fait de façon linéaire au voisinage de T , la nuance se faisant au niveau de formules de mise à jour, différentes de (15) :

$$\begin{cases} \hat{A}(T+1) = (1-\lambda)X_{T+1} + \lambda [\hat{A}(T) + \hat{B}(T)] & \text{où } 0 < \lambda < 1 \\ \hat{B}(T+1) = (1-\mu) [\hat{A}(T+1) - \hat{A}(T)] + \mu \hat{B}(T) & \text{où } 0 < \mu < 1 \end{cases} \quad (18)$$

La première relation est une moyenne pondérée de deux informations sur $A(T)$, correspondant au niveau de la série à la date T : l'observation X_{T+1} et la prévision faite en T ($\hat{A}(T) + \hat{B}(T)$). La seconde relation s'interprète comme une moyenne pondérée de deux informations sur $B(T)$, correspondant à la pente de la série à la date T : la différence entre les niveaux estimés en T et $T+1$, et la pente estimée en T .

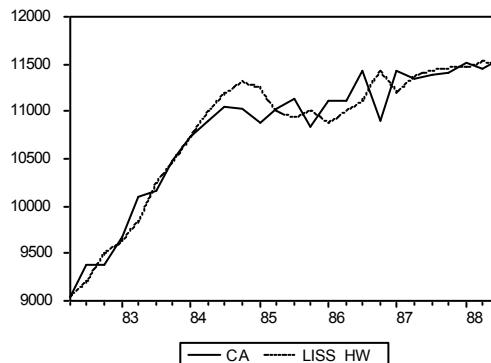
Toutefois, ces deux relations ne peuvent être utilisées qu'après initialisation, que l'on fera généralement de la façon suivante : $\hat{A}(2) = X_2$ et $\hat{B}(2) = X_2 - X_1$. La prévision à horizon h faite à la date T est donnée par

$$\hat{X}_T(h) = \hat{A}(T) + h\hat{B}(T)$$

Cette méthode peut être vue comme une généralisation du lissage exponentiel double, qui ne faisait intervenir qu'un coefficient, β (ou α). Cette dernière méthode correspond au cas particulier

$$\lambda = \beta^2 \quad \text{et} \quad \mu = 1 - \frac{(1-\beta)^2}{1-\beta^2} = \frac{2\beta}{1+\beta}$$

Exemple 36 Sur l'exemple précédant, on obtient



5.4.2 La méthode saisonnière additive

On suppose ici que la série (X_t) peut être approchée au voisinage de T par la série

$$Y_t = A + (t - T)B + S_t$$

où S_t est un facteur saisonnier. Les formules de mise à jour s'écrivent de la façon suivante, où s est le facteur de saisonnalisation (ou le nombre de saisons : $s = 4$ pour des données trimestrielles ou $s = 12$ pour des données mensuelles)

$$\begin{cases} \widehat{A}(T+1) = (1-\lambda)[X_{T+1} - S_{T+1-s}] + \lambda[\widehat{A}(T) + \widehat{B}(T)] & \text{où } 0 < \lambda < 1 & \text{(lissage de la moyenne)} \\ \widehat{B}(T+1) = (1-\mu)[\widehat{A}(T+1) - \widehat{A}(T)] + \mu\widehat{B}(T) & \text{où } 0 < \mu < 1 & \text{(lissage de la tendance)} \\ \widehat{S}_{T+1} = (1-\nu)[X_{T+1} - \widehat{A}(T+1)] + \nu S_{T+1-s} & \text{où } 0 < \nu < 1 & \text{(lissage de la saisonnalité)} \end{cases}$$

et la prévision à horizon h ($1 \leq h \leq s$) s'écrit

$$\widehat{X}_T(h) = \widehat{A}(T) + h\widehat{B}(T) + \widehat{S}_{T+h+s}$$

La encore, le problème d'initialisation va se poser, et on peut prendre

$$\begin{cases} \widehat{A}(s) = M_s(X_1, \dots, X_s) & \text{où } M_s \text{ est une moyenne pondérée} \\ \widehat{A}(s+1) = M_s(X_2, \dots, X_{s+1}) \\ \widehat{B}(s+1) = \widehat{A}(s+1) - \widehat{A}(s) \\ \widehat{S}_i = X_i - \widehat{A}(i) \end{cases}$$

Remarque 26 *Le lissage exponentiel est utilisable sous EViews,*

A gauche, on voit comment trouver l'option de lissage exponentiel. EViews propose alors plusieurs méthodes (fenêtre au centre) : lissage exponentiel simple, double, ou de Holt Winters. Les résultats sont alors présentés sous la forme de droite. EViews donne ici les paramètres λ (alpha) et μ (beta), ainsi que la variance de l'erreur de prévision.

5.5 Exemple de mise en pratique des méthodes de lissage

Comme nous allons le voir ici, les méthodes de lissage, afin de faire de la prévision, peuvent très facilement être mises en oeuvre, en particulier sur des tableurs (*Excel par exemple*). Nous allons voir comment faire de la prévision sur des données comportant de la saisonnalité, à l'aide des méthodes de lissage exponentiel.

5.5.1 Présentation des données

Les données sont mensuelles, et comportent une saisonnalité annuelle,

DATE	VENTES	MOYENNE MOBILE	DIFFERENCE	SAISONNALITE	VENTES CVS
Jan-99	2,853,123			126,280	2,726,843
Feb-99	2,797,469			78,024	2,719,445
Mar-99	2,773,701			20,897	2,752,804
Apr-99	2,735,895			-62,580	2,798,475
May-99	2,695,410			-121,586	2,816,996
Jun-99	2,652,799			-169,498	2,822,297
Jul-99	2,705,817	2,844,752	-138,935	-138,159	2,843,976
Aug-99	2,795,698	2,864,431	-68,733	-107,685	2,903,383
Sep-99	2,955,251	2,885,559	69,692	68,668	2,886,583
Oct-99	3,007,658	2,904,892	102,766	104,799	2,902,859
Nov-99	3,020,084	2,920,910	99,174	97,431	2,922,653
Dec-99	3,032,833	2,934,917	97,916	103,410	2,929,423
Jan-00	3,075,696	2,947,308	128,388	126,280	2,949,416
Feb-00	3,047,187	2,954,786	92,401	78,024	2,969,163
Mar-00	3,031,057	2,960,304	70,753	20,897	3,010,160
Apr-00	2,942,528	2,967,190	-24,662	-62,580	3,005,108
May-00	2,873,213	2,972,169	-98,956	-121,586	2,994,799
Jun-00	2,811,176	2,975,988	-164,812	-169,498	2,980,674
Jul-00	2,844,805	2,978,711	-133,906	-138,159	2,982,964
Aug-00	2,836,192	2,979,352	-143,160	-107,685	2,943,877
Sep-00	3,047,197	2,976,077	71,120	68,668	2,978,529
Oct-00	3,080,972	2,970,664	110,309	104,799	2,976,173
Nov-00	3,066,260	2,967,094	99,166	97,431	2,968,829
Dec-00	3,078,322	2,965,940	112,382	103,410	2,974,912
Jan-01	3,095,541	2,967,892	127,649	126,280	2,969,261
Feb-01	3,042,742	2,975,618	67,124	78,024	2,964,718
Mar-01	2,956,884	2,982,365	-25,481	20,897	2,935,987
Apr-01	2,886,788	2,983,808	-97,020	-62,580	2,949,368
May-01	2,843,283	2,984,021	-140,738	-121,586	2,964,869
Jun-01	2,813,409	2,984,116	-170,707	-169,498	2,982,907
Jul-01	2,889,432		-138,159	3,027,591	
Aug-01	2,976,979		-107,685	3,084,664	
Sep-01	3,068,328		68,668	2,999,660	
Oct-01	3,094,477		104,799	2,989,678	
Nov-01	3,057,867		97,431	2,960,436	
Dec-01	3,088,998		103,410	2,985,588	
Moyenne	2,932,641				2,932,641
Ecart-Type	133,328				85,190
Rapport	4.546%				2.905%

Pour lisser cette série, nous avons utilisé une moyenne mobile permettant d'annuler une saisonnalité annuelle (*période* 12) : $M_{2 \times 12}$. En particulier

$$MX_t = \frac{1}{12} \left[\frac{1}{2} X_{t-6} + X_{t-5} + X_{t-4} + \dots + X_t + \dots + X_{t+4} + X_{t+5} + \frac{1}{2} X_{t+6} \right]$$

et ainsi, sur l'exemple ci-dessus, on peut noter que

$$2\,864\,431 = \frac{1}{12} \left[\frac{2\,797\,469}{2} + 2\,773\,701 + 2\,735\,895 + \dots + 2\,795\,698 + \dots + 3\,032\,833 + 3\,075\,696 + \frac{3\,047\,187}{2} \right]$$

La composante saisonnière est alors obtenue comme différence entre la série brute X_t et la série désaisonnalisée MX_t . Cette série correspond dans le tableau ci-dessus à la variable DIFFERENCE. Par exemple,

$$92\,401 = 3\,047\,187 - 2\,954\,786$$

On considère alors la différence moyenne pour chaque mois :

$$\begin{aligned} \text{JAN} \quad \Delta^1 &= \frac{1}{2} [(X_{\text{JAN}00} - MX_{\text{JAN}00}) + (X_{\text{JAN}01} - MX_{\text{JAN}01})] \\ \text{FEV} \quad \Delta^2 &= \frac{1}{2} [(X_{\text{FEV}00} - MX_{\text{FEV}00}) + (X_{\text{FEV}01} - MX_{\text{FEV}01})] \end{aligned}$$

Par exemple, la différence moyenne pour Septembre est

$$\Delta^9 = \frac{1}{2} (69\,692 + 71\,120) = 70\,406$$

On peut noter que la somme de ces différences moyennes Δ^i ne vaut pas 0 : on va alors normaliser les Δ^i de façon à ce que leur somme soit nulle. On considère alors

$$\delta^i = \Delta^i - \frac{1}{12} \sum_{j=1}^{12} \Delta^j$$

Dans l'exemple considérée, la somme des Δ^i valait 20 865 : aussi, on va définir $\delta^i = \Delta^i - 20\,865/12$. Cette série δ^i va alors constituer la composante saisonnière de la série (X_t). Par exemple, la saisonnalité pour le mois de mars est -62

580. Et à partir de là, on construit la série corrigée des variations saisonnières (*CVS*) comme différence entre (X_t) et la composante saisonnière du mois correspondant.

C'est à partir de cette série corrigée des variations saisonnières que l'on va faire de la prévision. La méthodologie est alors la suivante. On sépare la série initiale (X_t) de la façon suivante :

$$X_t = Z_t + \Sigma_t = \text{série CVS} + \text{composante saisonnière}$$

et la prévision sera alors faite en considérant

$$\hat{X}_t = \hat{Z}_t + \Sigma_t$$

On extrapole la série corrigée des variations saisonnières (*par lissage*), et on rajoute ensuite la composante saisonnière.

5.5.2 Lissage "linéaire"

La méthode la plus simple pour faire de la prévision sur une série désaisonnalisée est d'utiliser une régression linéaire,

DATE	VENTES	SÉRIE CVS	APPROX. LINEAIRE	ERREUR	COMPOSANTE SAISONNIERE	PREVISION (constante)	PREVISION (droite)	ERREUR
Jan-99	2.853.123	2.726.843	2.820.908	-94.064	126.280	2.947.187	-107.972	-205.798
Feb-99	2.797.469	2.719.445	2.827.417	-107.972	78.024	2.955.411	-107.972	-213.986
Mar-99	2.773.701	2.752.804	2.833.297	-60.493	20.897	2.854.194	-80.493	-179.837
Apr-99	2.735.895	2.798.475	2.839.806	-41.331	-62.580	2.777.226	-41.331	-134.166
May-99	2.695.410	2.816.996	2.846.105	-29.110	-121.586	2.724.520	-29.110	-115.645
Jun-99	2.652.799	2.822.297	2.852.615	-30.317	-169.498	2.653.116	-30.317	-110.343
Jul-99	2.705.817	2.843.976	2.858.914	-14.938	-138.159	2.720.755	-14.938	-86.665
Aug-99	2.795.698	2.903.383	2.865.424	37.960	-107.685	2.757.738	37.960	-29.258
Sep-99	2.955.251	2.886.583	2.871.933	14.650	68.668	2.940.601	14.650	-46.057
Oct-99	3.007.658	2.902.859	2.878.232	24.627	104.799	2.903.031	24.627	-29.781
Nov-99	3.020.084	2.922.653	2.884.742	37.911	97.431	2.982.173	37.911	-9.988
Dec-99	3.032.833	2.928.423	2.891.041	38.382	103.410	2.994.451	38.382	-3.218
Jan-00	3.075.696	2.949.416	2.897.551	51.866	126.280	3.023.830	51.866	16.775
Feb-00	3.047.187	2.968.163	2.904.060	65.103	78.024	2.982.084	65.103	36.522
Mar-00	3.031.057	3.010.160	2.910.150	100.010	20.897	2.931.047	100.010	77.519
Apr-00	2.942.528	3.005.108	2.916.659	88.449	-62.580	2.854.079	88.449	72.467
May-00	2.873.213	2.994.799	2.922.958	71.840	-121.586	2.801.373	71.840	62.158
Jun-00	2.811.176	2.980.674	2.929.468	51.207	-169.498	2.759.969	51.207	48.034
Jul-00	2.844.805	2.982.964	2.935.767	47.197	-138.159	2.797.608	47.197	50.323
Aug-00	2.836.192	2.943.877	2.942.277	1.601	-107.685	2.834.591	1.601	11.236
Sep-00	3.047.197	2.978.529	2.948.786	29.743	68.668	3.017.454	29.743	45.889
Oct-00	3.080.972	2.976.173	2.955.085	21.088	104.799	3.059.884	21.088	43.533
Nov-00	3.066.260	2.968.829	2.961.595	7.234	97.431	3.059.026	7.234	36.188
Dec-00	3.078.322	2.974.912	2.967.894	7.018	103.410	3.071.304	7.018	42.271
Jan-01	3.095.541	2.989.261	2.974.404	-5.142	126.280	3.100.683	-5.142	36.620
Feb-01	3.042.742	2.964.718	2.980.913	-16.195	78.024	3.058.937	-16.195	32.077
Mar-01	2.956.884	2.935.987	2.965.763	-29.879	20.897	2.907.690	-29.879	3.349
Apr-01	2.886.788	2.945.368	2.993.302	-43.934	-62.580	2.930.722	-43.934	16.727
May-01	2.843.283	2.964.869	2.999.601	-34.733	-121.586	2.878.016	-34.733	32.228
Jun-01	2.813.409	2.982.907	3.006.111	-23.203	-169.498	2.836.612	-23.203	50.267
Jul-01	2.889.432	3.027.591	3.012.410	15.181	-138.159	2.871.261	15.181	94.950
Aug-01	2.976.979	3.084.664	3.018.920	65.745	-107.685	2.911.234	65.745	152.023
Sep-01	3.068.328	2.999.660	3.025.429	-25.769	68.668	3.094.097	-25.769	67.020
Oct-01	3.094.477	2.989.678	3.031.729	-42.050	104.799	3.136.527	-42.050	57.038
Nov-01	3.057.867	2.980.436	3.038.238	-77.802	97.431	3.135.669	-77.802	27.795
Dec-01	3.088.998	2.985.588	3.044.537	-58.949	103.410	3.147.947	-58.949	62.947
Jan-02		3.051.257			126.280	3.177.536		3.068.921
Feb-02		3.057.766			78.024	3.135.790		3.010.665
Mar-02		3.063.646			20.897	3.084.543		2.963.538
Apr-02		3.070.155			-62.580	3.007.575		2.910.061
May-02		3.076.454			-121.586	2.954.969		2.811.055
Jun-02		3.082.964			-169.498	2.913.466		2.763.142
Jul-02		3.089.263			-138.159	2.951.104		2.794.482
Aug-02		3.095.773			-107.685	2.988.087		2.824.956
Sep-02		3.102.282			68.668	3.170.950		3.001.308
Oct-02		3.108.582			104.799	3.213.380		3.037.439
Nov-02		3.115.091			97.431	3.212.522		3.030.072
Dec-02		3.121.390			103.410	3.224.800		3.036.051
Jan-03		3.127.900			126.280	3.254.180		3.041.665
Feb-03		3.134.409			78.024	3.212.433		3.010.665
Mar-03		3.140.289			20.897	3.161.186		2.953.538

DATE	VENTES	SÉRIE CVS	APPROX. LINEAIRE	ERREUR	COMPOSANTE SAISONNIERE	PREVISION (droite)	PREVISION (constante)	ERREUR
Jan-99	2.853.123	2.726.843	2.820.908	-94.064	126.280	2.947.187	-94.064	3.058.921
Feb-99	2.797.469	2.719.445	2.827.417	-107.972	78.024	2.955.411	-107.972	3.010.665
Mar-99	2.773.701	2.752.804	2.833.297	-60.493	20.897	2.854.194	-60.493	2.963.538
Apr-99	2.735.895	2.798.475	2.839.806	-41.331	-62.580	2.777.226	-41.331	2.870.061
May-99	2.695.410	2.816.996	2.846.105	-29.110	-121.586	2.724.520	-29.110	2.811.055
Jun-99	2.652.799	2.822.297	2.852.615	-30.317	-169.498	2.683.116	-30.317	2.763.142
Jul-99	2.705.817	2.843.976	2.858.914	-14.938	-138.159	2.720.755	-14.938	2.794.482
août-99	2.795.698	2.903.383	2.865.424	37.960	-107.685	2.757.738	37.960	2.824.956
sept-99	2.955.251	2.886.583	2.871.933	14.650	68.668	2.940.601	14.650	3.001.308
oct-99	3.007.658	2.902.859	2.878.232	24.627	104.799	2.903.031	24.627	3.037.439
nov-99	3.020.084	2.922.653	2.884.742	37.911	97.431	2.982.173	37.911	3.030.072
dec-99	3.032.833	2.928.423	2.891.041	38.382	103.410	2.994.451	38.382	3.036.051
Jan-00	3.075.696	2.949.416	2.897.551	51.866	126.280	3.023.830	51.866	3.058.921
Feb-00	3.047.187	2.968.163	2.904.060	65.103	78.024	2.982.084	65.103	3.010.665
Mar-00	3.031.057	3.010.160	2.910.150	100.010	20.897	2.931.047	100.010	2.963.538
Apr-00	2.942.528	3.005.108	2.916.659	88.449	-62.580	2.854.079	88.449	2.870.061
May-00	2.873.213	2.994.799	2.922.958	71.840	-121.586	2.801.373	71.840	2.811.055
Jun-00	2.811.176	2.980.674	2.929.468	51.207	-169.498	2.759.969	51.207	2.763.142
Jul-00	2.844.805	2.982.964	2.935.767	47.197	-138.159	2.797.608	47.197	2.794.482
août-00	2.836.192	2.943.877	2.942.277	1.601	-107.685	2.834.591	1.601	2.824.956
sept-00	3.047.197	2.978.529	2.948.786	29.743	68.668	3.017.454	29.743	3.001.308
oct-00	3.080.972	2.976.173	2.955.085	21.088	104.799	3.059.884	21.088	3.037.439
nov-00	3.066.260	2.968.829	2.961.595	7.234	97.431	3.059.026	7.234	3.030.072
dec-00	3.078.322	2.974.912	2.967.894	7.018	103.410	3.071.304	7.018	3.036.051
Jan-01	3.095.541	2.989.261	2.974.404	-5.142	126.280	3.100.683	-5.142	3.058.921
Feb-01	3.042.742	2.964.718	2.980.913	-16.195	78.024	3.058.937	-16.195	3.010.665
Mar-01	2.956.884	2.935.987	2.965.763	-29.879	20.897	2.907.690	-29.879	2.963.538
Apr-01	2.886.788	2.945.368	2.993.302	-43.934	-62.580	2.930.722	-43.934	2.920.061
May-01	2.843.283	2.964.869	2.999.601	-34.733	-121.586	2.878.016	-34.733	2.811.055
Jun-01	2.813.409	2.982.907	3.006.111	-23.203	-169.498	2.836.612	-23.203	2.763.142
Jul-01	2.889.432	3.027.591	3.012.410	15.181	-138.159	2.871.261	15.181	2.794.482
août-01	2.976.979	3.084.664	3.018.920	65.745	-107.685	2.911.234	65.745	2.824.956
sept-01	3.068.328	2.999.660	3.025.429	-25.769	68.668	3.094.097	-25.769	3.001.308
oct-01	3.094.477	2.989.678	3.031.729	-42.050	104.799	3.136.527	-42.050	3.037.439
nov-01	3.057.867	2.980.436	3.038.238	-77.802	97.431	3.135.669	-77.802	3.030.072
dec-01	3.088.998	2.985.588	3.044.537	-58.949	103.410	3.147.947	-58.949	3.036.051
Jan-02		3.051.257			126.280	3.177.536		3.068.921
Feb-02		3.057.766			78.024	3.135.790		3.010.665
Mar-02		3.063.646			20.897	3.084.543		2.963.538
Apr-02		3.070.155			-62.580	3.007.575		2.910.061
May-02		3.076.454			-121.586	2.954.969		2.811.055
Jun-02		3.082.964			-169.498	2.913.466		2.763.142
Jul-02		3.089.263			-138.159	2.951.104		2.794.482
Aug-02		3.095.773			-107.685	2.988.087		2.824.956
Sep-02		3.102.282			68.668	3.170.950		3.001.308
Oct-02		3.108.582			104.799	3.213.380		3.037.439
Nov-02		3.115.091			97.431	3.212.522		3.030.072
Dec-02		3.121.390			103.410	3.224.800		3.036.051
Jan-03		3.127.900			126.280	3.254.180		3.041.665
Feb-03		3.134.409			78.024	3.212.433		3.010.665
Mar-03		3.140.289			20.897	3.161.186		2.953.538

Pour cela, on part des séries de base X_t (*série brute*) et Z_t (*série corrigée des variations saisonnières*). Les séries sur la gauche donne l'approximation linéaire. Pour cela, on régresse la série Z_t sur les séries \mathbb{I} (*c'est à dire la constante*) et t (*la série "temps"*), et l'on obtient

$$\text{APPROX. LINEAIRE : } Z_t = A + Bt + \varepsilon_t \text{ avec } A = 209.98, B = -4772213$$

et où ε_t est la séries des erreurs (ERREUR)¹⁸. En particulier

$$\begin{cases} 2\ 858\ 914 = -4\ 772\ 213 + 209.98 \times 36\ 342 \quad (t = 36\ 342 \text{ correspond au } 1\text{er juillet } 1999) \\ 38\ 382 = 2\ 929\ 423 - 2\ 891\ 041 \end{cases}$$

La composante saisonnière (COMPOSANTE SAISONNIERE) est alors la série obtenue dans la partie précédente, et combinée à la série $\hat{Z}_t = A + Bt$: on obtient la première prévision de X_t : $\hat{X}_t = A + Bt + \Sigma_t$. (

5.5.3 Lissage exponentiel simple

Le lissage exponentiel simple est présenté ci-dessous, appliqué à la série corrigée des variations saisonnières. Pour cela, on se fixe au préalable une constante de lissage, par exemple $\lambda = 0.7$. La série lissée est alors définie par la façon suivante

$$\begin{cases} S_t = Z_{t-1} = 2\ 726\ 843 \text{ pour } t = Feb99 \\ S_t = \lambda Z_{t-1} + (1 - \lambda) S_{t-1} = 0.7 \times 3\ 005\ 108 + (1 - 0.7) \times 2\ 953\ 724 = 2\ 969\ 139 \text{ pour } t > Feb99 \\ S_t = S_T \text{ pour } t > T = Dec01 \end{cases}$$

L'erreur saisonnière est alors toujours la série (Σ_t) et la prévision est alors donnée par $\hat{X}_t = S_t + \Sigma_t$, soit dans l'exemple ci-dessous

$$3\ 051\ 423 = 2\ 925\ 143 + 126\ 280$$

DATE	VENTES	SERIE CVS	SERIE LISSEE	ERREUR SAISONNIERE	PREVISION (lissage)	ERREUR PREVISION
Jan-99	2,853,123	2,726,843				
Feb-99	2,797,469	2,719,445	2,726,843	78,024	2,804,867	-7,398
Mar-99	2,773,701	2,752,804	2,721,665	20,897	2,742,562	31,139
Apr-99	2,735,895	2,798,475	2,743,462	-62,580	2,680,882	55,013
May-99	2,695,410	2,816,996	2,781,971	-121,586	2,660,385	35,025
Jun-99	2,652,799	2,822,297	2,806,488	-169,498	2,636,990	15,809
Jul-99	2,705,817	2,843,976	2,817,555	-138,159	2,679,396	26,421
Aug-99	2,795,698	2,903,383	2,836,050	-107,685	2,728,364	67,334
Sep-99	2,955,251	2,886,583	2,883,183	68,668	2,951,851	3,400
Oct-99	3,007,658	2,902,859	2,885,563	104,799	2,990,362	17,296
Nov-99	3,020,084	2,922,653	2,897,671	97,431	2,995,102	24,982
Dec-99	3,032,833	2,929,423	2,915,158	103,410	3,018,568	14,265
Jan-00	3,075,696	2,949,416	2,925,143	126,280	3,051,423	24,273
Feb-00	3,047,187	2,969,163	2,942,134	78,024	3,020,158	27,029
Mar-00	3,031,057	3,010,160	2,961,055	20,897	2,981,952	49,105
Apr-00	2,942,528	3,005,108	2,995,428	-62,580	2,932,848	9,680
May-00	2,873,213	2,994,799	3,002,204	-121,586	2,880,618	-7,405
Jun-00	2,811,176	2,980,674	2,997,020	-169,498	2,827,522	-16,346
Jul-00	2,844,805	2,982,964	2,985,578	-138,159	2,847,419	-2,614
Aug-00	2,836,192	2,943,877	2,983,748	-107,685	2,876,063	-39,871
Sep-00	3,047,197	2,978,529	2,955,839	68,668	3,024,506	22,691
Oct-00	3,080,972	2,976,173	2,971,722	104,799	3,076,521	4,451
Nov-00	3,066,260	2,968,829	2,974,838	97,431	3,072,269	-6,009
Dec-00	3,078,322	2,974,912	2,970,632	103,410	3,074,042	4,280
Jan-01	3,095,541	2,989,261	2,973,628	126,280	3,099,908	-4,367
Feb-01	3,042,742	2,964,718	2,970,571	78,024	3,048,595	-5,853
Mar-01	2,956,884	2,935,987	2,966,474	20,897	2,987,371	-30,487
Apr-01	2,886,788	2,949,368	2,945,133	-62,580	2,882,553	4,235
May-01	2,843,283	2,964,869	2,981,971	-121,586	2,660,385	182,898
Jun-01	2,813,409	2,982,907	2,909,999	-169,498	2,740,501	72,908
Jul-01	2,889,432	3,027,591	2,961,035	-138,159	2,822,876	66,556
Aug-01	2,976,979	3,084,664	3,007,624	-107,685	2,899,939	77,040
Sep-01	3,068,328	2,999,660	3,061,552	68,668	3,130,220	-61,892
Oct-01	3,094,477	2,989,678	3,018,228	104,799	3,123,027	-28,550
Nov-01	3,057,867	2,960,436	2,998,243	97,431	3,095,675	-37,808
Dec-01	3,088,998	2,985,588	2,971,778	103,410	3,075,188	13,810
Jan-02		2,981,445	126,280	3,107,725		
Feb-02		2,981,445	78,024	3,059,469		
Mar-02		2,981,445	20,897	3,002,342		
Apr-02		2,981,445	-62,580	2,918,865		
May-02		2,981,445	-121,586	2,859,859		
Jun-02		2,981,445	-169,498	2,811,947		
Jul-02		2,981,445	-138,159	2,843,286		
Aug-02		2,981,445	-107,685	2,873,760		
Sep-02		2,981,445	68,668	3,050,112		
Oct-02		2,981,445	104,799	3,086,244		
Nov-02		2,981,445	97,431	3,078,876		
Dec-02		2,981,445	103,410	3,084,855		
Jan-03		2,981,445	126,280	3,107,725		
Feb-03		2,981,445	78,024	3,059,469		
Mar-03		2,981,445	20,897	3,002,342		

DATE	VENTES	SERIE CVS	SERIE LISSEE	ERREUR SAISONNIERE	PREVISION (lissage)	ERREUR PREVISION
Jan-99	2,853,123	2,726,843				
Feb-99	2,797,469	2,719,445	2,726,843	78,024	2,804,867	-7,398
Mar-99	2,773,701	2,752,804	2,721,665	20,897	2,742,562	31,139
Apr-99	2,735,895	2,798,475	2,743,462	-62,580	2,680,882	55,013
May-99	2,695,410	2,816,996	2,781,971	-121,586	2,660,385	35,025
Jun-99	2,652,799	2,822,297	2,806,488	-169,498	2,636,990	15,809
Jul-99	2,705,817	2,843,976	2,817,555	-138,159	2,679,396	26,421
Aug-99	2,795,698	2,903,383	2,836,050	-107,685	2,728,364	67,334
Sep-99	2,955,251	2,886,583	2,883,183	68,668	2,951,851	3,400
Oct-99	3,007,658	2,902,859	2,885,563	104,799	2,990,362	17,296
Nov-99	3,020,084	2,922,653	2,897,671	97,431	2,995,102	24,982
Dec-99	3,032,833	2,929,423	2,915,158	103,410	3,018,568	14,265
Jan-00	3,075,696	2,949,416	2,925,143	126,280	3,051,423	24,273
Feb-00	3,047,187	2,969,163	2,942,134	78,024	3,020,158	27,029
Mar-00	3,031,057	3,010,160	2,961,055	20,897	2,981,952	49,105
Apr-00	2,942,528	3,005,108	2,995,428	-62,580	2,932,848	9,680
May-00	2,873,213	2,994,799	3,002,204	-121,586	2,880,618	-7,405
Jun-00	2,811,176	2,980,674	2,997,020	-169,498	2,827,522	-16,346
Jul-00	2,844,805	2,982,964	2,985,578	-138,159	2,847,419	-2,614
Aug-00	2,836,192	2,943,877	2,983,748	-107,685	2,876,063	-39,871
Sep-00	3,047,197	2,978,529	2,955,839	68,668	3,024,506	22,691
Oct-00	3,080,972	2,976,173	2,971,722	104,799	3,076,521	4,451
Nov-00	3,066,260	2,968,829	2,974,838	97,431	3,072,269	-6,009
Dec-00	3,078,322	2,974,912	2,970,632	103,410	3,074,042	4,280
Jan-01	3,095,541	2,989,261	2,973,628	126,280	3,099,908	-4,367
Feb-01	3,042,742	2,964,718	2,970,571	78,024	3,048,595	-5,853
Mar-01	2,956,884	2,935,987	2,966,474	20,897	2,987,371	-30,487
Apr-01	2,886,788	2,949,368	2,945,133	-62,580	2,882,553	4,235
May-01	2,843,283	2,964,869	2,981,971	-121,586	2,660,385	182,898
Jun-01	2,813,409	2,982,907	2,909,999	-169,498	2,740,501	72,908
Jul-01	2,889,432	3,027,591	2,961,035	-138,159	2,822,876	66,556
Aug-01	2,976,979	3,084,664	3,007,624	-107,685	2,899,939	77,040
Sep-01	3,068,328	2,999,660	3,061,552	68,668	3,130,220	-61,892
Oct-01	3,094,477	2,989,678	3,018,228	104,799	3,123,027	-28,550
Nov-01	3,057,867	2,960,436	2,998,243	97,431	3,095,675	-37,808
Dec-01	3,088,998	2,985,588	2,971,778	103,410	3,075,188	13,810
Jan-02		2,981,445	126,280	3,107,725		
Feb-02		2,981,445	78,024	3,059,469		
Mar-02		2,981,445	20,897	3,002,342		
Apr-02		2,981,445	-62,580	2,918,865		
May-02		2,981,445	-121,586	2,859,859		
Jun-02		2,981,445	-169,498	2,811,947		
Jul-02		2,981,445	-138,159	2,843,286		
Aug-02		2,981,445	-107,685	2,873,760		
Sep-02		2,981,445	68,668	3,050,112		
Oct-02		2,981,445	104,799	3,086,244		
Nov-02		2,981,445	97,431	3,078,876		
Dec-02		2,981,445	103,410	3,084,855		
Jan-03		2,981,445	126,280	3,107,725		
Feb-03		2,981,445	78,024	3,059,469		
Mar-03		2,981,445	20,897	3,002,342		

5.5.4 Lissage exponentiel double

Les séries (S) et (S') sont définies récursivement de la façon suivante

$$\begin{cases} S_t = \lambda X_{t-1} + (1 - \lambda) S_{t-1} \text{ pour } t = 2, \dots, n + 1 \\ S'_t = \lambda S_{t-1} + (1 - \lambda) S'_{t-1} \text{ pour } t = 2, \dots, n + 1 \end{cases}$$

Les coefficients A et B sont définie par

$$A_t = 2S_t - S'_t \text{ et } B_t = \frac{\lambda}{1 - \lambda} [S_t - S'_t]$$

soit sur l'exemple ci-dessous

$$A_t = 2 \times 2\ 771\ 987 - 2\ 746\ 504 = 2\ 797\ 469 \text{ et } B_t = \frac{0.7}{1 - 0.7} (2\ 882\ 758 - 2\ 842\ 769)$$

La série "lissage brut" est donnée par

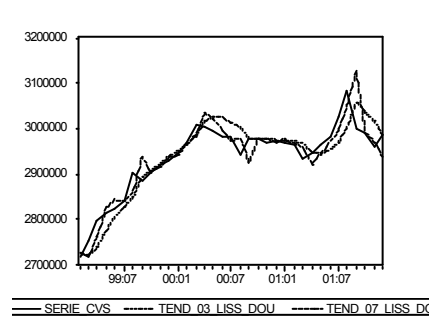
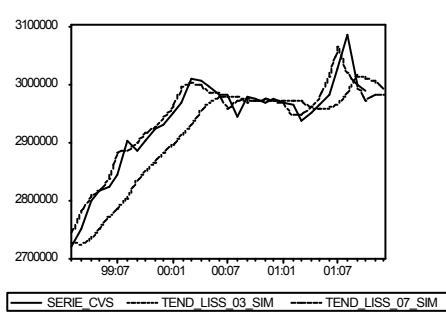
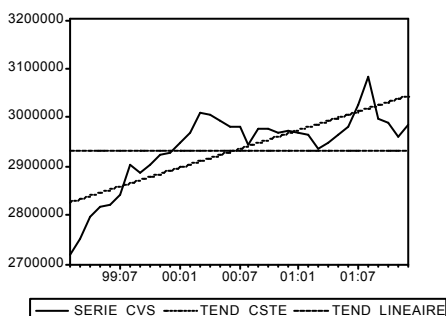
$$A_t + B_t = 2\ 877\ 040 + 18\ 480 = 2\ 895\ 520$$

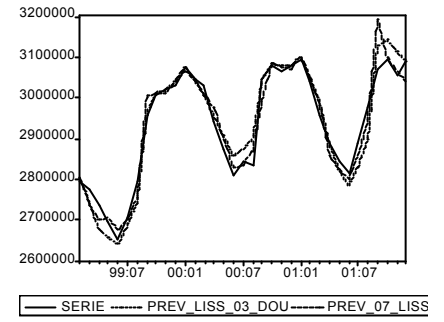
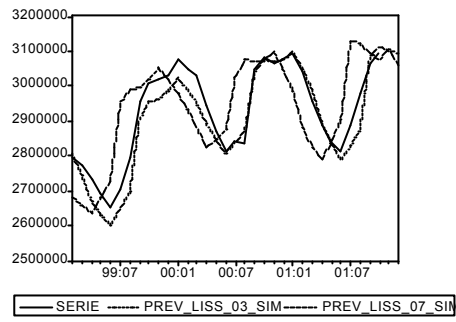
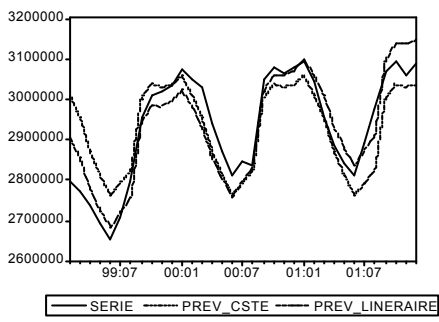
La série de la prévision est alors donnée, lorsqu'il y a n observations, par

$$\begin{cases} X_t^P = A_t + B_t \text{ pour } t = 2, \dots, n + 1 \\ X_{n+h}^P = A_{n+1} + h.B_{n+1} \text{ pour } h \geq 1 \end{cases}$$

DATE	VENTES	SERIE CVS	SERIE LISSEE S	SERIE LISSEE SS	COEFF. A	COEFF. B	LISSAGE BRUT	ERREUR SAISONNIERE	PREVISION (lissage)	ERREUR PREVISION
Jan-99	2,853,123	2,726,843								
Feb-99	2,797,469	2,719,445	2,726,843	2,726,843	2,726,843	0	2,726,843	78,024	2,804,867	-7,398
Mar-99	2,773,701	2,752,804	2,721,665	2,723,218	2,720,111	-3,625	2,716,486	20,897	2,737,383	36,318
Apr-99	2,735,895	2,798,475	2,743,462	2,737,389	2,749,535	14,171	2,763,706	-62,580	2,701,126	34,769
May-99	2,695,410	2,816,996	2,781,971	2,768,596	2,795,346	31,207	2,826,553	-121,586	2,704,967	-9,557
Jun-99	2,652,799	2,822,297	2,806,488	2,795,121	2,817,856	26,524	2,844,380	-169,498	2,674,882	-22,083
Jul-99	2,705,817	2,843,976	2,817,555	2,810,824	2,824,285	15,704	2,839,989	-138,159	2,701,830	3,987
Aug-99	2,795,698	2,903,383	2,836,050	2,828,482	2,843,617	17,658	2,861,275	-107,685	2,753,590	42,108
Sep-99	2,955,251	2,886,583	2,883,183	2,866,773	2,899,594	38,291	2,937,884	68,668	3,006,552	-51,301
Oct-99	3,007,658	2,902,859	2,885,563	2,879,926	2,891,201	13,153	2,904,354	104,799	3,009,153	-1,495
Nov-99	3,020,084	2,922,653	2,897,671	2,892,347	2,902,994	12,421	2,915,415	97,431	3,012,846	7,238
Dec-99	3,032,833	2,929,423	2,915,158	2,908,315	2,922,001	15,968	2,937,969	103,410	3,041,379	-8,546
Jan-00	3,075,696	2,949,416	2,925,143	2,920,095	2,930,192	11,780	2,941,972	126,280	3,068,252	7,444
Feb-00	3,047,187	2,989,163	2,942,134	2,935,523	2,948,746	15,428	2,964,174	78,024	3,042,198	4,989
Mar-00	3,031,057	3,010,160	2,961,055	2,953,395	2,968,714	17,872	2,986,587	20,897	3,007,484	23,573
Apr-00	2,942,528	3,005,108	2,995,428	2,982,818	3,008,038	29,423	3,037,461	-62,580	2,974,882	-32,354
May-00	2,873,213	2,994,799	3,002,204	2,996,388	3,008,020	13,570	3,021,590	-121,586	2,900,004	-82,791
Jun-00	2,811,176	2,980,674	2,997,020	2,996,831	2,997,210	442	2,997,652	-169,498	2,828,154	-16,978
Jul-00	2,844,805	2,982,964	2,985,578	2,988,954	2,982,202	-7,877	2,974,326	-138,159	2,836,167	8,638
Aug-00	2,836,192	2,943,877	2,983,748	2,985,310	2,982,187	-3,644	2,978,543	-107,685	2,870,857	-34,665
Sep-00	3,047,197	2,978,529	2,955,839	2,964,680	2,946,997	-20,830	2,926,367	68,668	2,995,035	52,162
Oct-00	3,080,972	2,976,173	2,971,722	2,969,610	2,973,835	4,930	2,978,764	104,799	3,083,563	-2,591
Nov-00	3,066,260	2,968,829	2,974,838	2,973,270	2,976,407	3,660	2,980,067	97,431	3,077,498	-11,238
Dec-00	3,078,322	2,974,912	2,970,632	2,971,423	2,969,840	-1,847	2,967,994	103,410	3,071,404	6,918
Jan-01	3,095,541	2,969,261	2,973,628	2,972,966	2,974,289	1,543	2,975,833	126,280	3,102,112	-6,571
Feb-01	3,042,742	2,964,718	2,970,571	2,971,290	2,969,853	-1,677	2,968,176	78,024	3,046,200	-3,458
Mar-01	2,956,884	2,935,987	2,966,474	2,967,919	2,965,029	-3,371	2,961,659	20,897	2,982,556	-25,672
Apr-01	2,886,788	2,949,368	2,945,133	2,951,969	2,938,297	-15,950	2,922,347	-62,580	2,859,767	27,021
May-01	2,843,283	2,964,869	2,948,097	2,949,259	2,946,936	-2,710	2,944,226	-121,586	2,822,640	20,643
Jun-01	2,813,409	2,982,907	2,959,837	2,956,664	2,963,011	7,405	2,970,416	-169,498	2,800,917	12,492
Jul-01	2,889,432	3,027,591	2,975,986	2,970,190	2,981,783	13,526	2,995,309	-138,159	2,857,150	32,282
Aug-01	2,976,979	3,084,664	3,012,110	2,999,534	3,024,686	29,344	3,054,030	-107,685	2,946,344	30,635
Sep-01	3,068,328	2,999,660	3,062,898	3,043,889	3,081,907	44,355	3,126,262	68,668	3,194,930	-126,602
Oct-01	3,094,477	2,989,678	3,018,632	3,026,209	3,011,055	-17,680	2,993,375	104,799	3,098,173	-3,696
Nov-01	3,057,867	2,960,436	2,998,364	3,006,718	2,990,011	-19,491	2,970,520	97,431	3,067,951	-10,084
Dec-01	3,088,998	2,985,588	2,971,814	2,982,285	2,961,343	-24,432	2,936,911	103,410	3,040,321	48,677
Jan-02			2,981,456	2,981,705	2,981,207	-581	2,980,626	126,280	3,106,906	-3,106,906
Feb-02					2,981,207	-1,161	2,980,046	78,024	3,058,069	-3,058,069
Mar-02					2,981,207	-1,742	2,979,465	20,897	3,000,362	-3,000,362
Apr-02					2,981,207	-2,323	2,978,884	-62,580	2,916,305	-2,916,305
May-02					2,981,207	-2,903	2,978,304	-121,586	2,856,718	-2,856,718
Jun-02					2,981,207	-3,484	2,977,723	-169,498	2,808,225	-2,808,225
Jul-02					2,981,207	-4,065	2,977,142	-138,159	2,838,983	-2,838,983
Aug-02					2,981,207	-4,645	2,976,562	-107,685	2,868,877	-2,868,877
Sep-02					2,981,207	-5,226	2,975,981	68,668	3,044,649	-3,044,649
Oct-02					2,981,207	-5,806	2,975,401	104,799	3,080,199	-3,080,199
Nov-02					2,981,207	-6,387	2,974,820	97,431	3,072,251	-3,072,251
Dec-02					2,981,207	-6,968	2,974,239	103,410	3,077,649	-3,077,649
Jan-03					2,981,207	-7,548	2,973,659	126,280	3,099,938	-3,099,938
Feb-03					2,981,207	-8,129	2,973,078	78,024	3,051,102	-3,051,102
Mar-03					2,981,207	-8,710	2,972,497	20,897	2,993,395	-2,993,395

Les graphiques ci-dessous correspondent, respectivement à la comparaison entre les lissages corrigés des variations saisonnières et les lissages de la série brute (*en haut et en bas respectivement*), avec à gauche le lissage par une constante et par une droite, au centre un lissage exponentiel simple et à droite un lissage exponentiel double,





”Prévoir consiste à projeter dans l’avenir ce qu’on a perçu dans le passé.”
Henri Bergson, Le possible et le réel (1930) .

6 Introduction aux modèles linéaires *ARIMA*

6.1 Rappels sur les espaces \mathcal{L}^2

6.1.1 Propriétés topologiques

On considère le processus (X_t) défini sur l’espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs dans \mathbb{R} .

Définition 28 L’espace $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ est l’espace des variables de carré intégrable (variances-covariances finies).

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

Propriété 21 L^2 est un espace de Hilbert, muni du **produit scalaire** $\langle \cdot, \cdot \rangle$ et la **norme associée** $\|\cdot\|$

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

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

Théorème 5 Théorème de projection Si \mathcal{H} est un sous espace fermé de L^2 , pour toute variable $Y \in L^2$, il existe une unique variable aléatoire $\hat{Y} \in \mathcal{H}$ tel que

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

caractérisé par $\hat{Y} \in \mathcal{H}$ et $Y - \hat{Y} \in \mathcal{H}^\perp$. On notera aussi $\Pi_{\mathcal{H}}(Y)$.

Remarque 27 X_n converge vers X au sens de \mathcal{L}^2 si

$$\lim_{n \rightarrow \infty} \|X_n - X\| = 0, \text{ c'est à dire } \begin{cases} \lim_{n \rightarrow \infty} \mathbb{E}X_n = \mathbb{E}X \\ \lim_{n \rightarrow \infty} V(X_n - X) = 0 \end{cases}$$

On peut alors définir la variable aléatoire $Y = \sum_{n \in \mathbb{Z}} a_n X_n$ comme limite, dans L^2 de $Y_{p,q}$

$$Y_{p,q} = \sum_{n=-p}^q a_n X_n \text{ et } Y = \lim_{p,q \rightarrow +\infty} Y_{p,q} = \sum_{n \in \mathbb{Z}} a_n X_n.$$

6.1.2 Rappel sur les vecteurs et processus gaussiens

Pour un vecteur aléatoire $Z = (Z_1, \dots, Z_d)'$, on définit son espérance par $\mathbb{E}(Z) = (\mathbb{E}(Z_1), \dots, \mathbb{E}(Z_d))'$ et sa matrice de variance-covariance (si elles existent) par $V(Z) = \mathbb{E}((Z - \mathbb{E}(Z))(Z - \mathbb{E}(Z))')$. Cette matrice est hermitienne positive. De plus, si A est la matrice d’une application linéaire de \mathbb{R}^d dans \mathbb{R}^p , le vecteur AZ admet pour espérance $A\mathbb{E}(Z)$ et pour matrice de variance-covariance $AV(Z)A'$.

Définition 29 Le vecteur $X = (X_1, \dots, X_d)$ est un vecteur gaussien si toute combinaison des X_i est une variable gaussienne, i.e. pour tout $a \in \mathbb{R}^d$, aX est une variable gaussienne. Sa densité s’écrit alors

$$f(x) = \frac{1}{(2\pi)^{d/2} \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(x - \mu)' \Sigma^{-1}(x - \mu)\right),$$

où $\mu \in \mathbb{R}^d$ et Σ est une matrice hermitienne positive $d \times d$.

Si X est un vecteur gaussien, son moment à l’ordre p existe et de plus, $\mathbb{E}(X) = \mu$ et $V(X) = \Sigma$.

Définition 30 Le processus (X_t) est un processus gaussien si tout système fini extrait est un vecteur aléatoire gaussien, i.e. pour tout n , pour tout t_1, \dots, t_n , $(X_{t_1}, \dots, X_{t_n})$ est un vecteur gaussien.

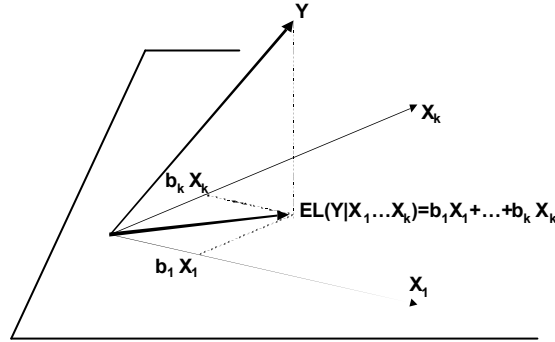
6.1.3 Régression affine dans \mathcal{L}^2

(i) régression sur un nombre fini de variables

La **régression linéaire théorique** de Y sur X_1, \dots, X_n est la projection orthogonale dans $\mathcal{L}^2(\cdot, \mathcal{A}, \mathbb{P})$ de X sur $\mathcal{H}^* = \text{Vect}(X_1, \dots, X_n)$, et la **régression affine théorique** de Y sur X_1, \dots, X_n est la projection orthogonale dans $\mathcal{L}^2(\cdot, \mathcal{A}, \mathbb{P})$ de Y sur $\mathcal{H} = \text{Vect}(\mathbb{I}, X_1, \dots, X_n)$. On note alors

$$\widehat{Y} = EL(Y|\mathbb{I}, X_1, \dots, X_n) = \Pi_{\mathcal{H}}(Y)$$

où $EL(\cdot|\cdot)$ désigne l'**espérance linéaire**. Cette variable est la meilleure approximation (au sens de L^2) de Y par une combinaison linéaire de $\mathbb{I}, X_1, \dots, X_n$,



Propriété 22 Soit γ le vecteur $[\text{cov}(X, X_i)]_{i=0,1,\dots,n}$ et Σ la matrice $[\text{cov}(X_i, X_j)]_{i,j=0,1,\dots,n}$. Alors

$$\widehat{X} = EL(X|\mathbb{I}, X_1, \dots, X_n) = a_0 + a_1X_1 + \dots + a_nX_n,$$

où $a = (a_0, a_1, \dots, a_n)$ vérifie $a = \Sigma^{-1}\gamma$.

Preuve. \widehat{X} peut s'écrire $\alpha_0 + \alpha_1X_1 + \dots + \alpha_nX_n$ car $\widehat{X} \in \mathcal{H}$, et vérifie

$$\begin{cases} \langle X - \widehat{X}, \mathbb{I} \rangle = 0 \\ \langle X - \widehat{X}, X_i \rangle = 0 \text{ pour } i = 1, \dots, n, \end{cases}$$

car $X - \widehat{X} \in \mathcal{H}^\perp$, c'est à dire

$$\begin{cases} \mathbb{E}(X - \widehat{X}) = 0 \text{ soit } \mathbb{E}(X) = \mathbb{E}(\widehat{X}) \\ \mathbb{E}([X - \widehat{X}] X_i) = 0 \text{ pour } i = 1, \dots, n. \end{cases}$$

D'où

$$\begin{cases} \mathbb{E}(X) = \mathbb{E}(\widehat{X}) = \alpha_0 + \alpha_1\mathbb{E}(X_1) + \dots + \alpha_n\mathbb{E}(X_n) & (1) \\ \mathbb{E}(XX_i) = \mathbb{E}(\widehat{X}X_i) = \alpha_0\mathbb{E}(X_i) + \alpha_1\mathbb{E}(X_1X_i) + \dots + \alpha_n\mathbb{E}(X_nX_i) & (2) \end{cases}$$

(1) donne $\alpha_0 = \mathbb{E}(X) - \alpha_1\mathbb{E}(X_1) - \dots - \alpha_n\mathbb{E}(X_n)$ et par substitution dans (2),

$$\begin{aligned} \mathbb{E}(XX_i) &= \mathbb{E}(X)\mathbb{E}(X_i) - (\alpha_1\mathbb{E}(X_1) + \dots + \alpha_n\mathbb{E}(X_n))\mathbb{E}(X_i) \\ &\quad + \alpha_1\mathbb{E}(X_1X_i) + \dots + \alpha_n\mathbb{E}(X_nX_i) \end{aligned}$$

donc, pour $i = 1, \dots, n$

$$\text{cov}(XX_i) = \alpha_1\text{cov}(X_1X_i) + \dots + \alpha_n\text{cov}(X_nX_i),$$

ce qui donne le système

$$\begin{cases} \text{cov}(XX_1) = \alpha_1\text{cov}(X_1X_1) + \dots + \alpha_n\text{cov}(X_nX_1) \\ \text{cov}(XX_2) = \alpha_1\text{cov}(X_1X_2) + \dots + \alpha_n\text{cov}(X_nX_2) \\ \dots \\ \text{cov}(XX_n) = \alpha_1\text{cov}(X_1X_n) + \dots + \alpha_n\text{cov}(X_nX_n), \end{cases}$$

qui s'écrit sous forme matricielle

$$\gamma = \Sigma\alpha.$$

■

Remarque 28 On peut noter que $EL(X|\mathbb{I}, X_1, \dots, X_n) = EL(X|X_1, \dots, X_n)$ si et seulement si $\mathbb{E}(X) = 0$ et $\mathbb{E}(X_j) = 0$ pour $j = 1, 2, \dots, n$.

(ii) **régression sur un nombre infini de variables**

On considère cette fois ci X_1, \dots, X_n, \dots des variables de L^2 , et $X_0 = \mathbb{I} \in L^2$. Soit \mathcal{H} l'adhérence de l'espace engendré par les combinaisons linéaires des X_i : $\mathcal{H} = \overline{\mathcal{L}}(\mathbb{I}, X_1, \dots, X_n, \dots)$.

On considère alors $\widehat{X}_n = EL(X|\mathbb{I}, X_1, \dots, X_n)$. La projection sur \mathcal{H} est alors la limite (dans L^2) des variables \widehat{X}_n

$$\widehat{X} = \lim_{n \rightarrow \infty} EL(X|\mathbb{I}, X_1, \dots, X_n).$$

6.1.4 La notion d'innovation

Opérateur de projection linéaire Etant données une variable aléatoire Y et une famille de variables aléatoires $\{X_1, \dots, X_n\}$, on définit l'opérateur de projection linéaire Π de la façon suivante :

$$\Pi(Y|\{X_1, \dots, X_n\}) = \alpha_1 X_1 + \dots + \alpha_n X_n,$$

où les α_i sont les solutions du programme d'optimisation

$$(\alpha_1, \dots, \alpha_n) = \arg \min_{a_1, \dots, a_n} \{Var[Y - (a_1 X_1 + \dots + a_n X_n)]\}.$$

En d'autres termes, $\Pi(Y|\{X_1, \dots, X_n\})$ est la meilleure prévision linéaire de Y basée sur $\{X_1, \dots, X_n\}$, au sens où la variance de l'erreur sera minimale. On peut noter que cet opérateur est linéaire, au sens où

$$\Pi(\lambda Y + \mu Z|\{X_1, \dots, X_n\}) = \lambda \Pi(Y|\{X_1, \dots, X_n\}) + \mu \Pi(Z|\{X_1, \dots, X_n\}).$$

De plus, l'erreur de prévision $Y - \Pi(Y|\{X_1, \dots, X_n\})$ est non corrélée avec toute fonction linéaire des X_1, \dots, X_n . Enfin, si $cov(X_1, X_2) = 0$, alors $\Pi(Y|\{X_1, X_2\}) = \Pi(Y|\{X_1\}) + \Pi(Y|\{X_2\})$.

Il est possible de projeter sur une suite infinie de variables aléatoires $\{X_1, \dots, X_n, \dots\}$, en notant

$$\Pi(Y|\{X_1, \dots, X_n, \dots\}) = \lim_{k \rightarrow \infty} \Pi(Y|\{X_1, \dots, X_k\}).$$

Cette limite existant pour toute suite de variables aléatoires $\{X_1, \dots, X_n, \dots\}$.

Prévision linéaire

Définition 31 Soit $(X_t)_{t \in \mathbb{Z}}$ un processus de L^2 . On appelle **meilleur prévision linéaire** de X_t sachant son passé la **régression linéaire (théorique)** de X_t sur son passé $\mathcal{H} = Vect(\mathbb{I}, X_{t-1}, X_{t-2}, \dots)$, et sera notée $EL(X_t|\mathbb{I}, X_{t-1}, X_{t-2}, \dots)$.

Définition 32 Le **processus d'innovation** du processus (X_t) est le processus (ε_t) défini par

$$\varepsilon_t = X_t - EL(X_t|\mathbb{I}, X_{t-1}, X_{t-2}, \dots).$$

Propriété 23 Soit (Y_t) un bruit blanc $BB(0, \sigma^2)$, le processus stationnaire (X_t) définit par¹⁹

$$X_t = Y_t - \rho Y_{t-1} \text{ pour } |\rho| < 1,$$

alors

$$EL(X_t|\mathbb{I}, X_{t-1}, X_{t-2}, \dots) = - \sum_{i=1}^{\infty} \rho^i X_{t-i}$$

et le processus d'innovation est (Y_t) .

Preuve. Le processus (X_t) est stationnaire en tant que moyenne mobile de bruit blanc. Définissons alors

$$S_{t,n} = - \sum_{i=1}^n \rho^i X_{t-i}.$$

¹⁹Ceci correspond à un processus $MA(1)$, dont la racine est à l'intérieur du cercle unité.

A t fixé, la suite $(S_{t,n})$ est une suite de Cauchy dans L^2 puisque

$$\|S_{t,n} - S_{t,m}\| = \left\| \sum_{i=m+1}^n \rho^i X_{t-i} \right\| \leq \|X_t\| \sum_{i=m+1}^n |\rho^i|,$$

qui tend vers 0 quand m et n tendent vers l'infini. $(S_{t,n})$ converge donc dans L^2 vers $S_t = \sum_{i=1}^{\infty} \rho^i X_{t-i}$, élément de $\text{Vect}(\mathbb{I}, X_{t-1}, X_{t-2}, \dots)$.

Or $X_t = Y_t - \rho Y_{t-1}$, donc $X_t + S_{n,t} = Y_t + \rho^{n+1} Y_{t-n-1}$ et donc

$$X_t + \sum_{i=1}^{\infty} \rho^i X_{t-i} = Y_t,$$

puisque $\|\rho^{n+1} Y_{t-n-1}\| \leq \sigma^2 |\rho|^{n+1} \rightarrow 0$ quand $n \rightarrow \infty$, d'où $X_t = S_t + Y_t$.

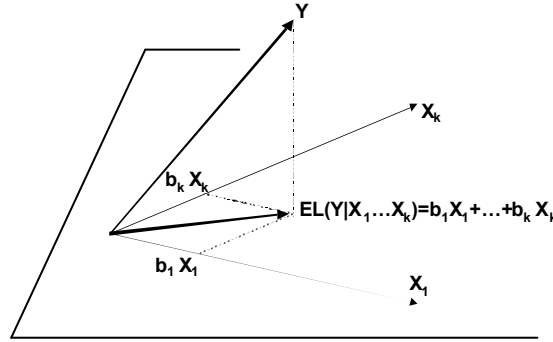
Or $\langle X_s, Y_t \rangle = 0$ pour tout $s < t$, $\langle \mathbb{I}, Y_t \rangle = 0$ et $S_t \in \text{Vect}(\mathbb{I}, X_{t-1}, X_{t-2}, \dots)$, donc $S_t = EL(X_t | \mathbb{I}, X_{t-1}, X_{t-2}, \dots)$ et (ε_t) est le processus d'innovation. ■

Remarque 29 Soit (Y_t) un bruit blanc BB $(0, \sigma^2)$, le processus stationnaire (X_t) satisfaisant

$$X_t - \rho X_{t-1} = Y_t, \text{ avec } |\rho| > 1,$$

Comme nous le verrons par la suite (propriété (??)) le processus Y_t ainsi défini ne correspond pas à l'innovation du processus X_t . Il est possible de montrer (en utilisant la densité spectrale) que le processus ε_t défini par $\varepsilon_t = X_t - \rho^{-1} X_{t-1}$ est effectivement un bruit blanc. En fait, (ε_t) correspond au processus d'innovation associé au processus (X_t) .

Du fait de cette définition, l'innovation possède un certain nombre de propriétés



Comme on peut le voir sur le schéma ci-dessus, si ε_t est l'innovation, alors elle est orthogonale au passé de X_t c'est à dire que

$$\begin{cases} \mathbb{E}(\varepsilon_t X_{t-1}) = \mathbb{E}(\varepsilon_t X_{t-2}) = \dots = \mathbb{E}(\varepsilon_t X_{t-h}) = \dots = 0 \\ \text{mais } \mathbb{E}(\varepsilon_t X_t) \neq 0. \end{cases}$$

De plus, on aura également que ε_{t+k} sera également orthonogonale au passé de X_t , pour $k \geq 0$,

$$\mathbb{E}(\varepsilon_{t+k} X_{t-1}) = \mathbb{E}(\varepsilon_{t+k} X_{t-2}) = \dots = \mathbb{E}(\varepsilon_{t+k} X_{t-h}) = \dots = 0.$$

Remarque 30 De façon rigoureuse, il conviendrait d'introduire la notion de processus régulier : on dira que le processus stationnaire (X_t) , centré, est régulier s'il existe un bruit blanc (ε_t) tel que, pour tout $t \in \mathbb{Z}$, l'adhérence des passés (linéaires) $\mathcal{H}_X^t = \text{Vect}(\mathbb{I}, X_{t-1}, X_{t-2}, \dots)$ et $\mathcal{H}_\varepsilon^t = \text{Vect}(\mathbb{I}, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots)$ coïncident : $\mathcal{H}_X^t = \mathcal{H}_\varepsilon^t$. On peut alors montrer si (X_t) est un processus stationnaire régulier, et si (ε_t) est un bruit blanc tel que, à chaque date t les passés coïncident, alors on a la décomposition

$$\mathcal{H}_X^t = \mathcal{H}_X^{t-1} \oplus \mathbb{R}\varepsilon_t, \text{ pour tout } t,$$

où \oplus désigne une somme directe orthogonale, et le processus bruit blanc est alors unique : il est appelé innovation du processus (X_t) . Le fait que les deux espaces coïncident implique, en particulier, que si (ε_t) est l'innovation du processus (X_t) alors

$$EL(X_{T+k} | X_T, X_{T-1}, \dots) = EL(X_{T+h} | \varepsilon_T, \varepsilon_{T-1}, \dots).$$

Compléments à l'aide des espaces \mathcal{H} Etant donné un processus (X_t) , on notera $\mathcal{H}(X)$ le sous-espace de Hilbert de L^2 correspondant à l'adhérence, dans L^2 , de l'espace des combinaisons linéaires finies d'éléments de (X_t) . On notera $\mathcal{H}_T(X)$ le sous-espace de Hilbert de L^2 correspondant à l'adhérence, dans L^2 , de l'espace des combinaisons linéaires finies d'éléments de (X_t) avec $t \leq T$.

Définition 33 On appelle processus d'innovation la suite $\varepsilon_t = X_t - \Pi_{\mathcal{H}_{t-1}(X)}(X_t)$.

Ce processus est alors une suite orthogonale (pour le produit scalaire $\langle \cdot, \cdot \rangle$), et on a l'inclusion $\mathcal{H}_{t-1}(\varepsilon) \subset \mathcal{H}_{t-1}(X)$.

6.2 Polynômes d'opérateurs retard L et avance F

6.2.1 Rappels sur les opérateurs retards

Nous avons défini précédemment l'opérateur retard L par $L : X_t \mapsto L(X_t) = LX_t = X_{t-1}$ et l'opérateur avance F par $F : X_t \mapsto F(X_t) = FX_t = X_{t+1}$. On notera alors

$$L^p = \underbrace{L \circ L \circ \dots \circ L}_{p \text{ fois}} \text{ où } p \in \mathbb{N},$$

avec la convention $L^0 = \mathbb{I}$ et $L^{-1} = F$. Et de façon analogue, $L^{-p} = F^p$ pour $p \in \mathbb{N}$.

6.2.2 Inversibilité des polynômes $P(L)$

Soit $A(\cdot)$ un polynôme, on cherche $B(\cdot)$ tel que $A(\cdot) \circ B(\cdot) = B(\cdot) \circ A(\cdot) = 1$.

(i) **inversibilité de $P(L) = 1 - \lambda L$**

Propriété 24 (i) Si $|\lambda| < 1$ alors $1 - \lambda L$ est inversible, et de plus,

$$(1 - \lambda L)^{-1} = \sum_{k=0}^{\infty} \lambda^k L^k.$$

(ii) Si $|\lambda| > 1$ alors $1 - \lambda L$ est inversible, et de plus,

$$(1 - \lambda L)^{-1} = \sum_{k=1}^{\infty} \frac{1}{\lambda^k} F^k.$$

(iii) Si $|\lambda| = 1$, alors $1 - \lambda L$ n'est pas inversible.

Preuve. (i) Si $|\lambda| < 1$ alors

$$(1 - \lambda)^{-1} = \sum_{k=0}^{+\infty} \lambda^k = \frac{1}{1 - \lambda} < +\infty,$$

donc $A(L) = \sum_{k=0}^{+\infty} \lambda^k L^k$ est bien défini. De plus:

$$(1 - \lambda L)A(L) = \lim_{k \rightarrow +\infty} (1 - \lambda L) \left(\sum_{j=0}^k \lambda^j L^j \right) = \lim_{k \rightarrow +\infty} 1 - \lambda^{k+1} L^{k+1} = 1,$$

ce qui signifie que A est le polynôme inverse associé à $(1 - \lambda L)$.

(ii) De façon analogue, si $|\lambda| > 1$ alors $1 - \lambda L = -\lambda \left(L - \frac{1}{\lambda} \right) = -\lambda L \left(1 - \frac{F}{\lambda} \right)$. On a alors :

$$(\lambda L)^{-1} = \frac{1}{\lambda} F \text{ et } \left(1 - \frac{F}{\lambda} \right)^{-1} = \sum_{k=0}^{+\infty} \frac{1}{\lambda^k} F^k \text{ car } \left| \frac{1}{\lambda} \right| < 1$$

En combinant ces deux résultats :

$$\begin{aligned} (1 - \lambda L)^{-1} &= (-\lambda L)^{-1} \left(1 - \frac{F}{\lambda} \right)^{-1} = -\frac{1}{\lambda} F \left(\sum_{k=0}^{+\infty} \frac{1}{\lambda^k} F^k \right) \\ &= -\sum_{k=1}^{+\infty} \frac{1}{\lambda^k} F^k = -\sum_{k=-\infty}^{-1} \lambda^k L^k, \end{aligned}$$

ce qui correspond au résultat souhaité.

(iii) En effet, il n'existe pas de polynôme $A(L) = \sum_{k \in \mathbb{Z}} a_k L^k$, $\sum_{k \in \mathbb{Z}} |a_k| < +\infty$ tel que $(1-L)A(L) = 1$. En effet, si un tel polynôme existait,

$$(1-L)A(L) = 1 \Rightarrow |a_k| = |a_{k-1}| \not\rightarrow 0 \text{ quand } k \rightarrow \infty,$$

et donc $\sum_{k \in \mathbb{Z}} |a_k| = +\infty$. ■

Exemple 37 Soit (X_t) et (Y_t) deux processus stationnaires tels que $Y_t = X_t - \lambda X_{t-1} = (1 - \lambda L)X_t$, où $\lambda < 1$. Cette relation s'inverse en

$$X_t = (1 - \lambda L)^{-1} Y_t = Y_t + \lambda Y_{t-1} + \dots + \lambda^k Y_{t-k} + \dots$$

Exemple 38 Dans le cas où $\lambda = 1$ (**racine unité**) on se retrouve en présence d'une marche aléatoire $Y_t = X_t - X_{t-1}$ (non stationnaire).

(ii) **inversibilité des polynômes en L**

Tout polynôme $A(L) = 1 + a_1 L + \dots + a_n L^n$ (normalisé tel que $A(0) = 1$), peut s'écrire

$$A(z) = a_n (z - z_1)(z - z_2) \dots (z - z_n),$$

correspondant à la décomposition en éléments simples ($z_i = \text{racines du polynôme}$). On peut écrire

$$A(L) = \prod_{i=1}^n (1 - \lambda_i L) \text{ où } \lambda_i = \frac{1}{z_i}$$

Propriété 25 Si pour tout i , $|\lambda_i| \neq 1$, alors $A(L)$ est inversible. Pour cela, notons

$$A(L) = \prod (1 - \lambda_i L) = \underbrace{\prod_{|\lambda_i| < 1} (1 - \lambda_i L)}_{A_1(L)} \underbrace{\prod_{|\lambda_i| > 1} \left(1 - \frac{1}{\lambda_i} F\right)}_{A_2(L)} \underbrace{\prod_{|\lambda_i| > 1} (-\lambda_i L)}_{A_3(L)},$$

puisque $(1 - \lambda_i L) = -\lambda_i L (1 - F/\lambda_i)$, et alors, l'inverse de $A(L)$ est donné par $A(L)^{-1}$

$$A(L)^{-1} = \prod (1 - \lambda_i L)^{-1} = A_1(L)^{-1} A_2(L)^{-1} A_3(L)^{-1} = \underbrace{\prod_{|\lambda_i| < 1} (1 - \lambda_i L)^{-1}}_{\sum \alpha_k L^k} \underbrace{\prod_{|\lambda_i| > 1} \left(1 - \frac{1}{\lambda_i} F\right)^{-1}}_{\sum \beta_k F^k} \left[\prod_{|\lambda_i| > 1} \lambda_i \right] F^n,$$

où $n = \text{card} \{i, |\lambda_i| > 1\}$.

Preuve. En effet, $\forall i$, $(1 - \lambda_i L)^{-1}$ est bien défini, de la forme $\sum_{k \in \mathbb{Z}} a_{i,k} L^k$ et $A(L)^{-1} = \prod_{i=1}^p (1 - \lambda_i L)^{-1}$ est donc aussi défini. Toutefois, $A(L)^{-1}$ peut contenir des termes en L^k , $k > 0$ qui sont des termes concernant le futur

Si $|\lambda_i| < 1$ pour tout i alors $(1 - \lambda_i L)^{-1} = \sum_{k=0}^{+\infty} \lambda_i^k L^k$ et :

$$A(L)^{-1} = \prod_{i=1}^p (1 - \lambda_i L)^{-1} = \sum_{k=0}^{+\infty} a_k L^k \text{ où } \sum_{k=0}^{+\infty} |a_k| < +\infty.$$

Par ailleurs,

$$A(z) = \prod_{i=1}^p (1 - \lambda_i z) \text{ et } A(z)A(z)^{-1} = 1 \Leftrightarrow \prod_{i=1}^p (1 - \lambda_i z) \left(\sum_{k=0}^{+\infty} a_k z^k \right) = 1,$$

de telle sorte que $A(0)A(0)^{-1} = 1 \times a_0 = 1 \Rightarrow a_0 = 1$. S'il existe i tel que $\lambda_i \in \mathbb{C} \setminus \mathbb{R}$ alors $A(L) = (1 - \lambda_i)(1 - \bar{\lambda}_i)P(L)$ et

$$(1 - \lambda_i)^{-1}(1 - \bar{\lambda}_i)^{-1} = \left(\sum_{k=0}^{+\infty} \lambda_i^k L^k \right) \left(\sum_{k=0}^{+\infty} \bar{\lambda}_i^k L^k \right) = \sum_{k=0}^{+\infty} \alpha_k L^k \quad \alpha_k \in \mathbb{R}, \alpha_0 = 1, \sum_{k=0}^{+\infty} |a_k| < +\infty.$$

■

Remarque 31 Si des racines sont inférieures à 1 (en module), cette décomposition fait intervenir le futur de la variable.

Pour déterminer, en pratique, l'inverse d'un polynôme $A(L)$, supposons qu'il s'écrive $A(L) = \prod_{j=1}^p (1 - \lambda_j L)$, de telle sorte que

$$A(L)^{-1} = \prod_{j=1}^p \left(\sum_{k=0}^{+\infty} \lambda_j^k L^k \right)$$

On peut utiliser directement cette méthode de calcul pour p petit ($p = 1, 2$) mais elle s'avère fastidieuse en général. On note,

$$A(L) \left(\sum_{k=0}^{+\infty} a_k L^k \right) = (1 + \varphi_1 L + \dots + \varphi_p L^p) \left(\sum_{k=0}^{+\infty} a_k L^k \right) = 1$$

Les a_k sont obtenus par récurrence puis identification.

$$\phi(L)^{-1} = \prod_{j=1}^p \frac{1}{1 - \lambda_j L}$$

On décompose alors cette fraction rationnelle en éléments simples, $1 = \phi(z)Q_r(z) + z^{r+1}R_r(z)$ avec $\lim_{r \rightarrow +\infty} Q_r(z) = A^{-1}(z)$.

6.3 Compléments sur les séries stationnaires : les autocorrélations

Pour rappels, Un processus (X_t) est **stationnaire** (au second ordre) si pour tout t , $\mathbb{E}(X_t^2) < +\infty$, pour tout t , $\mathbb{E}(X_t) = \mu$, constante indépendante de t et, pour tout t et pour tout h , $\text{cov}(X_t, X_{t+h}) = \gamma(h)$, indépendante de t .

6.3.1 Autocovariance et autocorrélation

Pour une série stationnaire (X_t) , on définit la **fonction d'autocovariance** $h \mapsto \gamma_X(h) = \text{cov}(X_t, X_{t-h})$ pour tout t , et on définit la **fonction d'autocorrélation** $h \mapsto \rho_X(h) = \gamma_X(h) / \gamma_X(0)$ pour tout t , soit

$$\rho_X(h) = \text{corr}(X_t, X_{t-h}) = \frac{\text{cov}(X_t, X_{t-h})}{\sqrt{V(X_t)}\sqrt{V(X_{t-h})}} = \frac{\gamma_X(h)}{\gamma_X(0)}$$

Définition 34 On appellera **matrice d'autocorrélation** du vecteur $(X_t, X_{t-1}, \dots, X_{t-h+1})$

$$\mathcal{R}(h) = \begin{bmatrix} 1 & \rho(1) & \rho(2) & & \rho(h-1) \\ \rho(1) & 1 & \rho(1) & & \rho(h-2) \\ \rho(2) & \rho(1) & 1 & \ddots & \rho(h-3) \\ & & \ddots & \ddots & \ddots \\ \rho(h-1) & \rho(h-2) & \rho(h-3) & & 1 \end{bmatrix} = \begin{bmatrix} & & & & \rho(h-1) \\ & & & & \vdots \\ & & & & \rho(1) \\ [\rho(h-1) & \dots & \rho(1)] & & 1 \end{bmatrix}$$

On peut noter que $\det \mathcal{R}(h) \geq 0$ pour tout $h \in \mathbb{N} \setminus \{0\}$. Cette propriété implique un certain nombre de contraintes sur les $\rho_X(i)$. Par exemple, la relation $\det \mathcal{R}(2) \geq 0$ implique la contrainte suivante sur le couple $(\rho(1), \rho(2))$:

$$[1 - \rho(2)] [1 + \rho(2) - 2\rho(1)^2] \geq 0,$$

ce qui fait qu'il ne peut y avoir de chute brutale de $\rho(1)$ à $\rho(2)$: il est impossible d'avoir $\rho(2) = 0$ si $\rho(1) \geq 1/\sqrt{2}$.

Ces fonctions sont estimées, pour un échantillon X_1, \dots, X_T , de la façon suivante :

$$\hat{\gamma}(h) = \frac{1}{T-h} \sum_{t=1}^{T-h} X_t X_{t-h} \text{ et } \hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)},$$

(quand le processus est centré, sinon, il faut considérer $(X_t - \mu)(X_{t-h} - \mu)$).

6.3.2 Autocorrélations partielles

Les deux précédentes mesures de dépendance entre X_t et X_{t+h} ne faisaient intervenir que les variables X_t et X_{t+h} . Nous allons introduire ici une notion faisant intervenir les variables intermédiaires. Nous supposons, sans perte de généralité que le processus (X_t) est centré : $\mathbb{E}(X_t) = 0$ pour tout t .

Définition 35 Pour une série stationnaire (X_t) , on définit la **fonction d'autocorrélation partielle** $h \mapsto \psi_X(h)$ par

$$\psi_X(h) = \text{corr}(\widehat{X}_t, \widehat{X}_{t-h}),$$

où

$$\begin{cases} \widehat{X}_{t-h} = X_{t-h} - EL(X_{t-h}|X_{t-1}, \dots, X_{t-h+1}) \\ \widehat{X}_t = X_t - EL(X_t|X_{t-1}, \dots, X_{t-h+1}). \end{cases}$$

On regarde ici la projection (ou l'espérance linéaire) les deux valeurs extrêmes X_t et X_{t-h} sur l'ensemble des valeurs intermédiaires $\Xi_{h-1}^{t-1} = \{X_{t-1}, \dots, X_{t-h+1}\}$. Cette projection peut s'écrire, dans le cas de X_t

$$EL(X_t|X_{t-1}, \dots, X_{t-h+1}) = a_1(h-1)X_{t-1} + a_2(h-1)X_{t-2} + \dots + a_{h-1}(h-1)X_{t-h+1}.$$

On peut aussi écrire, en rajoutant X_{t-h} , et en projetant ainsi sur Ξ_h^{t-1} ,

$$EL(X_t|X_{t-1}, \dots, X_{t-h}) = a_1(h)X_{t-1} + a_2(h)X_{t-2} + \dots + a_{h-1}(h)X_{t-h+1} + a_h(h)X_{t-h}. \quad (19)$$

Il est alors possible de montrer que

$$EL(X_t|X_{t-1}, \dots, X_{t-h+1}) = a_1(h-1)X_{t-1} + a_2(h-1)X_{t-2} + \dots + a_{h-1}(h-1)EL(X_{t-h}|X_{t-1}, \dots, X_{t-h+1}).$$

On a alors

$$\sum_{i=1}^{h-1} a_i(h-1)X_{t-i} = \sum_{i=1}^{h-1} a_i(h)X_{t-i} + a_h(h) \sum_{i=1}^{h-1} a_{h-i}(h-1)X_{t-i}.$$

Aussi, on a le résultat suivant, permettant d'obtenir les coefficients de façon récursive

Propriété 26 Pour $j = 1, \dots, h-1$

$$a_j(h) = a_j(h-1) - a_h(h) + a_{h-j}(h-1) \quad (20)$$

Toutefois, cette méthode récursive n'est possible qu'à condition de connaître $a_h(h)$. Pour cela, on peut utiliser le résultat suivant,

Lemme 1 En notant ρ_i le coefficient d'autocorrélation, $\rho_i = \text{corr}(X_t, X_{t-i})$, alors

$$a_h(h) = \frac{\rho(h) - \sum_{i=1}^{h-1} \rho(h-i) a_i(h-1)}{1 - \sum_{i=1}^{h-1} \rho(i) a_i(h-1)}. \quad (21)$$

Preuve. De (19), on peut déduire

$$\rho(h) = \rho(h-1) a_1(h) + \dots + \rho(1) a_{h-1}(h) + a_h(h), \text{ puisque } \rho(0) = 0,$$

c'est à dire

$$a_h(h) = \rho(h) - [\rho(h-1) a_1(h) + \dots + \rho(1) a_{h-1}(h)] = \rho(h) - \sum_{i=1}^{h-1} \rho(h-i) a_i(h).$$

En utilisant (20), on peut écrire

$$a_h(h) = \rho(h) - \left[\sum_{i=1}^{h-1} \rho(h-i) a_i(h-1) - a_h(h) \cdot \sum_{i=1}^{h-1} \rho(i) a_i(h-1) \right].$$

■

On peut d'ailleurs noter que l'on a la relation suivante

$$a(h) = \begin{bmatrix} a_1(h) \\ \vdots \\ a_h(h) \end{bmatrix} = \mathcal{R}(h)^{-1} \begin{bmatrix} \rho(1) \\ \vdots \\ \rho(h) \end{bmatrix}.$$

Définition 36 L'algorithme récursif basé sur (20), (21) et la condition initiale $a_1(1) = \rho(1)$ est appelé **algorithme de Durbin**.

De (20) on peut en déduire en particulier que

$$a_1(h) = a_1(h-1) + a_h(h) a_{h-1}(h-1),$$

et de (21), que pour $h \geq 2$,

$$a_h(h) = \frac{-\rho(1) a_{h-1}(h-1)}{1 - \rho(1) a_1(h-1)}.$$

Ces deux équation permettent d'obtenir récursivement les deux coefficients extrêmes $a_1(h)$ et $a_h(h)$ pour tout h .

Propriété 27 Soit (X_t) un processus stationnaire, alors $\psi_X(0) = 1$, et, pour $h \geq 1$, $\psi_X(h)$ est le coefficient relatif à X_{t-h} dans la projection de X_t sur $X_{t-1}, \dots, X_{t-h+1}, X_{t-h}$, soit $a_h(h)$.

Preuve. Cette propriété s'obtient en notant que

$$EL(X_t | X_{t-1}, \dots, X_{t-h}) - EL(X_t | X_{t-1}, \dots, X_{t-h+1}) = a_h(h) [X_{t-h} - EL(X_{t-h} | X_{t-1}, \dots, X_{t-h+1})].$$

■

Théorème 6 Il est équivalent de connaître la fonction d'autocorrélation $(\rho_X(h))$ ou la fonction d'autocorrélation partielle $(\psi_X(h))$.

Preuve. (i) L'algorithme de Durbin a montré que la connaissance des $\rho_X(h)$ permet de construire de façon récursive les fonctions $\psi_X(h)$.

(ii) Réciproquement, la relation inverse s'obtient par récurrence, en notant que $a_1(1) = \psi_X(1) = \rho_X(1)$, et que

$$\mathcal{R}(h-1) \begin{bmatrix} a_1(h) \\ \vdots \\ a_{h-1}(h) \end{bmatrix} + \begin{bmatrix} \rho(h-1) \\ \vdots \\ \rho(1) \end{bmatrix} a_h(h) = \begin{bmatrix} \rho(1) \\ \vdots \\ \rho(h-1) \end{bmatrix},$$

et

$$\begin{bmatrix} \rho(h-1) & \dots & \rho(1) \end{bmatrix} \begin{bmatrix} a_1(h) \\ \vdots \\ a_{h-1}(h) \end{bmatrix} + a_h(h) = \rho(h).$$

■

Exemple 39 En particulier, on peut noter que

$$\psi_X(1) = \rho_X(1) \text{ et } \psi_X(2) = \frac{\rho_X(2) - \rho_X(1)^2}{1 - \rho_X(1)^2}$$

Une autre formulation consiste à dire que la fonction d'autocorrélation partielle mesure la corrélation entre X_t et X_{t-h} une fois retirée l'influence des variables antérieures à X_{t-h} . En reprenant les notations de la partie précédente,

$$\mathcal{R}(h) = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(h-3) & \rho(h-2) & \rho(h-1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(h-4) & \rho(h-3) & \rho(h-2) \\ \rho(2) & \rho(1) & 1 & \ddots & \rho(h-5) & \rho(h-4) & \rho(h-3) \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \rho(h-3) & \rho(h-4) & \rho(h-5) & \dots & 1 & \rho(1) & \rho(2) \\ \rho(h-2) & \rho(h-3) & \rho(h-4) & \dots & \rho(1) & 1 & \rho(1) \\ \rho(h-1) & \rho(h-2) & \rho(h-3) & \dots & \rho(2) & \rho(1) & 1 \end{bmatrix}$$

et on introduit de façon analogue la matrice $\mathcal{R}^*(h)$ obtenue en remplaçant la dernière colonne de $\mathcal{R}(h)$ par le vecteur $[\rho(1), \dots, \rho(h)]'$,

$$\mathcal{R}^*(h) = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(h-3) & \rho(h-2) & \rho(1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(h-4) & \rho(h-3) & \rho(2) \\ \rho(2) & \rho(1) & 1 & \dots & \rho(h-5) & \rho(h-4) & \rho(3) \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \rho(h-3) & \rho(h-4) & \rho(h-5) & \dots & 1 & \rho(1) & \rho(h-2) \\ \rho(h-2) & \rho(h-3) & \rho(h-4) & \dots & \rho(1) & 1 & \rho(h-1) \\ \rho(h-1) & \rho(h-2) & \rho(h-3) & \dots & \rho(2) & \rho(1) & \rho(h) \end{bmatrix}$$

Il est alors possible de montrer simplement que

$$\psi_X(h) = \frac{|\mathcal{R}^*(h)|}{|\mathcal{R}(h)|} \text{ pour tout } h.$$

6.3.3 Densité spectrale

Comme nous l'avon déjà mentionné, il est équivalent de connaître la fonction d'autocorrélation et la densité spectrale du processus.

Propriété 28 Soit (X_t) un processus stationnaire de la forme $X_t = m + \sum_{j=0}^{+\infty} a_j \varepsilon_{t-j}$ où (ε_t) est un bruit blanc et $\sum_{j=0}^{+\infty} |a_j| < +\infty$, alors $\sum_{h \in \mathbb{Z}} |\gamma_X(h)| < +\infty$.

Preuve.

$$\sum_{h \in \mathbb{Z}} |\gamma_X(h)| = \sum_{h \in \mathbb{Z}} \left| \sum_{j,k} a_j a_k \gamma_\varepsilon(h+j-k) \right|.$$

Or, comme (ε_t) est un bruit blanc,

$$\gamma_\varepsilon(h+j-k) = \begin{cases} 0 & \text{si } h+j-k \neq 0 \\ \sigma_\varepsilon^2 & \text{si } h+j-k = 0, \end{cases}$$

et donc,

$$\sum_{h \in \mathbb{Z}} |\gamma_X(h)| = \sum_{h \in \mathbb{Z}} \left| \sigma_\varepsilon^2 \sum_j a_j a_{h+j} \right| \leq \sigma_\varepsilon^2 \sum_{h,j} |a_j| |a_{h+j}| = \sigma_\varepsilon^2 \left(\sum_j |a_j| \right)^2 < +\infty.$$

■

Propriété 29 La densité spectrale du processus (X_t) est définie par

$$f_X(\omega) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_X(h) \exp(i\omega h) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_X(h) \cos(\omega h).$$

Preuve. En effet,

$$\begin{aligned} f_X(\omega) &= \frac{1}{2\pi} \left[\gamma_X(0) + \sum_{h>0} \gamma_X(h) e^{i\omega h} + \sum_{h<0} \gamma_X(-h) e^{i\omega h} \right] = \frac{1}{2\pi} \left[\gamma_X(0) + \sum_{h>0} \gamma_X(h) e^{i\omega h} + \sum_{h>0} \underbrace{\gamma_X(-h)}_{=\gamma_X(h)} e^{-i\omega h} \right] \\ &= \frac{1}{2\pi} \left[\gamma_X(0) + \sum_{h>0} \gamma_X(h) \underbrace{(e^{i\omega h} + e^{-i\omega h})}_{=2 \cos(\omega h)} \right] = \frac{1}{2\pi} \left[\gamma_X(0) + \sum_{h \neq 0} \gamma_X(h) \cos(\omega h) \right] \\ &= \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_X(h) \cos(\omega h). \end{aligned}$$

■

On peut d'ailleurs noter que si (ε_t) est un bruit blanc de variance σ^2 , on a alors

$$(\varepsilon_t) \sim \mathcal{BB}(0, \sigma^2) \Rightarrow f_\varepsilon(\omega) = \frac{\sigma_\varepsilon^2}{2\pi}.$$

Propriété 30 Avec les notations précédentes, on a le théorème d'injectivité suivant,

$$\forall h \in \mathbb{Z}, \quad \gamma_X(h) = \int_{[-\pi; \pi]} f_X(\omega) e^{-i\omega h} d\omega = \int_{[-\pi; \pi]} f_X(\omega) \cos(\omega h) d\omega.$$

Preuve. En effet,

$$\begin{aligned} \int_{[-\pi; \pi]} f_X(\omega) e^{-i\omega h} d\omega &= \frac{1}{2\pi} \int_{[-\pi; \pi]} \left(\sum_{k \in \mathbb{Z}} \gamma_X(k) e^{i\omega k} \right) e^{-i\omega h} d\omega \\ &= \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \gamma_X(k) \underbrace{\left(\int_{[-\pi; \pi]} e^{i\omega(k-h)} d\omega \right)}_{= \begin{cases} 0 & \text{si } k \neq h \\ 2\pi & \text{si } k = h \end{cases}} \quad (\text{d'après Fubini}) \\ &= \gamma_X(h). \end{aligned}$$

■

Propriété 31 Soient (ε_t) un bruit blanc, et considérons les processus (X_t) et (Y_t) définis par

$$X_t = \sum_{j \in \mathbb{Z}} a_j \varepsilon_{t-j} \quad \text{et} \quad Y_t = \sum_{k \in \mathbb{Z}} b_k X_{t-k} \quad \text{où} \quad \sum_j |a_j|, \sum_j |b_j| < +\infty,$$

alors $Y_t = \sum_{k \in \mathbb{Z}} c_k \varepsilon_{t-k}$, et de plus,

$$f_Y(\omega) = f_X(\omega) \left| \sum_{k \in \mathbb{Z}} b_k e^{i\omega k} \right|^2.$$

Preuve. Il suffit de noter que

$$\begin{aligned} Y_t = \sum_{k \in \mathbb{Z}} b_k X_{t-k} &= \sum_{k \in \mathbb{Z}} b_k \left(\sum_{j \in \mathbb{Z}} a_j \varepsilon_{t-k-j} \right) = \sum_{j, k \in \mathbb{Z}} a_j b_k \varepsilon_{t-(k+j)} \\ &= \sum_{j, h \in \mathbb{Z}} a_j b_{h-j} \varepsilon_{t-h} = \sum_{h \in \mathbb{Z}} \underbrace{\left(\sum_{j \in \mathbb{Z}} a_j b_{h-j} \right)}_{=c_h} \varepsilon_{t-h}. \end{aligned}$$

et de plus, la densité spectrale s'écrit

$$\begin{aligned} f_Y(\omega) &= \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_Y(h) e^{i\omega h} = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \left(\sum_{j, k \in \mathbb{Z}} b_j b_k \gamma_X(h+j-k) \right) e^{i\omega h} \\ &= \frac{1}{2\pi} \sum_{h, j, k \in \mathbb{Z}} b_j b_k \gamma_X(h+j-k) e^{i\omega(h+j-k)} e^{-i\omega j} e^{i\omega k} \\ &= \frac{1}{2\pi} \left(\sum_{l \in \mathbb{Z}} \gamma_X(l) e^{i\omega l} \right) \left(\sum_{j \in \mathbb{Z}} b_j e^{i\omega j} \right) \left(\sum_{k \in \mathbb{Z}} b_k e^{-i\omega k} \right) \\ &= f_X(\omega) \left| \sum_{k \in \mathbb{Z}} b_k e^{i\omega k} \right|^2. \end{aligned}$$

■

6.3.4 Autocorrélations inverses

Cette notion a été introduite en 1972 par Cleveland, et Chatfield en a précisé les principales caractéristiques en 1979. Etant donné un processus (X_t) stationnaire, de fonction d'autocovariance γ_X et de densité spectrale f_X , il se peut que $1/f_X$ soit interprétable comme une densité spectrale (*par exemple dès lors que $1/f_X$ est continue*).

Définition 37 La fonction d'autocovariance inverse γ_{iX} est la fonction d'autocovariance associée au spectre inverse $1/f$,

$$\gamma_{iX}(h) = \int_{-\pi}^{+\pi} \frac{1}{f_X(\omega)} \exp(i\omega h) d\omega \quad \text{ou} \quad \frac{1}{f_X(\omega)} = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_{iX}(h) \exp(-i\omega h).$$

De la même façon que précédemment, on peut alors définir une autocorrélation inverse,

Définition 38 La fonction d'autocorrélation inverse ρ_{iX} est définie par,

$$\rho_{iX}(h) = \frac{\gamma_{iX}(h)}{\gamma_{iX}(0)}.$$

Considérons une série (X_t) stationnaire, de processus d'autocovariance $\gamma(h)$ pour $h \in \mathbb{Z}$. On définit alors la "fonction génératrice d'autocovariance" comme le polynôme (d'opérateurs retards) suivant

$$\Gamma(L) = \dots + \gamma(-1)L^{-1} + \gamma(0)\mathbb{I} + \gamma(1)L + \gamma(2)L^2 + \dots = \sum_{k=-\infty}^{+\infty} \gamma(k)L^k,$$

et de façon similaire, on peut définir la fonction génératrice d'autocorrélation. La fonction génératrice d'autocovariance inverse, notée $\Gamma_i(L)$ est défini par $\Gamma_i(L)\Gamma(L) = \mathbb{I}$ et est telle que

$$\Gamma_i(L) = \dots + \gamma_i(-1)L^{-1} + \gamma_i(0)\mathbb{I} + \gamma_i(1)L + \gamma_i(2)L^2 + \dots = \sum_{k=-\infty}^{+\infty} \gamma_i(k)L^k = \Gamma(L)^{-1}$$

Exemple 40 Dans le cas d'un processus dit ARMA(p, q) (voir partie (6.6)), définit par une relation de la forme $\Phi(L)X_t = \Theta(L)\varepsilon_t$ où ε_t est un bruit blanc, et où Φ et Θ sont respectivement des polynômes de degré p et q . La fonction génératrice d'autocovariance inverse est donnée par

$$\Gamma_i(L) = \frac{\Phi(L)\Phi(L^{-1})}{\Theta(L)\Theta(L^{-1})} \cdot \frac{1}{\sigma^2} \quad \text{où } \sigma^2 \text{ est la variance du bruit blanc } \varepsilon_t$$

Dans le cas où la composante moyenne mobile n'existe pas ($\Theta = \mathbb{I}$, on parle alors de processus AR(p)), on peut alors en déduire simplement que l'autocovariance inverse est donnée par

$$\gamma_i(h) = \frac{\sum_{j=0}^{p-h} \phi_j \phi_{j+h}}{\sum_{j=0}^p \phi_j^2} \quad \text{pour } h \leq p \text{ et } \gamma_i(h) = 0 \text{ pour } h \geq p$$

avec la convention $\phi_0 = 1$. Aussi, pour les processus AR(p), les autocorrélations inverses s'annulent au delà du retard p (de la même façon que les autocorrélations partielles). Bhansali a montré en 1980 que pour un bruit blanc, les autocorrélations inverses empiriques suivent un bruit blanc de loi normale de moyenne nulle et de variance $1/n$. Ainsi, la significativité des coefficients d'autocorrélation inverse peut être testée, au seuil de 5%, en la comparant avec $1.96/\sqrt{n}$.

Il est également possible de définir les autocorrélations partielles inverses (en utilisant une construction analogue à celle développée dans la partie précédente, en remplaçant les ρ par les ρ_i). Comme l'a montré Bhansali (1980 – 1983) et Cleveland et Parzen, les autocorrélations partielles inverses peuvent être obtenus à l'aide de méthodes récursives (proches de celle de Durbin).

Remarque 32 On peut noter la "correspondance" suivante

autocorrélations	autocorrélations partielles
↓	↓
autocorrélations partielles inverses	autocorrélations inverses

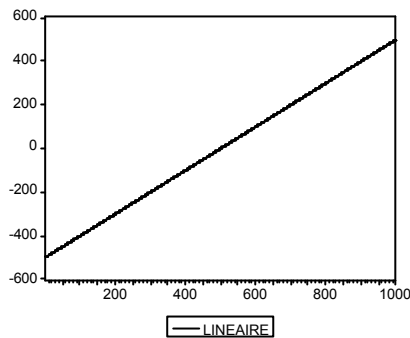
En fait, comme nous le verrons par la suite, sur l'identification des modèles ARMA, les autocorrélations permettent de détecter (entre autres) si une série est intégrée, et s'il faut la différencier, alors que les autocorrélations partielles permettent de vérifier que la série n'a pas été surdifférenciée.

Les autocorrélations ρ et les autocorrélations inverses ρ_i sont identiques si et seulement si X est un bruit blanc

6.3.5 Complément : autocorrélogrammes de fonctions déterministes

Nous allons rappeler ici les formes des autocorrélogrammes, et des autocorrélogrammes partiels de séries non-stationnaires, et déterministes.

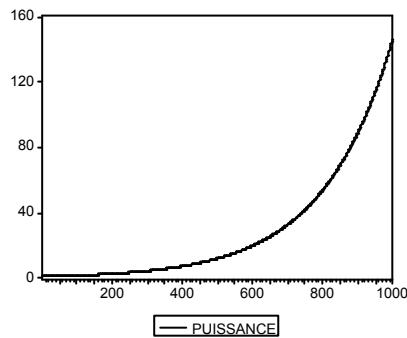
Exemple 41 *Fonction linéaire* $X_t = a + bt$



Sample: 1 100
Included observations: 100

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.970	0.970	96.941	0.000	
2	0.940	-0.015	188.91	0.000	
3	0.910	-0.015	276.00	0.000	
4	0.880	-0.015	359.30	0.000	
5	0.850	-0.015	435.92	0.000	
6	0.820	-0.015	508.96	0.000	
7	0.791	-0.015	577.52	0.000	
8	0.761	-0.015	641.73	0.000	
9	0.731	-0.015	701.70	0.000	
10	0.702	-0.015	757.55	0.000	
11	0.673	-0.015	809.40	0.000	
12	0.643	-0.015	857.39	0.000	
13	0.614	-0.015	901.64	0.000	
14	0.585	-0.015	942.29	0.000	
15	0.557	-0.015	979.48	0.000	
16	0.528	-0.015	1013.4	0.000	
17	0.500	-0.015	1044.1	0.000	
18	0.472	-0.015	1071.7	0.000	
19	0.444	-0.015	1096.5	0.000	
20	0.416	-0.015	1118.6	0.000	
21	0.388	-0.015	1138.1	0.000	
22	0.361	-0.016	1155.1	0.000	

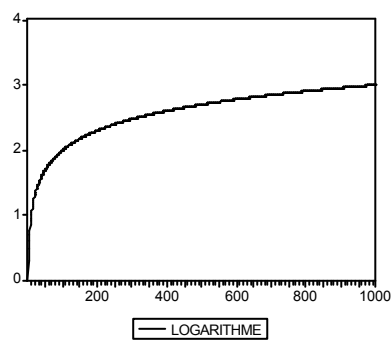
Exemple 42 *Fonction puissance* $X_t = (1 + r)^t$



Sample: 1 100
Included observations: 100

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.970	0.970	96.878	0.000	
2	0.939	-0.015	188.73	0.000	
3	0.909	-0.015	275.65	0.000	
4	0.879	-0.015	357.74	0.000	
5	0.849	-0.015	435.11	0.000	
6	0.819	-0.015	507.87	0.000	
7	0.789	-0.015	576.13	0.000	
8	0.759	-0.015	640.02	0.000	
9	0.729	-0.015	699.65	0.000	
10	0.700	-0.015	755.15	0.000	
11	0.670	-0.015	806.65	0.000	
12	0.641	-0.015	854.28	0.000	
13	0.612	-0.015	898.18	0.000	
14	0.583	-0.015	938.49	0.000	
15	0.554	-0.015	975.34	0.000	
16	0.526	-0.015	1008.9	0.000	
17	0.497	-0.015	1039.3	0.000	
18	0.469	-0.015	1066.6	0.000	
19	0.441	-0.015	1091.1	0.000	
20	0.413	-0.015	1112.9	0.000	
21	0.386	-0.015	1132.2	0.000	
22	0.359	-0.015	1149.0	0.000	

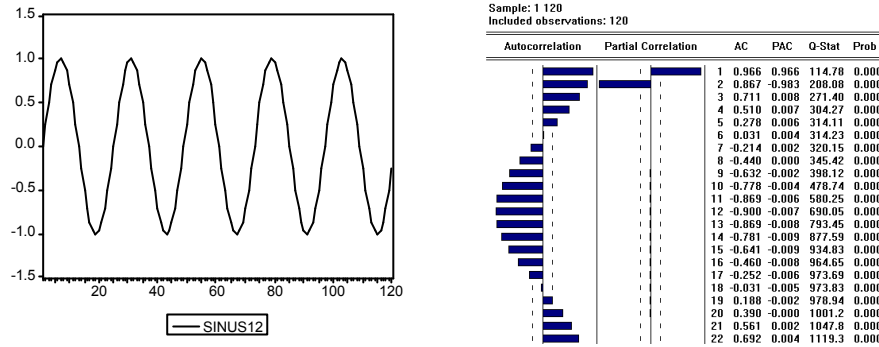
Exemple 43 *Fonction logarithmique* $X_t = \log(t)$



Sample: 1 100
Included observations: 100

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.911	0.911	85.554	0.000	
2	0.844	0.082	159.74	0.000	
3	0.788	0.044	225.11	0.000	
4	0.740	0.028	283.25	0.000	
5	0.696	0.019	335.27	0.000	
6	0.656	0.013	382.00	0.000	
7	0.619	0.009	424.09	0.000	
8	0.585	0.006	462.05	0.000	
9	0.553	0.003	496.30	0.000	
10	0.522	0.002	527.22	0.000	
11	0.493	-0.000	555.11	0.000	
12	0.466	-0.001	580.24	0.000	
13	0.439	-0.003	602.84	0.000	
14	0.414	-0.003	623.12	0.000	
15	0.389	-0.004	641.28	0.000	
16	0.365	-0.005	657.48	0.000	
17	0.342	-0.006	671.88	0.000	
18	0.320	-0.006	684.63	0.000	
19	0.299	-0.007	695.85	0.000	
20	0.278	-0.007	705.68	0.000	
21	0.257	-0.008	714.22	0.000	
22	0.237	-0.008	721.58	0.000	

Exemple 44 *Fonction sinusoïdale* $X_t = \sin(\pi t/12)$



6.4 Les processus autorégressifs : $AR(p)$

Définition 39 On appelle processus autoregressif d'ordre p , noté $AR(p)$, un processus stationnaire (X_t) vérifiant une relation du type

$$X_t - \sum_{i=1}^p \phi_i X_{t-i} = \varepsilon_t \text{ pour tout } t \in \mathbb{Z}, \quad (22)$$

où les ϕ_i sont des réels et (ε_t) est un bruit blanc de variance σ^2 . (22) est équivalent à l'écriture

$$\Phi(L) X_t = \varepsilon_t \text{ où } \Phi(L) = \mathbb{I} - \phi_1 L - \dots - \phi_p L^p$$

Il convient de faire toutefois attention aux signes, certains ouvrages ou logiciels considérant des polynômes de la forme $\mathbb{I} + \phi_1 L + \dots + \phi_p L^p$.

Remarque 33 En toute généralité, un processus $AR(p)$ vérifie une relation de la forme $\Phi(L) X_t = \mu + \varepsilon_t$ où μ est un terme constant. De cette forme générale, il est possible de se ramener à (22) par une simple translation : il suffit de considérer non pas X_t mais $Y_t = X_t - m$ où $m = \mu/\Phi(1)$. En effet, $\Phi(L)(Y_t + m) = \mu + \varepsilon_t$ peut se réécrire $\Phi(L)Y_t + \Phi(1)m = \mu + \varepsilon_t$ c'est à dire $\Phi(L)Y_t = \varepsilon_t$. m correspond ici à l'espérance de (X_t) .

6.4.1 Réécriture de la forme $AR(p)$

Comme nous l'avons vu dans la partie (6.2.2), si l'on souhaite inverser un polynôme (en l'occurrence, présenter X_t comme une fonction des ε_t), il convient de regarder les racines du polynôme Φ , en particulier leur position par rapport à 1 (en module). Comme nous allons le voir dans cette partie, il est possible, lorsque les racines de Φ sont de module différent de 1, qu'il est toujours possible de supposer les racines de module supérieur à 1, quitte à changer la forme du bruit blanc.

Ecriture sous la forme $MA(\infty)$ quand les racines de Φ sont de module strictement supérieur à 1 On suppose $\Phi(L)X_t = \mu + \varepsilon_t$ où $\Phi(L) = 1 - (\varphi_1 L + \dots + \varphi_p L^p)$ et aussi que $|z| \leq 1 \Rightarrow \phi(z) \neq 0$ (de telle sorte que les racines de Φ sont de module strictement supérieur à 1). D'après les résultats énoncés dans la partie sur les polynômes d'opérateurs retards, (X_t) admet une représentation $MA(\infty)$ i.e.

$$X_t = m + \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k} \quad \text{où } a_0 = 1, a_k \in \mathbb{R}, \sum_{k=0}^{+\infty} |a_k| < +\infty.$$

On sait que $\Phi(L)(X_t - m) = \varepsilon_t$, donc $X_t - m = \Phi(L)^{-1}(\varepsilon_t)$.

Propriété 32 Sous ces hypothèses, $\overline{\mathcal{L}}(X_t) = \overline{\mathcal{L}}(\underline{\varepsilon}_t)$, où $\overline{\mathcal{L}}(X_t) = \overline{\mathcal{L}}(1, X_t, X_{t-1}, \dots, X_{t-p}, \dots)$ et $\overline{\mathcal{L}}(\underline{\varepsilon}_t) = \overline{\mathcal{L}}(1, \varepsilon_t, \varepsilon_{t-1}, \dots)$, et de plus (ε_t) est l'innovation de (X_t) .

Preuve. (i) $X_t = \mu + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t$, qui peut se réécrire $X_t = \eta + \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k}$ donc $X_t \in \overline{\mathcal{L}}(\underline{\varepsilon}_t) = \overline{\mathcal{L}}(1, \varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-k}, \dots)$. Donc $\forall k \geq 0, X_{t-k} \in \overline{\mathcal{L}}(\underline{\varepsilon}_{t-k}) \subset \overline{\mathcal{L}}(\underline{\varepsilon}_t)$. On en déduit que $\overline{\mathcal{L}}(1, X_t, X_{t-1}, \dots, X_{t-k}, \dots) \subset \overline{\mathcal{L}}(\underline{\varepsilon}_t)$ et donc $\overline{\mathcal{L}}(X_t) \subset \overline{\mathcal{L}}(\underline{\varepsilon}_t)$. Le second espace étant fermé, on en déduit que $\overline{\mathcal{L}}(X_t) \subset \overline{\mathcal{L}}(\underline{\varepsilon}_t)$.

De la même façon et comme $\varepsilon_t = X_t - (\mu + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p})$, on obtient l'inclusion réciproque et finalement $\overline{\mathcal{L}}(X_t) = \overline{\mathcal{L}}(\varepsilon_t)$.

(ii) L'innovation de (X_t) vaut, par définition, $X_t - \widehat{X}_t$, où

$$\begin{aligned}\widehat{X}_t &= EL(X_t | X_{t-1}) = EL(X_t | 1, X_{t-1}, \dots, X_{t-k}, \dots) \\ &= EL(\underbrace{\mu + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p}}_{\in \overline{\mathcal{L}}(X_{t-1})} + \varepsilon_t | X_{t-1}) = \mu + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + EL(\varepsilon_t | X_{t-1}).\end{aligned}$$

Comme $\overline{\mathcal{L}}(X_{t-1}) = \overline{\mathcal{L}}(\varepsilon_{t-1})$, on a $EL(\varepsilon_t | X_{t-1}) = EL(\varepsilon_t | \varepsilon_{t-1}) = 0$ car (ε_t) est un bruit blanc. Finalement $\widehat{X}_t = \mu + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p}$ et $X_t - \widehat{X}_t = \varepsilon_t$: (ε_t) est bien l'innovation de (X_t) . ■

Si (X_t) est un processus $AR(p)$, $\Phi(L)X_t = \mu + \varepsilon_t$ où les racines de Φ sont à l'extérieur du disque unité, on dit que la représentation $\Phi(L)X_t = \mu + \varepsilon_t$ est la **représentation canonique** de (X_t) .

Écriture sous la forme $MA(\infty)$ quand certaines racines de Φ sont de module strictement inférieur à 1

On suppose que le processus (X_t) s'écrit $\Phi(L)X_t = \mu + \varepsilon_t$ avec

$$\Phi(L) = \prod_{j=1}^p (1 - \lambda_j L) = \left[\prod_{j/|\lambda_j| < 1} (1 - \lambda_j L) \right] \left[\prod_{j/|\lambda_j| > 1} (1 - \lambda_j L) \right]$$

On peut alors montrer que l'on n'aura pas $\overline{\mathcal{L}}(X_t) = \overline{\mathcal{L}}(\varepsilon_t)$, et donc (ε_t) n'est pas l'innovation.

Pour obtenir la représentation canonique il faut changer le polynôme Φ et le bruit blanc. On pose

$$\Phi^*(z) = \left[\prod_{j/|\lambda_j| < 1} (1 - \lambda_j z) \right] \left[\prod_{j/|\lambda_j| > 1} \left(1 - \frac{z}{\lambda_j}\right) \right]$$

de telle sorte que Φ^* a toutes ses racines de module strictement supérieur à 1.

Propriété 33 Soit (η_t) le processus tel que $\eta_t = \Phi^*(L)X_t$. Alors (η_t) est un bruit blanc.

Preuve. En effet, la densité spectrale de (η_t) est $f_\eta(\omega) = f_X(\omega) |\Phi^*(e^{i\omega})|^2$. Et comme $\Phi(L)X_t = \varepsilon_t$, on a aussi :

$$f_X(\omega) |\Phi(e^{i\omega})|^2 = f_\varepsilon(\omega) = \frac{\sigma_\varepsilon^2}{2\pi}$$

On peut alors écrire

$$\begin{aligned}f_\eta(\omega) &= \frac{\sigma_\varepsilon^2}{2\pi} \frac{1}{|\phi(e^{i\omega})|^2} |\phi^*(e^{i\omega})|^2 = \frac{\sigma_\varepsilon^2}{2\pi} \frac{\left[\prod_{j/|\lambda_j| < 1} |1 - \lambda_j e^{i\omega}|^2 \right] \left[\prod_{j/|\lambda_j| > 1} \left|1 - \frac{e^{i\omega}}{\lambda_j}\right|^2 \right]}{\left[\prod_{j/|\lambda_j| < 1} |1 - \lambda_j e^{i\omega}|^2 \right] \left[\prod_{j/|\lambda_j| > 1} |1 - \lambda_j e^{i\omega}|^2 \right]} \\ &= \frac{\sigma_\varepsilon^2}{2\pi} \prod_{j, |\lambda_j| > 1} \frac{1}{|\lambda_j|^2} \underbrace{\frac{|\lambda_j - e^{i\omega}|^2}{|1 - \lambda_j e^{i\omega}|^2}}_{=1} = \frac{\sigma_\varepsilon^2}{2\pi} \prod_{j, |\lambda_j| > 1} \frac{1}{|\lambda_j|^2}\end{aligned}$$

On a donc

$$f_\eta(\omega) = \frac{\sigma_\varepsilon^2 \alpha}{2\pi} = \frac{\sigma_\eta^2}{2\pi} \text{ avec } \alpha = \prod_{j, |\lambda_j| > 1} \frac{1}{|\lambda_j|^2} < 1$$

et finalement (η_t) est un bruit blanc. ■

La représentation $\Phi^*(L)X_t = \eta_t$ est alors la représentation canonique de (X_t) et (η_t) est l'innovation de (X_t) .

6.4.2 Propriétés des autocorrélations - les équations de Yule-Walker

Le processus (X_t) s'écrit

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t. \quad (23)$$

En multipliant par X_t , on obtient

$$\begin{aligned} X_t^2 &= \phi_1 X_{t-1} X_t + \phi_2 X_{t-2} X_t + \dots + \phi_p X_{t-p} X_t + \varepsilon_t X_t \\ &= \phi_1 X_{t-1} X_t + \phi_2 X_{t-2} X_t + \dots + \phi_p X_{t-p} X_t + \varepsilon_t (\phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t) \\ &= \phi_1 X_{t-1} X_t + \phi_2 X_{t-2} X_t + \dots + \phi_p X_{t-p} X_t + \varepsilon_t^2 + [\phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p}] \varepsilon_t, \end{aligned}$$

d'où, en prenant l'espérance

$$\gamma(0) = \phi_1 \gamma(1) + \phi_2 \gamma(2) + \dots + \phi_p \gamma(p) + \sigma^2 + 0,$$

le dernière terme étant nul car ε_t est supposé indépendant du passé de X_t , $\{X_{t-1}, X_{t-2}, \dots, X_{t-p}, \dots\}$. De plus, en multipliant (23) par X_{t-h} , en prenant l'espérance et en divisant par $\gamma(0)$, on obtient

$$\rho(h) - \sum_{i=1}^p \phi_i \rho(h-i) = 0 \text{ pour tout } h > 0.$$

Cette suite d'équations définit le système d'équation dit de Yule-Walker :

Propriété 34 Soit (X_t) un processus $AR(p)$ d'autocorrélation $\rho(h)$. Alors

$$\begin{bmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \\ \vdots \\ \rho(p-1) \\ \rho(p) \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \ddots & \rho(p-1) \\ \rho(1) & 1 & \rho(1) & & \rho(p-2) \\ \rho(2) & \rho(1) & 1 & \ddots & \rho(p-3) \\ \ddots & & \ddots & \ddots & \\ \ddots & & & \ddots & 1 & \rho(1) \\ \rho(p-1) & \rho(p-2) & \rho(p-3) & & \rho(1) & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \\ \phi_{p-1} \\ \phi_p \end{bmatrix}$$

De plus les $\rho(h)$ décroissent exponentiellement vers 0.

Preuve. En effet, $\forall h > 0$, $\rho(h) - \phi_1 \rho(h-1) - \dots - \phi_p \rho(h-p) = 0$. Le polynôme caractéristique de cette relation de récurrence est :

$$z^p - \phi_1 z^{p-1} - \dots - \phi_{p-1} z - \phi_p = z^p \left(1 - \frac{\phi_1}{z} - \dots - \frac{\phi_{p-1}}{z^{p-1}} - \frac{\phi_p}{z^p} \right) = z^p \Phi\left(\frac{1}{z}\right),$$

avec $\Phi(L)X_t = \varepsilon_t$ et $\Phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$. Les racines du polynôme caractéristique sont les $\lambda_i = \frac{1}{z_i}$ (les z_i étant les racines de Φ) avec $|\lambda_i| < 1$. La forme générale de la solution est, si z_1, \dots, z_n sont des racines distinctes de Φ de multiplicités respectives m_1, \dots, m_n

$$\rho(h) = \sum_{i=1}^n \sum_{k=0}^{m_i-1} \alpha_{ik} \lambda_i^k h^k$$

c'est à dire que $\rho(h)$ décroît vers 0 exponentiellement avec h . ■

Par inversion, il est possible d'exprimer les ϕ_i en fonction des $\rho(h)$. La méthodologie développée dans la partie (6.3.2) permet d'obtenir les autocorrélations partielles $\psi(h)$. Il est possible de montrer le résultat suivant

Propriété 35 (i) Pour un processus $AR(p)$ les autocorrélations partielles sont nulles au delà de rang p , $\psi(h) = 0$ pour $h > p$.

(ii) Pour un processus $AR(p)$ les autocorrélations inverses sont nulles au delà de rang p , $\rho_i(h) = 0$ pour $h > p$.

Preuve. (i) Si (X_t) est un processus $AR(p)$ et si $\Phi(L)X_t = \mu + \varepsilon_t$ est sa représentation canonique, en notant $\psi(h)$ le coefficient de X_{t-h} dans $EL(X_t | X_{t-1}, \dots, X_{t-h})$ alors,

$$X_t = \mu + \underbrace{\phi_1 X_{t-1} + \dots + \phi_p X_{t-p}}_{\in \mathcal{L}(1, X_{t-1}, \dots, X_{t-p}) \subset \mathcal{L}(1, X_{t-1}, \dots, X_{t-h})} + \varepsilon_t$$

de telle sorte que

$$\begin{aligned} EL(X_t | X_{t-1}, \dots, X_{t-h}) &= \mu + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + EL(\varepsilon_t | X_{t-1}, \dots, X_{t-h}) \\ &= \mu + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + 0 \end{aligned}$$

Aussi, si $h > p$, le coefficient de X_{t-h} est 0. et si $h = p$, le coefficient de X_{t-p} est $\varphi_p \neq 0$.

(ii) Les autocorrélation inverses sont définies par $\rho_i(h) = \gamma_i(h)/\gamma_i(0)$ où

$$\gamma_i(h) = \int_{-\pi}^{\pi} \frac{1}{f_X(\omega)} e^{i\omega h} d\omega.$$

Si $\Phi(L)X_t = \varepsilon_t$ la densité spectrale de (X_t) vérifie

$$f_X(\omega) |\Phi(e^{i\omega})|^2 = f_\varepsilon(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \text{ donc } f_X(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{1}{|\Phi(e^{i\omega})|^2}.$$

Par conséquent,

$$\frac{1}{f_X(\omega)} = \frac{2\pi}{\sigma_\varepsilon^2} |\Phi(e^{i\omega})|^2$$

Aussi, si $\Phi(z) = 1 + \psi z + \dots + \psi_p z^p$ (i.e. $\psi_0 = 1$ et $\psi_k = -\phi_k$),

$$\frac{1}{f_X(\omega)} = \frac{2\pi}{\sigma_\varepsilon^2} \left(\sum_{k=0}^p \psi_k e^{i\omega k} \right) \left(\sum_{k=0}^p e^{-i\omega k} \right) = \frac{2\pi}{\sigma_\varepsilon^2} \sum_{0 \leq k, l \leq p} \psi_k \psi_l e^{i\omega(k-l)},$$

et donc, les autocovariances inverses sont définies par

$$\gamma_i(h) = \frac{2\pi}{\sigma_\varepsilon^2} \sum_{0 \leq k, l \leq p} \psi_k \psi_l \underbrace{\int_{-\pi}^{\pi} e^{i\omega(k-l+h)} d\omega}_{=0 \text{ sauf si } k-l+h=0}.$$

Or $k-l \in [-p; p]$ donc si $h > p$, $\gamma_i(h) = 0$. En revanche si $h = p$,

$$\gamma_i(p) = \frac{4\pi^2}{\sigma_\varepsilon^2} \psi_0 \psi_p = -\frac{4\pi^2}{\sigma_\varepsilon^2} \varphi_p \neq 0.$$

■

Cette méthode pourra être utilisée pour identifier les processus $AR(p)$.

6.4.3 Le processus $AR(1)$

La forme général des processus de type $AR(1)$ est

$$X_t - \phi X_{t-1} = \varepsilon_t \text{ pour tout } t \in \mathbb{Z},$$

où (ε_t) est un bruit blanc de variance σ^2 .

(i) si $\phi = \pm 1$, le processus (X_t) n'est pas stationnaire. Par exemple, pour $\phi = 1$, $X_t = X_{t-1} + \varepsilon_t$ peut s'écrire

$$X_t - X_{t-h} = \varepsilon_t + \varepsilon_{t-1} + \dots + \varepsilon_{t-h+1},$$

et donc $\mathbb{E}(X_t - X_{t-h})^2 = h\sigma^2$. Or pour un processus stationnaire, il est possible de montrer que $\mathbb{E}(X_t - X_{t-h})^2 \leq 4V(X_t)$. Puisqu'il est impossible que pour tout h , $h\sigma^2 \leq 4V(X_t)$, le processus n'est pas stationnaire.

Si $|\phi| \neq 1$, il existe un unique processus stationnaire tel que $X_t - \phi X_{t-1} = \varepsilon_t$ pour tout $t \in \mathbb{Z}$, ou

$$(1 - \phi L) X_t = \varepsilon_t.$$

(ii) si $|\phi| < 1$ alors on peut inverser le polynôme, et

$$X_t = (1 - \phi L)^{-1} \varepsilon_t = \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i} \text{ (en fonction du passé de } (\varepsilon_t)). \quad (24)$$

(iii) si $|\phi| > 1$ alors on peut inverser le polynôme, et

$$X_t = -\frac{1}{\phi} F \left(1 - \frac{1}{\phi} F \right)^{-1} \varepsilon_t = -\sum_{i=1}^{\infty} \phi^{-i} \varepsilon_{t+i} \text{ (en fonction du futur de } (\varepsilon_t)).$$

La représentation canonique est alors

$$X_t - \frac{1}{\phi} X_{t-1} = \eta_t,$$

où

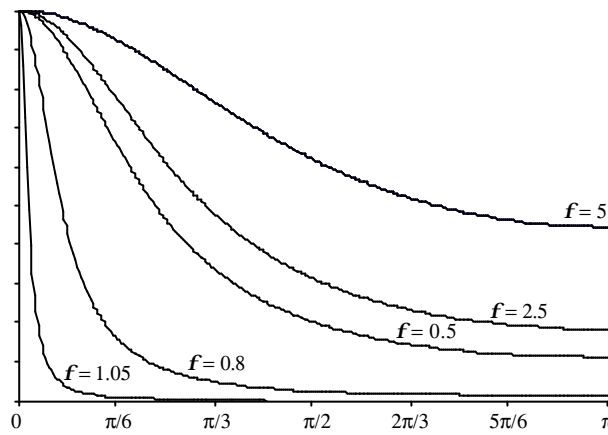
$$\eta_t = (1 - \phi F)(1 - \phi L)^{-1} \varepsilon_t = -\phi \varepsilon_{t+1} + (1 - \phi^2) \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}.$$

Propriété 36 La fonction d'autocorrélation est donnée par $\rho(h) = \phi^h$.

Preuve. Cette expression s'obtient à partir des relations du type (24), ou en notant que $\rho(h) = \phi \rho(h - 1)$. ■
La densité spectrale d'un processus AR(1) est de la forme

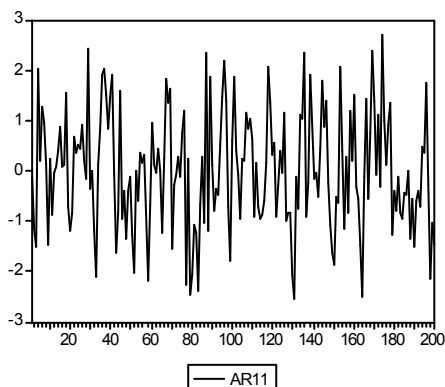
$$f(\omega) = \frac{\sigma^2}{2\pi} \frac{1}{1 + \phi^2 - 2\phi \cos \omega},$$

qui correspond au graphique ci-dessous, avec $\phi > 0$ (les courbes sont ici présentées dans la même base 1 : $f(0) = 1$),



Les valeurs les plus importantes sont obtenues aux basses fréquences, les fonctions ϕ étant décroissantes sur $[0, \pi]$. Dans les trois exemples ci-dessous, les ε_t seront pris gaussiens, $\varepsilon_t \sim \mathcal{N}(0, 1)$

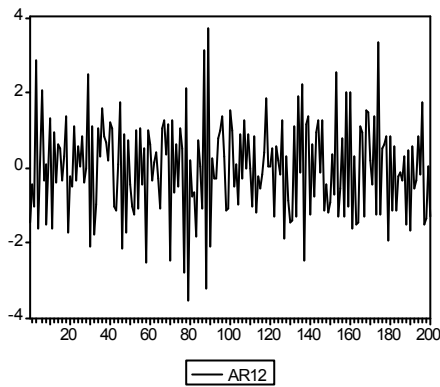
Exemple 45 Processus AR(1), $\phi = 0.2$ - Dans le cas où ϕ est relativement faible



Sample: 1 1500
Included observations: 1500

	Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.181	0.181	49.418	0.000		
2	0.022	-0.011	50.138	0.000		
3	-0.005	-0.007	50.170	0.000		
4	-0.009	-0.007	50.289	0.000		
5	-0.036	-0.034	52.207	0.000		
6	-0.017	-0.005	52.644	0.000		
7	0.008	0.013	52.753	0.000		
8	0.026	0.023	53.805	0.000		
9	0.025	0.016	54.778	0.000		
10	0.024	0.016	55.645	0.000		
11	0.009	0.001	55.759	0.000		
12	-0.061	-0.064	61.313	0.000		
13	-0.085	-0.063	72.326	0.000		
14	-0.043	-0.015	75.180	0.000		
15	0.002	0.015	75.189	0.000		
16	0.028	0.026	76.418	0.000		
17	0.018	0.003	76.887	0.000		
18	-0.011	-0.024	77.084	0.000		
19	-0.010	-0.008	77.244	0.000		
20	-0.019	-0.014	77.821	0.000		
21	0.003	0.017	77.837	0.000		
22	-0.009	-0.005	77.950	0.000		

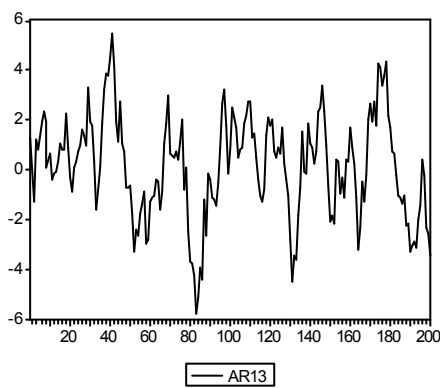
Exemple 46 *Processus AR(1), $\phi = -0.5$ - Dans le cas où ϕ négatif*



Sample: 1 1500
Included observations: 1500

	Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	-0.510	-0.510	390.47	0.000		
2	0.252	-0.011	485.86	0.000		
3	-0.138	-0.018	514.48	0.000		
4	0.079	0.004	523.77	0.000		
5	-0.060	-0.022	529.21	0.000		
6	0.020	-0.028	529.84	0.000		
7	-0.011	-0.009	530.01	0.000		
8	0.017	0.014	530.46	0.000		
9	0.002	0.022	530.46	0.000		
10	0.004	0.014	530.49	0.000		
11	0.018	0.031	531.00	0.000		
12	-0.026	-0.007	531.99	0.000		
13	-0.028	-0.065	533.17	0.000		
14	-0.009	-0.058	533.28	0.000		
15	0.003	-0.019	533.30	0.000		
16	0.009	0.012	533.42	0.000		
17	0.013	0.029	533.66	0.000		
18	-0.020	-0.011	534.25	0.000		
19	0.017	-0.005	534.70	0.000		
20	-0.030	-0.026	535.05	0.000		
21	0.029	0.007	537.34	0.000		
22	-0.032	-0.007	538.89	0.000		

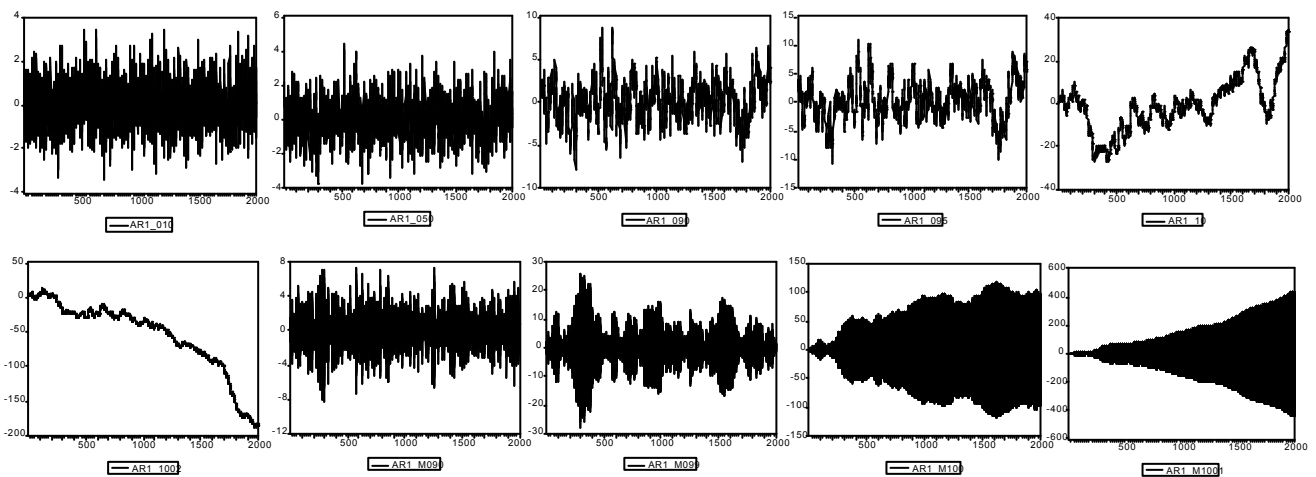
Exemple 47 *Processus AR(1), $\phi = 0.85$ - Dans le cas où ϕ est proche de 1*



Sample: 1 1500
Included observations: 1500

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

Remarque 34 Les graphiques ci-dessous représentent des simulations de processus AR(1), avec différents coefficients ϕ , respectivement, de gauche à droite et de bas en haut 0.1, 0.5, 0.9, 0.95, 1, puis 1.002, -0.5, -0.9, -1 et -1.001,



Comme on peut le voir, on retrouve une marche aléatoire pour ± 1 , et des processus "explosifs" quand $|\phi| > 1$.

A retenir 6 Un processus AR(1) : $X_t = \phi X_{t-1} + \varepsilon_t$ sera auto-corrélé positivement si $0 < \phi < 1$, et autocorrélé négativement si $-1 < \phi < 0$. Cette série va osciller autour de 0, en s'en écartant suivant la valeur ε_t du processus d'innovation (si $-1 < \phi < +1$). Si $\phi = +1$, on obtient une marche aléatoire, et si $\phi > +1$ ou $\phi < -1$ le processus

n'est pas stationnaire, et on obtient un modèle qui explosera (à moyen terme). La valeur ϕ , dans le cas où le processus est stationnaire, est la corrélation entre deux dates consécutives $\phi = \text{corr}(X_t, X_{t-1})$.

$$AR(1) : \begin{cases} \text{Fonction d'autocorrélation} \\ \text{Fonction d'autocorrélation partielle} \end{cases} \begin{cases} \phi > 0 & \text{décroissance exponentielle} \\ \phi < 0 & \text{sinusoïde amortie} \\ \text{première non nulle (signe = signe de } \rho) \\ \text{toutes nulles après} \end{cases}$$

6.4.4 Le processus AR(2)

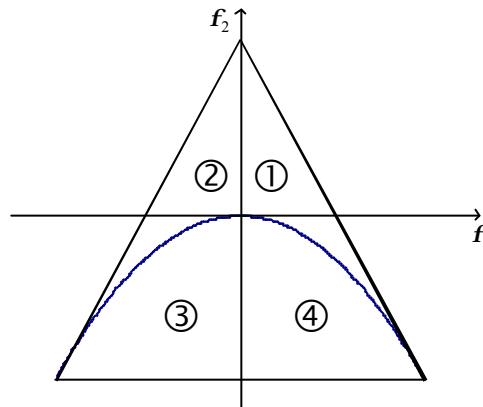
Ces processus sont également appelés modèles de Yule, dont la forme générale est

$$(1 - \phi_1 L - \phi_2 L^2) X_t = \varepsilon_t,$$

où les racines du polynôme caractéristique $\Phi(z) = 1 - \phi_1 z - \phi_2 z^2$ sont supposées à l'extérieur du disque unité (de telle sorte que le processus ε_t corresponde à l'innovation). Cette condition s'écrit

$$\begin{cases} 1 - \phi_1 + \phi_2 > 0 \\ 1 + \phi_1 - \phi_2 > 0 \\ \phi_1^2 + 4\phi_2 > 0, \end{cases}$$

c'est à dire que le couple (ϕ_1, ϕ_2) doit se trouver dans une des 4 parties ci-dessous,



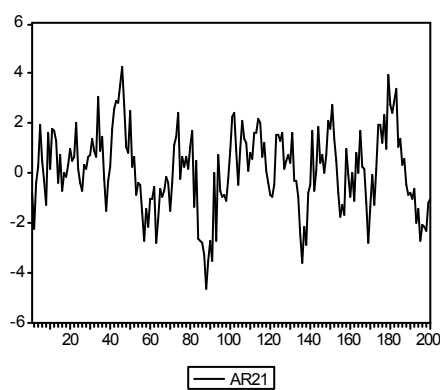
La fonction d'autocorrélation satisfait l'équation de récurrence

$$\rho(h) = \phi_1 \rho(h - 1) + \phi_2 \rho(h - 2) \text{ pour } h \geq 2,$$

et la fonction d'autocorrélation partielle vérifie

$$a(h) = \begin{cases} \rho(1) & \text{pour } h = 1 \\ \left[\rho(2) - \rho(1)^2 \right] / \left[1 - \rho(1)^2 \right] & \text{pour } h = 2 \\ 0 & \text{pour } h \geq 3. \end{cases}$$

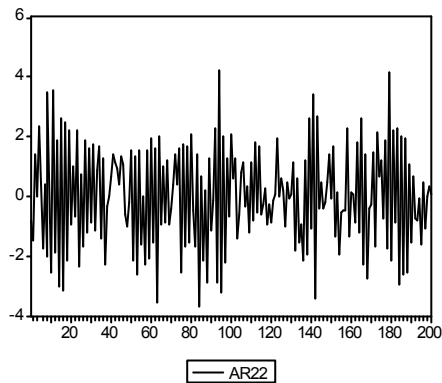
Exemple 48 Processus AR(0.5, 0.3) - cas 1 -



Sample: 1 5000
Included observations: 5000

	Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.699	0.699	2441.3	0.000		
2	0.634	0.286	4455.8	0.000		
3	0.510	-0.013	5757.6	0.000		
4	0.433	-0.000	6694.5	0.000		
5	0.350	-0.017	7307.9	0.000		
6	0.297	0.011	7748.3	0.000		
7	0.249	0.010	8057.8	0.000		
8	0.213	0.009	8284.7	0.000		
9	0.184	0.011	8455.2	0.000		
10	0.160	0.005	8593.8	0.000		
11	0.131	-0.015	8679.3	0.000		
12	0.110	-0.005	8731.3	0.000		
13	0.090	-0.002	8772.3	0.000		
14	0.077	0.005	8802.4	0.000		
15	0.073	0.018	8828.8	0.000		
16	0.053	-0.021	8842.8	0.000		
17	0.048	0.002	8854.3	0.000		
18	0.033	-0.010	8859.8	0.000		
19	0.025	-0.007	8862.9	0.000		
20	0.007	-0.022	8863.1	0.000		
21	0.006	0.010	8863.3	0.000		
22	-0.001	0.004	8863.3	0.000		

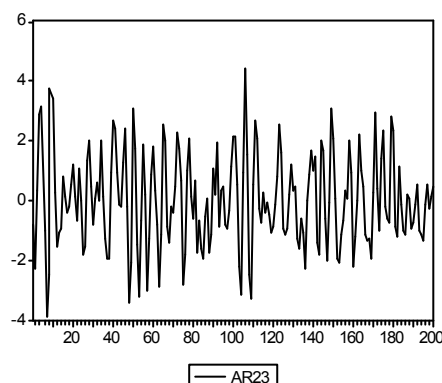
Exemple 49 Processus $AR(-0.5, 0.3)$ - cas 2 -



Sample: 1 5000
Included observations: 5000

	Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
1			1	-0.718	-0.718	2577.6	0.000
2			0.661	0.302	4766.6	0.000	
3			-0.555	-0.015	6305.7	0.000	
4			0.478	-0.006	7450.1	0.000	
5			-0.415	-0.014	8314.1	0.000	
6			0.345	-0.028	8909.1	0.000	
7			-0.298	-0.003	9353.9	0.000	
8			0.247	-0.010	9659.3	0.000	
9			-0.208	0.004	9876.5	0.000	
10			0.178	0.000	10036.	0.000	
11			-0.144	0.012	10140.	0.000	
12			0.115	-0.017	10206.	0.000	
13			-0.092	0.003	10248.	0.000	
14			0.064	-0.020	10269.	0.000	
15			-0.039	0.022	10276.	0.000	
16			0.024	0.003	10279.	0.000	
17			-0.010	0.000	10280.	0.000	
18			0.002	0.002	10280.	0.000	
19			0.009	0.007	10280.	0.000	
20			-0.020	-0.014	10282.	0.000	
21			0.019	-0.012	10284.	0.000	
22			-0.023	0.003	10287.	0.000	

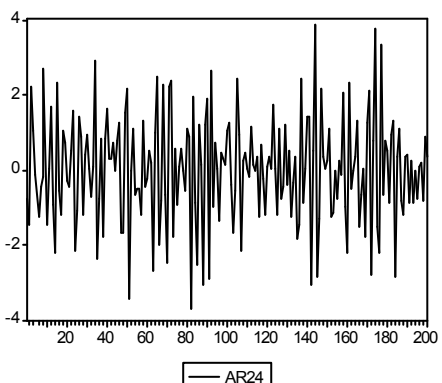
Exemple 50 Processus $AR(-0.5, -0.7)$ - cas 3 -



Sample: 1 5000
Included observations: 5000

	Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
1			1	0.294	0.294	431.72	0.000
2			-0.556	-0.703	1976.5	0.000	
3			0.483	0.004	3142.1	0.000	
4			0.140	-0.003	3252.4	0.000	
5			0.405	-0.015	4072.7	0.000	
6			0.091	-0.006	4114.4	0.000	
7			-0.252	-0.034	4431.8	0.000	
8			-0.198	-0.013	4627.6	0.000	
9			0.084	0.001	4662.8	0.000	
10			0.192	0.004	4846.7	0.000	
11			0.040	0.003	4854.6	0.000	
12			-0.118	0.004	4925.0	0.000	
13			-0.095	-0.011	4970.6	0.000	
14			0.034	-0.000	4976.4	0.000	
15			0.087	-0.005	5014.1	0.000	
16			0.020	-0.003	5016.1	0.000	
17			-0.040	0.027	5024.2	0.000	
18			-0.033	-0.010	5029.8	0.000	
19			0.004	-0.002	5029.9	0.000	
20			0.021	-0.001	5032.0	0.000	
21			0.020	0.018	5033.9	0.000	
22			-0.006	-0.022	5034.1	0.000	

Exemple 51 Processus $AR(0.5, -0.7)$ - cas 4 -



Sample: 1 5000
Included observations: 5000

	Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
1			1	-0.291	-0.291	424.76	0.000
2			-0.560	-0.704	1992.6	0.000	
3			0.483	-0.001	3159.9	0.000	
4			0.155	0.002	3280.8	0.000	
5			-0.425	-0.015	4183.7	0.000	
6			0.102	-0.005	4235.8	0.000	
7			0.241	-0.027	4526.5	0.000	
8			-0.194	-0.013	4714.8	0.000	
9			-0.070	-0.000	4739.6	0.000	
10			0.168	-0.008	4881.3	0.000	
11			-0.031	0.007	4886.1	0.000	
12			-0.091	0.023	4927.6	0.000	
13			0.052	-0.016	4941.1	0.000	
14			0.031	-0.014	4946.0	0.000	
15			-0.034	0.009	4951.8	0.000	
16			-0.012	-0.005	4952.5	0.000	
17			0.023	0.007	4955.2	0.000	
18			0.005	0.000	4955.3	0.000	
19			-0.009	0.021	4955.7	0.000	
20			-0.017	-0.014	4957.1	0.000	
21			0.016	0.000	4958.4	0.000	
22			0.011	-0.008	4959.0	0.000	

Dans le cas d'un modèle $AR(2)$ avec constante, de la forme $(1 - \phi_1 L - \phi_2 L^2) X_t = \phi_0 + \varepsilon_t$ on peut alors noter que l'espérance de X_t est

$$\mathbb{E}(X_t) = \frac{\phi_0}{1 - \phi_1 - \phi_2} \text{ dès lors que } \phi_1 + \phi_2 \neq 1.$$

En utilisant les équations de Yule Walker, nous avons noté que la fonction d'autocorrélation vérifiait la relation de récurrence suivante,

$$\begin{cases} \rho(0) = 1 \text{ et } \rho(1) = \phi_1 / (1 - \phi_2), \\ \rho(h) = \phi_1 \rho(h-1) + \phi_2 \rho(h-2) \text{ pour } h \geq 2, \end{cases}$$

c'est à dire que le comportement de cette suite peut être décrit en étudiant le polynôme caractéristique associé, $x^2 - \phi_1 x - \phi_2 = 0$. Dans le cas où le polynôme admet deux racines réelles, ω_1 et ω_2 où $\omega = \left(\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}\right) / 2$, alors le polynôme autorégressif peut s'écrire $(1 - \phi_1 L - \phi_2 L^2) = (1 - \omega_1 L)(1 - \omega_2 L)$: le modèle $AR(1)$ peut être vu alors comme un modèle $AR(1)$ appliqué à un processus $AR(1)$. L'autocorrélogramme présente une décroissance suivant un mélange d'exponentielles. Quand les racines sont complexes (*conjuguées*), alors les $\rho(h)$ présentent une évolution sinusoïdale amortie. On obtient alors des cycles stochastiques, de longueur moyenne $2\pi / \cos^{-1}(\phi_1/2\sqrt{-\phi_2})$.

A retenir 7 *Le comportement d'un processus $AR(2)$: $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$ dépendra fortement des racines de son équation caractéristique $1 - \phi_1 z - \phi_2 z^2 = 0$. Le cas le plus intéressant est celui où l'équation caractéristique a deux racines complexes conjuguées $r \exp(\pm i\theta)$ pour $r < 1$: le processus est alors stationnaire (et oscille alors autour de 0, sans exploser, de la même façon que les processus $AR(1)$ dans le cas où $|\phi| < 1$). Le processus est alors quasi-cyclique, de fréquence θ , avec un bruit aléatoire.*

$$AR(2) : \begin{cases} \text{Fonction d'autocorrélation} & \text{décroissance exponentielle ou sinusoïde amortie} \\ \text{Fonction d'autocorrélation partielle} & \begin{cases} \text{deux premières non nulles} \\ \text{toutes nulles après} \end{cases} \end{cases}$$

6.5 Les processus moyenne-mobile : $MA(q)$

Définition 40 *On appelle processus moyenne mobile ('moving average') d'ordre q , noté $MA(q)$, un processus stationnaire (X_t) vérifiant une relation du type*

$$X_t = \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \text{ pour tout } t \in \mathbb{Z}, \quad (25)$$

où les θ_i sont des réels et (ε_t) est un bruit blanc de variance σ^2 . (25) est équivalent à l'écriture

$$X_t = \Theta(L) \varepsilon_t \text{ où } \Theta(L) = \mathbb{I} + \theta_1 L + \dots + \theta_q L^q.$$

Remarque 35 *Encore une fois, nous allons utiliser dans cette partie des modèles de la forme (25), toutefois, dans certains ouvrages, la convention est d'écrire ces modèles sous la forme $\Theta(L) = \mathbb{I} - \theta_1 L - \dots - \theta_q L^q$. En particulier pour les logiciels d'économétrie, il convient de vérifier le signe attribué aux coefficients de la forme MA (cf exercice 15 de l'examen de 2002/2003).*

Contrairement aux processus $AR(p)$, les processus $MA(q)$ sont toujours des processus stationnaires. Les processus $MA(\infty)$ sont stationnaires si et seulement si $\sum_{i=1}^{\infty} \theta_i^2$ est finie. Pour rappel, un processus $AR(p)$ est stationnaire si les racines du polynôme retard sont à l'extérieur du cercle unité.

De la même façon que pour les $AR(p)$, il est possible d'inverser le polynôme Θ dans le cas où ses racines sont de module différent de 1 (*quitte à changer des bruit blanc, comme pour les processus AR*). Supposons que Θ n'ait pas de racines de module égal à 1, et considérons le polynôme Θ^* obtenu en remplaçant les racines de Θ de module inférieur à 1 par leur inverse. Le processus (η_t) défini par la relation $X_t = \Theta^*(L) \eta_t$ est là aussi un bruit blanc, dont la variance Σ^2 est définie par

$$\Sigma^2 = \sigma^2 \left[\prod_{i=r+1}^p |\alpha_i|^2 \right]^{-1},$$

où α_i sont les racines de module inférieur à 1. Aussi, la variance de (η_t) est ici supérieure à celle de (ε_t) . Par la suite, on supposera que le processus MA est sous forme canonique, c'est à dire que toutes les racines de Θ sont de module supérieur à 1.

6.5.1 Propriétés des autocorrélations

La fonction d'autocovariance est donnée par

$$\begin{aligned} \gamma(h) &= \mathbb{E}(X_t X_{t-h}) \\ &= \mathbb{E}([\varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}] [\varepsilon_{t-h} + \theta_1 \varepsilon_{t-h-1} + \dots + \theta_q \varepsilon_{t-h-q}]) \\ &= \begin{cases} [\theta_h + \theta_{h+1} \theta_1 + \dots + \theta_q \theta_{q-h}] \sigma^2 & \text{si } 1 \leq h \leq q \\ 0 & \text{si } h > q, \end{cases} \end{aligned}$$

avec, pour $h = 0$, la relation

$$\gamma(0) = [1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2] \sigma^2.$$

Cette dernière relation peut se réécrire

$$\gamma(k) = \sigma^2 \sum_{j=0}^q \theta_j \theta_{j+k} \text{ avec la convention } \theta_0 = 1.$$

D'où la fonction d'autocovariance,

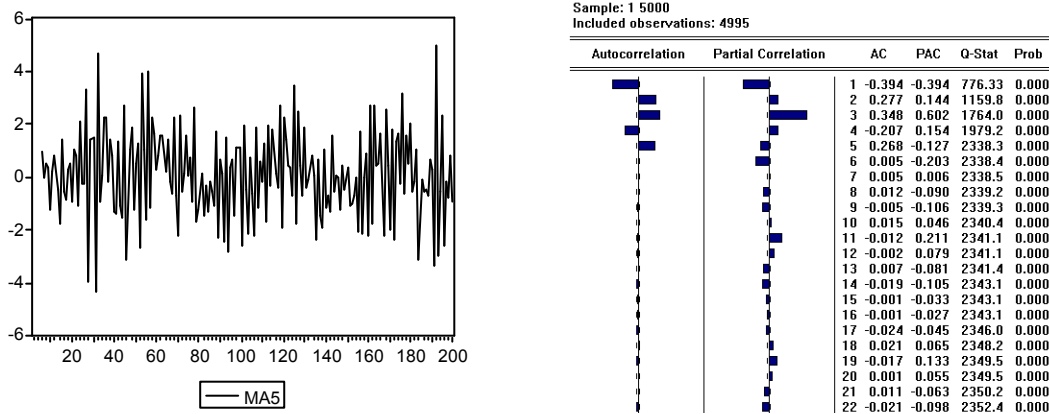
$$\rho(h) = \frac{\theta_h + \theta_{h+1}\theta_1 + \dots + \theta_q\theta_{q-h}}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} \text{ si } 1 \leq h \leq q,$$

et $\rho(h) = 0$ pour $h > q$.

On peut noter en particulier que $\gamma(q) = \sigma^2\theta_q \neq 0$, alors que $\gamma(q+1) = 0$. Cette propriété sera relativement pratique pour faire l'estimation de l'ordre de processus MA .

Exemple 52 *Le graphique ci-dessous montre l'évolution d'un processus $MA(5)$, avec un bruit blanc gaussien, de variance 1, avec à droite, l'autocorrélogramme associé, pour*

$$X_t = \varepsilon_t - 0.7\varepsilon_{t-1} + 0.8\varepsilon_{t-2} + 0.3\varepsilon_{t-3} - 0.1\varepsilon_{t-4} + 0.7\varepsilon_{t-5},$$



On retrouve bien sur cette simulation de processus $MA(5)$ le fait que les autocorrélations s'annulent pour $h > 5$.

A retenir 8 *Les processus MA sont toujours stationnaire, mais ils ne sont inversibles que si les racines du polynôme caractéristique sont à l'extérieur du disque unité.*

On peut noter que s'il n'y a pas de résultat particulier pour les autocorrélations partielles, on a malgré tout le résultat suivant

Propriété 37 *Si (X_t) suit un processus $MA(q)$, $X_t = \Theta(L)\varepsilon_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{t-q}$, alors les autocorrélations inverses $\rho_i(h)$ satisfont les équations de Yule-Walker inverse,*

$$\rho_i(h) + \theta_1\rho_i(h-1) + \dots + \theta_q\rho_i(h-q) = 0 \text{ pour } h = 1, 2, \dots, q.$$

En particulier, $\rho_i(h)$ décroît exponentiellement avec h .

Preuve. Par définition $\rho_i(h) = \gamma_i(h)/\gamma_i(0)$ avec $\gamma_i(h) = \int_{-\pi}^{\pi} \frac{1}{f_X(\omega)} e^{i\omega h} d\omega$. On peut alors écrire, si $X_t = \Theta(L)\varepsilon_t$

$$f_X(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} |\Theta(e^{i\omega})|^2 \text{ et donc } \frac{1}{f_X(\omega)} = \frac{2\pi}{\sigma_\varepsilon^2 |\Theta(e^{i\omega})|^2}.$$

Soit $(Y_t)_{t \in \mathbb{Z}}$ un processus tel que $\Theta(L)Y_t = \eta_t$ i.e. (Y_t) suit un processus $AR(q)$, et

$$\frac{\sigma_\eta^2}{2\pi} = f_Y(\omega) |\Theta(e^{i\omega})|^2.$$

Donc :

$$f_Y(\omega) = \frac{\sigma_\eta^2}{2\pi} \frac{1}{|\Theta(e^{i\omega})|^2},$$

de telle sorte que

$$f_Y(\omega) = \frac{1}{f_X(\omega)} \iff \frac{2\pi}{\sigma_\varepsilon^2} = \frac{\sigma_\eta^2}{2\pi} \iff \sigma_\eta^2 = \frac{4\pi^2}{\sigma_\varepsilon^2}.$$

■

6.5.2 Le processus $MA(1)$

La forme générale des processus de type $MA(1)$ est

$$X_t = \varepsilon_t + \theta\varepsilon_{t-1}, \text{ pour tout } t \in \mathbb{Z},$$

où (ε_t) est un bruit blanc de variance σ^2 . Les autocorrélations sont données par

$$\rho(1) = \frac{\theta}{1 + \theta^2}, \text{ et } \rho(h) = 0, \text{ pour } h \geq 2.$$

On peut noter que $-1/2 \leq \rho(1) \leq 1/2$: les modèles $MA(1)$ ne peuvent avoir de fortes autocorrélations à l'ordre 1. L'autocorrélation partielle à l'ordre h est donnée par

$$\psi(h) = \frac{(-1)^h \theta^h (\theta^2 - 1)}{1 - \theta^{2(h+1)}},$$

et plus généralement, les coefficients de régression sont donnés par

$$a_i(h) = -\frac{(-1)^i \theta^i}{1 - \theta^{2h+2}} + \frac{(-1)^i \theta^{2h+2-i}}{1 - \theta^{2h+2}},$$

dans le cas où $\theta \neq 1$.

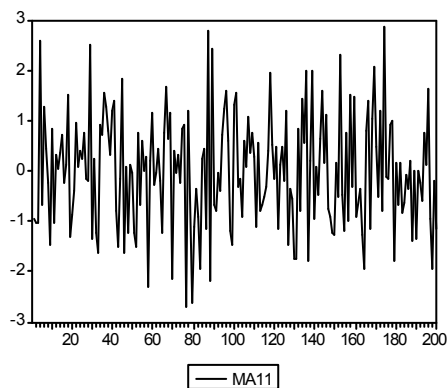
La densité spectrale d'un processus $MA(1)$ est de la forme

$$f_X(\omega) = \frac{\sigma^2}{2\pi} [(1 + \theta^2) + 2\theta \cos \omega],$$

correspondant à un trend déterministe, auquel vient s'ajouter une constante. De cette dernière expression, on peut en déduire aisément que les autocorrélations inverses, dans le cas d'un processus $MA(1)$ vérifient

$$\rho_i(h) = \left[\frac{\theta}{1 + \theta^2} \right]^h \text{ pour tout } h \geq 1.$$

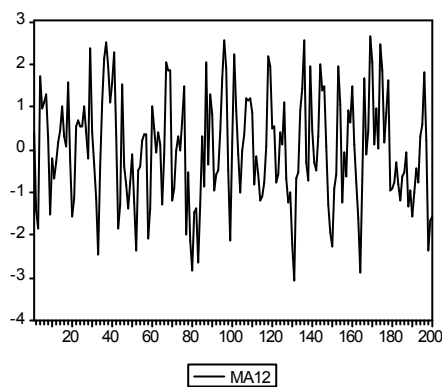
Exemple 53 *Processus $MA(-0.2)$ - Dans le cas où θ est relativement faible*



Sample: 1 1500
Included observations: 1500

	Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	-0.204	-0.204	62.668	0.000		
2	-0.008	-0.052	62.759	0.000		
3	-0.006	-0.019	62.819	0.000		
4	0.007	0.001	62.893	0.000		
5	-0.031	-0.031	64.340	0.000		
6	-0.008	-0.022	64.435	0.000		
7	0.005	-0.003	64.479	0.000		
8	0.017	0.017	64.941	0.000		
9	0.009	0.017	65.059	0.000		
10	0.013	0.020	65.312	0.000		
11	0.022	0.031	66.068	0.000		
12	-0.039	-0.028	68.360	0.000		
13	-0.055	-0.069	72.885	0.000		
14	-0.018	-0.047	73.367	0.000		
15	0.007	-0.011	73.432	0.000		
16	0.022	0.021	74.167	0.000		
17	0.012	0.020	74.388	0.000		
18	-0.015	-0.013	74.733	0.000		
19	0.002	-0.008	74.739	0.000		
20	-0.020	-0.023	75.338	0.000		
21	0.014	0.009	75.630	0.000		
22	-0.013	-0.003	75.872	0.000		

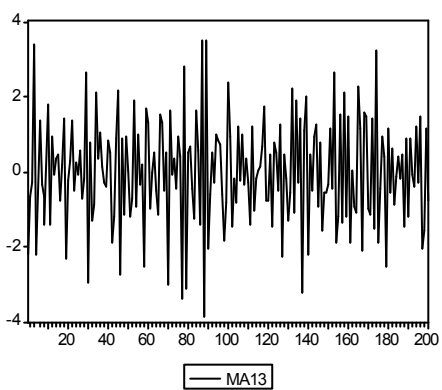
Exemple 54 *Processus* $MA(0.5)$ - Dans le cas où θ positif



Sample: 1 1500
Included observations: 1500

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.383	0.383	220.63	0.000	
2	-0.024	-0.200	221.46	0.000	
3	-0.013	0.090	221.72	0.000	
4	-0.017	-0.063	222.18	0.000	
5	-0.039	-0.008	224.44	0.000	
6	-0.023	-0.007	225.23	0.000	
7	0.011	0.020	225.41	0.000	
8	0.031	0.019	226.85	0.000	
9	0.033	0.018	228.50	0.000	
10	0.032	0.019	230.09	0.000	
11	0.005	-0.015	230.12	0.000	
12	-0.071	-0.075	237.85	0.000	
13	-0.103	-0.040	253.89	0.000	
14	-0.055	-0.006	258.56	0.000	
15	0.005	0.019	258.59	0.000	
16	0.035	0.023	260.49	0.000	
17	0.021	-0.008	261.15	0.000	
18	-0.009	-0.020	261.30	0.000	
19	-0.018	-0.009	261.81	0.000	
20	-0.018	-0.008	262.31	0.000	
21	-0.002	0.017	262.32	0.000	
22	-0.004	-0.007	262.34	0.000	

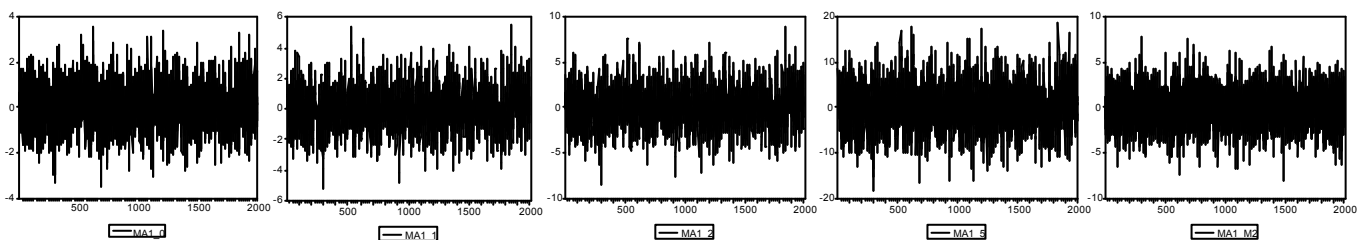
Exemple 55 *Processus* $MA(-0.85)$ - Dans le cas où θ est proche de -1



Sample: 1 1500
Included observations: 1500

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

Remarque 36 Les graphiques ci-dessous représentent des simulations de processus $MA(1)$, avec différents coefficients θ , respectivement, de gauche à droite 0, 1, 2, 5, -1 et -2



Comme on peut le voir, ces processus sont toujours stationnaires, quel que soit θ .

A retenir 9

$$MA(1) : \begin{cases} \text{Fonction d'autocorrélation} \\ \text{Fonction d'autocorrélation partielle} \end{cases} \begin{cases} \text{première non nulle (signe = signe de } \theta) \\ \text{toutes nulles après} \\ \theta > 0 \text{ décroissance exponentielle} \\ \theta < 0 \text{ sinusoïde amortie} \end{cases}$$

6.5.3 Le processus $MA(2)$

La forme générale de (X_t) suivant un processus $MA(2)$ est

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}.$$

La fonction d'autocorrélation est donnée par l'expression suivante

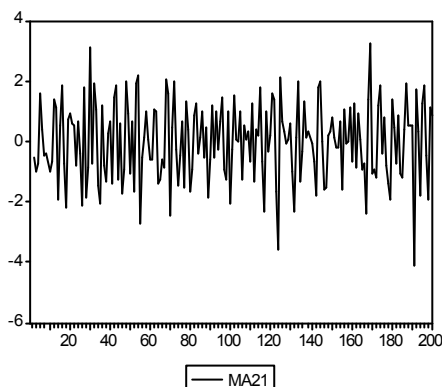
$$\rho(h) = \begin{cases} \theta_1 [1 + \theta_2] / [1 + \theta_1^2 + \theta_2^2] & \text{pour } h = 1 \\ \theta_2 / [1 + \theta_1^2 + \theta_2^2] & \text{pour } h = 2 \\ 0 & \text{pour } h \geq 3, \end{cases}$$

et la densité spectrale est donnée par

$$f_X(\omega) = |1 + \theta_1 e^{i\omega} + \theta_2 e^{2i\omega}|^2.$$

Les configurations possibles sont données dans les 4 exemples ci-dessous

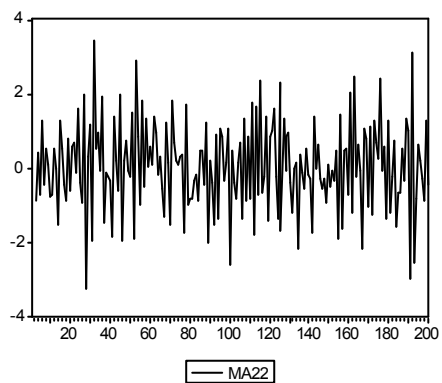
Exemple 56 *Processus* $MA(-0.5, -0.3)$ - cas 1 - les deux premières autocorrélations sont négatives (et nulles ensuite)



Sample: 1 5000
Included observations: 4998

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

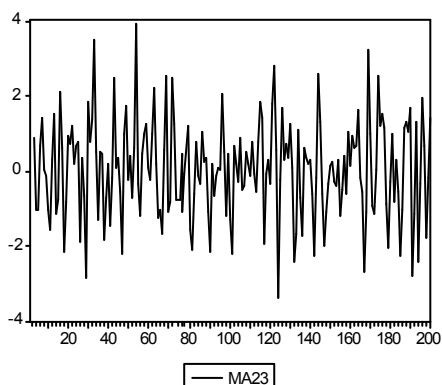
Exemple 57 *Processus* $MA(-0.5, 0.3)$ - cas 2 - les deux premières autocorrélations sont telles que $\rho(1) \leq 0$ et $\rho(2) \geq 0$ (puis nulles ensuite)



Sample: 1 5000
Included observations: 4998

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.488 -0.488	1193.0	0.000	
		2 0.346 0.141	1792.0	0.000	
		3 -0.001 0.205	1792.0	0.000	
		4 -0.002 0.050	1792.0	0.000	
		5 0.002 -0.109	1792.0	0.000	
		6 0.007 -0.073	1792.3	0.000	
		7 -0.007 0.016	1792.5	0.000	
		8 0.013 0.056	1793.3	0.000	
		9 -0.003 0.022	1793.4	0.000	
		10 0.021 0.009	1795.5	0.000	
		11 -0.015 -0.017	1796.6	0.000	
		12 0.004 -0.026	1796.7	0.000	
		13 -0.003 0.009	1796.7	0.000	
		14 -0.008 0.000	1797.1	0.000	
		15 -0.012 -0.016	1797.8	0.000	
		16 0.004 -0.007	1797.9	0.000	
		17 -0.018 -0.011	1799.4	0.000	
		18 0.019 0.018	1801.3	0.000	
		19 -0.008 0.020	1801.6	0.000	
		20 0.002 -0.004	1801.6	0.000	
		21 0.009 -0.004	1802.0	0.000	
		22 -0.021 -0.025	1804.3	0.000	

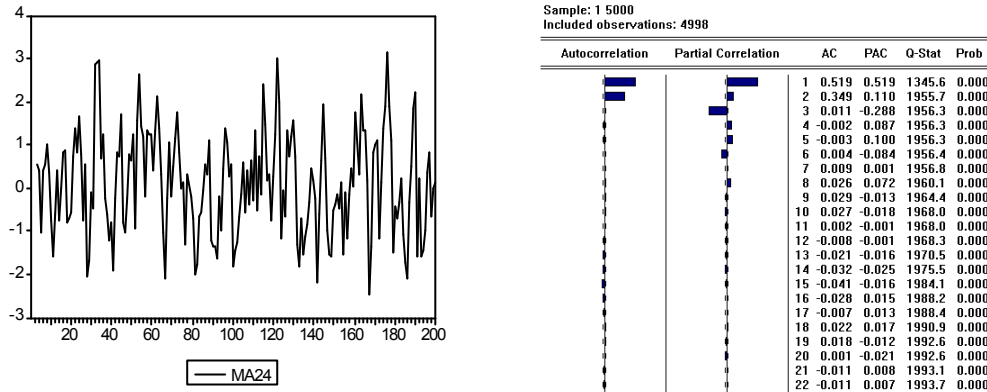
Exemple 58 *Processus* $MA(0.5, -0.7)$ - cas 3 - les deux premières autocorrélations sont telles que $\rho(1) \geq 0$ et $\rho(2) \leq 0$ (puis nulles ensuite)



Sample: 1 5000
Included observations: 4998

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.186 0.186	173.38	0.000	
		2 -0.306 0.353	641.17	0.000	
		3 -0.008 0.159	641.48	0.000	
		4 -0.022 -0.205	643.82	0.000	
		5 0.004 0.145	643.90	0.000	
		6 0.006 -0.141	644.08	0.000	
		7 -0.011 0.106	644.66	0.000	
		8 0.003 -0.097	644.70	0.000	
		9 0.032 0.122	649.86	0.000	
		10 0.023 -0.077	652.44	0.000	
		11 -0.017 0.070	653.95	0.000	
		12 -0.011 0.065	654.62	0.000	
		13 0.010 0.069	655.14	0.000	
		14 -0.009 -0.076	655.56	0.000	
		15 -0.028 0.036	659.37	0.000	
		16 -0.028 -0.084	663.39	0.000	
		17 -0.003 0.055	663.43	0.000	
		18 0.035 -0.031	669.48	0.000	
		19 0.025 0.056	672.50	0.000	
		20 -0.005 -0.041	672.60	0.000	
		21 -0.029 0.013	676.74	0.000	
		22 -0.020 0.036	678.69	0.000	

Exemple 59 *Processus* $MA(0.5, 0.7)$ - cas 4 - les deux premières autocorrélations sont négatives (et nulles ensuite)



6.6 Les processus $ARMA(p, q)$

Cette classe de processus généralise les processus $AR(p)$ et $MA(q)$.

Définition 41 On appelle processus $ARMA(p, q)$, un processus stationnaire (X_t) vérifiant une relation du type

$$X_t - \sum_{i=1}^p \phi_i X_{t-i} = \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \text{ pour tout } t \in \mathbb{Z}, \quad (26)$$

où les θ_i sont des réels et (ε_t) est un bruit blanc de variance σ^2 . (22) est équivalent à l'écriture

$$\Phi(L)X_t = \Theta(L)\varepsilon_t \text{ où } \begin{cases} \Theta(L) = \mathbb{I} + \theta_1 L + \dots + \theta_q L^q \\ \Phi(L) = \mathbb{I} - \phi_1 L - \dots - \phi_p L^p \end{cases}$$

On supposera de plus de les polyômes Φ et Θ n'ont pas de racines en module strictement supérieures à 1 (écriture sous forme canonique), et n'ont pas de racine commune. On supposera de plus que les degrés de Θ et Φ sont respectivement q et p , au sens où $\theta_q \neq 0$ et $\phi_p \neq 0$. On dira dans ce cas que cette écriture est la forme minimale.

Les processus $ARMA(p, q)$ peuvent donc se mettre

(i) sous la forme $MA(\infty)$ en écrivant $X_t = \Phi(L)^{-1} \Theta(L) \varepsilon_t$, si toutes les racines de Φ sont à l'extérieur du disque unité.

(ii) ou sous forme $AR(\infty)$ en écrivant $\Phi(L) \Theta(L)^{-1} X_t = \varepsilon_t$, si toutes les racines de Θ sont à l'extérieur du disque unité.

Remarque 37 Un processus $AR(p)$ est un processus $ARMA(p, 0)$ et un processus $MA(q)$ est un processus $ARMA(0, q)$.

6.6.1 Propriétés des autocorrélations

Propriété 38 Soit (X_t) un processus $ARMA(p, q)$, alors les autocovariances $\gamma(h)$ satisfont

$$\gamma(h) - \sum_{i=1}^p \phi_i \gamma(h-i) = 0 \text{ pour } h \geq q+1. \quad (27)$$

Preuve. La forme $ARMA(p, q)$ de (X_t) est

$$X_t - \sum_{j=1}^p \phi_j X_{t-j} = \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}$$

En multipliant par X_{t-h} , où $h \geq q+1$, et en prenant l'espérance, on obtient (27). ■

De plus, on a la relation suivante

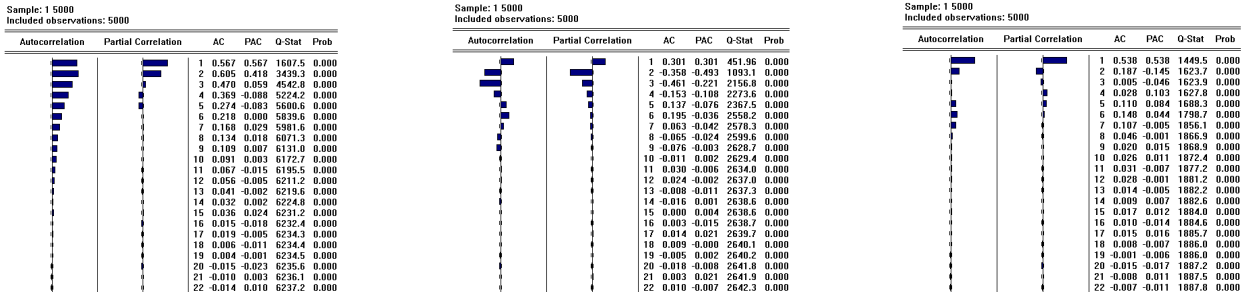
Propriété 39 Soit (X_t) un processus $ARMA(p, q)$, alors les autocorrélations $\gamma(h)$ satisfont

$$\gamma(h) - \sum_{i=1}^p \phi_i \gamma(h-i) = \sigma^2 [\theta_h + h_1 \theta_{h+1} + \dots + h_{q-h} \theta_q] \text{ pour } 0 \leq h \leq q, \quad (28)$$

où les h_i correspondent aux coefficients de la forme $MA(\infty)$ de (X_t) ,

$$X_t = \sum_{j=0}^{+\infty} h_j \varepsilon_{t-j}.$$

Exemple 60 Les sorties ci-dessous correspondent aux autocorrélogrammes de processus $ARMA(2, 1)$, $ARMA(1, 2)$ et $ARMA(2, 2)$ respectivement de gauche à droite



Pour établir une règle quant au comportement de la fonction d'autocorrélation, deux cas sont à envisager,

- Si $p > q$, la fonction d'autocorrélation se comporte comme un mélange de fonctions exponentielles/sinusoïdales amorties
- Si $q \geq p$, le $q-p+1$ premières valeurs de l'autocorrélogramme ont un comportement propre, et pour $k \geq q-p+1$, l'autocorrélogramme tend vers 0.

Des propriétés symétriques existent pour l'autocorrélogramme partiel.

Remarque 38 Avec les notations (26), la variance de X_t est donnée par

$$V(X_t) = \gamma(0) = \frac{1 + \theta_1^2 + \dots + \theta_q^2 + 2\phi_1\theta_1 + \dots + \phi_h\theta_h}{1 - \phi_1^2 - \dots - \phi_p^2} \sigma^2 \text{ où } h = \min(p, q).$$

6.6.2 Densité spectrale des processus $ARMA(p, q)$

Propriété 40 La densité spectrale du processus $ARMA(X_t)$ stationnaire est une fraction rationnelle en $\exp(i\omega)$, et est donnée par

$$f_X(\omega) = \frac{\sigma^2 |\Theta(\exp[i\omega])|^2}{2\pi |\Phi(\exp[i\omega])|^2}.$$

6.6.3 Les processus $ARMA(1, 1)$

Soit (X_t) un processus $ARMA(1, 1)$ défini par

$$X_t - \phi X_{t-1} = \varepsilon_t + \theta \varepsilon_{t-1}, \text{ pour tout } t,$$

où $\phi \neq 0$, $\theta \neq 0$, $|\phi| < 1$ et $|\theta| < 1$. Ce processus peut de mettre sous forme $AR(\infty)$, puisque

$$(1 - \phi L)(1 + \theta L)^{-1} X_t = \Pi(L) X_t = \varepsilon_t,$$

où

$$\Pi(L) = (1 - \phi L) \left[1 - \theta L + \theta^2 L^2 + \dots + (-1)^h \theta^h L^h + \dots \right],$$

aussi

$$\Pi(L) = \sum_{i=0}^{+\infty} \pi_i L^i \text{ où } \begin{cases} \pi_0 = 1 \\ \pi_i = (-1)^i [\phi + \theta] \theta^{i-1} \text{ pour } i \geq 1. \end{cases}$$

La fonction d'autocorrélation s'écrit

$$\begin{cases} \rho(1) = (1 + \phi\theta)(\phi + \theta) / [1 + \theta^2 + 2\phi\theta] \\ \rho(h) = \phi^h \rho(1) \text{ pour } h \geq 2, \end{cases}$$

et la fonction d'autocorrélations partielles a le même comportement qu'une moyenne mobile, avec comme valeur initiale $a(1)\rho(1)$. La sortie ci-dessous montre l'évolution de l'autocorrélogramme dans le cas d'un $ARMA(1, 1)$

Sample: 1 5000
Included observations: 5000

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.388	0.388	752.58	0.000
		2	0.299	0.175	1200.9	0.000
		3	0.229	0.080	1464.1	0.000
		4	0.180	0.042	1626.2	0.000
		5	0.121	-0.003	1699.8	0.000
		6	0.099	0.011	1748.8	0.000
		7	0.076	0.005	1777.9	0.000
		8	0.063	0.008	1797.6	0.000
		9	0.058	0.015	1814.6	0.000
		10	0.053	0.013	1829.0	0.000
		11	0.034	-0.008	1834.8	0.000
		12	0.026	-0.004	1838.0	0.000
		13	0.014	-0.009	1839.0	0.000
		14	0.015	0.004	1840.0	0.000
		15	0.027	0.021	1843.8	0.000
		16	0.007	-0.014	1844.0	0.000
		17	0.014	0.007	1845.0	0.000
		18	0.006	-0.006	1845.2	0.000
		19	0.003	-0.004	1845.2	0.000
		20	-0.019	-0.026	1847.0	0.000
		21	-0.006	0.006	1847.2	0.000
		22	-0.005	0.004	1847.4	0.000

A retenir 10

$$ARMA(1, 1) : \begin{cases} \text{Fonction d'autocorrélation} & \text{décroissance après le premier retard} \\ \text{Fonction d'autocorrélation partielle} & \text{décroissance exponentielle ou sinusoïde amortie} \end{cases}$$

6.7 Introduction aux modèles linéaires non-stationnaires

Tous les processus décrits dans les parties précédantes sont des processus stationnaires (X_t) . En fait, la plupart des résultats obtenus dans les parties précédantes reposent sur l'hypothèse (*très forte*) de stationnarité. C'est le cas par exemple de la loi des grands nombres telle qu'elle est formulée dans le théorème ergodique. On peut toutefois noter qu'un certain nombre de processus, très simples, sont non-stationnaires.

Exemple 61 Marche aléatoire - La marche aléatoire est définie de la façon suivante : soit $\varepsilon_1, \dots, \varepsilon_t, \dots$ une suite de variables i.i.d. et on supposera que ε_t ademet une variance finie, notée σ^2 . On supposera également les ε_t centrés. Une marche aléatoire Y_t vérifie

$$Y_t = Y_{t-1} + \varepsilon_t, \text{ pour tout } t,$$

avec la convention $Y_0 = 0$. On peut noter que $Y_t = Y_0 + Y_1 + \dots + Y_{t-1}$. On a alors

$$\begin{cases} \mathbb{E}(Y_t) = 0 \\ V(Y_t) = \sigma^2 t \end{cases}, \text{ cov}(Y_s; Y_t) = \sigma^2 (s \wedge t) \text{ et } \text{corr}(Y_s, Y_t) = \frac{s \wedge t}{\sqrt{st}} \text{ pour } s, t \geq 0.$$

En notant \mathcal{F}_t la filtration générée par les Y_0, \dots, Y_t , c'est à dire $\mathcal{F}_t = \sigma\{Y_0, \dots, Y_t\}$, on peut montrer que

$$\mathbb{E}(Y_s | \mathcal{F}_t) = Y_t \text{ pour tout } s \geq t \geq 0.$$

Aussi, le processus (Y_t) , muni de sa filtration naturelle, est une martingale. La marche aléatoire est "stationnaire en moyenne", mais pas en variance. La non stationnarité de cette série pose de gros problème statistique : considérons par exemple la moyenne définie sur les n premières observations, soit

$$\bar{Y}_n = \frac{1}{n} \sum_{t=1}^n Y_t,$$

alors, de façon triviale, $\mathbb{E}(\bar{Y}_n) = 0$ mais $V(\bar{Y}_n) = \mathcal{O}(n)$. Plus précisément, la variance de cette moyenne est

$$V(\bar{Y}_n) = \sigma^2 \frac{n(n+1)(2n+1)}{6n^2}.$$

Exemple 62 Tendance linéaire - Un processus "tendance linéaire" est défini de la façon suivante : soit $\varepsilon_1, \dots, \varepsilon_t, \dots$ une suite de variables i.i.d. et on supposera que ε_t admet une variance finie, notée σ^2 . On supposera également les ε_t centrés. Une tendance linéaire Y_t vérifie

$$Y_t = \alpha t + \varepsilon_t \text{ pour tout } t, \text{ où } \alpha \in \mathbb{R}.$$

Ce processus vérifie

$$\begin{cases} \mathbb{E}(Y_t) = \alpha t \\ V(Y_t) = \sigma^2 \end{cases} \quad \text{cov}(Y_s; Y_t) = \text{corr}(Y_s, Y_t) = 0 \text{ pour } s, t \geq 0,$$

et $\mathbb{E}(Y_s | \mathcal{F}_t) = \alpha s$ pour tout $s \geq t \geq 0$. C'est à dire que ce processus n'est pas une martingale, et les variables du processus sont indépendantes (au sens non-corrélées). En notant comme précédemment \bar{Y}_n la moyenne des n premières observations, on a

$$\mathbb{E}(\bar{Y}_n) = \alpha \frac{n+1}{2} \text{ et } V(\bar{Y}_n) = \frac{\sigma^2}{n} \rightarrow 0 \text{ quand } n \rightarrow \infty.$$

Exemple 63 Marché aléatoire avec drift - Ce processus est défini comme mélange des deux précédents : soit X_t une marche aléatoire, soit $X_t = X_{t-1} + \varepsilon_t$, alors Y_t , marche aléatoire avec drift, est défini par

$$\begin{aligned} Y_t &= \alpha t + X_t \text{ pour tout } t, \text{ où } \alpha \in \mathbb{R} \\ &= [\alpha + \varepsilon_1] + [\alpha + \varepsilon_2] + \dots + [\alpha + \varepsilon_t]. \end{aligned}$$

On a alors les propriétés suivantes

$$\begin{cases} \mathbb{E}(Y_t) = \alpha t \\ V(Y_t) = \sigma^2 t \end{cases} \quad \text{cov}(Y_s; Y_t) = \sigma^2 (s \wedge t) \text{ et } \text{corr}(Y_s, Y_t) = \frac{s \wedge t}{\sqrt{st}} \text{ pour } s, t \geq 0,$$

et $\mathbb{E}(Y_s | \mathcal{F}_t) = \alpha s + X_t = \alpha [s - t] + Y_t$ pour tout $s \geq t \geq 0$.

Les processus stationnaires ayant beaucoup de propriétés, il peut apparaître intéressant de trouver une transformation "simple" du processus non-stationnaire que le rendrait stationnaire. La méthode la plus couramment utilisée est de prendre des différences :

Exemple 64 Marche aléatoire - Soit (Y_t) une marche aléatoire,

$$Y_t = Y_{t-1} + \varepsilon_t, \text{ pour tout } t,$$

alors $Z_t = Y_t - Y_{t-1}$ est stationnaire (et $Z_t = \varepsilon_t$).

Exemple 65 Tendance linéaire - Une tendance linéaire Y_t vérifie

$$Y_t = \alpha t + \varepsilon_t, \text{ pour tout } t, \text{ où } \alpha \in \mathbb{R},$$

alors $Z_t = Y_t - Y_{t-1} = \alpha + \varepsilon_t - \varepsilon_{t-1}$: il s'agit d'un processus MA(1) (non inversible, mais stationnaire comme tout processus MA).

Exemple 66 Marché aléatoire avec drift - Soit Y_t , marche aléatoire avec drift,

$$Y_t = \alpha t + X_t = [\alpha + \varepsilon_1] + [\alpha + \varepsilon_2] + \dots + [\alpha + \varepsilon_t],$$

alors $Z_t = Y_t - Y_{t-1} = \alpha + X_t - X_{t-1} = \alpha + \varepsilon_t$ est stationnaire.

C'est cette importance de la différenciation (dont l'intégration est l'opération duale) qui a permis de passer des modèles ARMA aux modèles ARIMA.

6.8 Les processus ARIMA (p, d, q)

L'hypothèse de stationnarité, présente - sous certaines conditions - dans les modèles ARMA, n'est que rarement vérifiée pour des séries économiques. En revanche, on peut considérer les différences premières $\Delta X_t = X_t - X_{t-1}$, ou des différences à des ordres plus élevés

$$\begin{cases} \Delta X_t = X_t - X_{t-1} = (1 - L) X_t \\ \Delta^d X_t = (1 - L)^d X_t \end{cases}$$

Définition 42 Un processus (X_t) est un processus ARIMA (p, d, q) - autorégressif moyenne mobile intégré - s'il vérifie une équation du type

$$\Phi(L)(1-L)^d X_t = \Theta(L)\varepsilon_t \text{ pour tout } t \geq 0$$

où

$$\begin{cases} \Phi(L) = \mathbb{I} - \phi_1 L - \phi_2 L^2 + \dots - \phi_p L^p \text{ où } \phi_p \neq 0 \\ \Theta(L) = \mathbb{I} + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q \text{ où } \theta_q \neq 0 \end{cases}$$

sont des polynômes dont les racines sont de module supérieur à 1, et où les conditions initiales

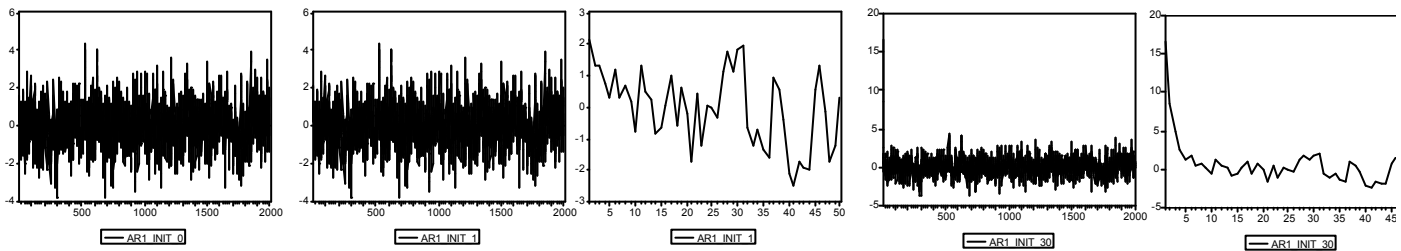
$$Z_{-1} = \{X_{-1}, \dots, X_{-p}, \varepsilon_{-1}, \dots, \varepsilon_{-q}\}$$

sont non-corrélées avec $\varepsilon_0, \dots, \varepsilon_t, \dots$ et où le processus (ε_t) est un bruit blanc de variance σ^2 .

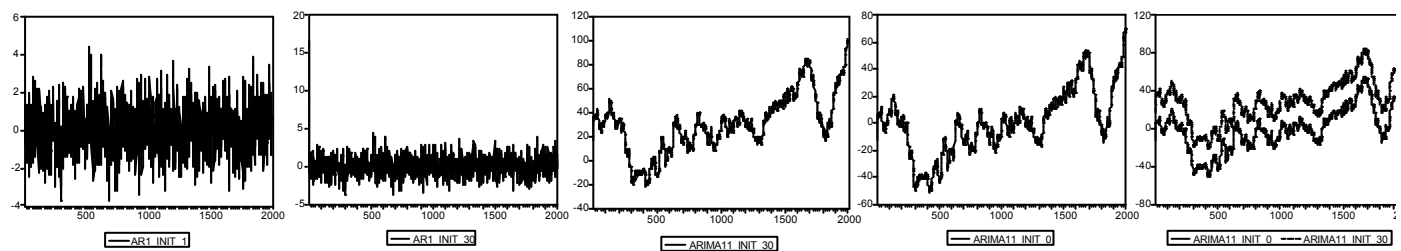
Remarque 39 Si les processus ARMA peuvent être définis sur \mathbb{Z} , il n'en est pas de même pour les processus ARIMA qui doivent commencer à une certaine date ($t = 0$ par convention), avec des valeurs initiales (q valeurs pour les ε_t , et $p+d$ pour X_t). En effet, si l'on considère un processus X_t , ARIMA(0, 1, 0) (= marche aléatoire), soit $(1-L)X_t = \varepsilon_t$. On peut écrire

$$X_t = X_0 + \sum_{k=1}^t \varepsilon_k \text{ mais pas } X_t = \sum_{k=-\infty}^t \varepsilon_k$$

car cette somme ne converge pas dans L^2 . Cette importance de l'initialisation peut se comprendre sur les graphique ci-dessous : considérer un processus AR(1) simulé (ou un processus ARMA de façon plus générale), à partir de la date $t = 0$: on peut noter qu'à relativement court terme les processus (X_t) et (Y_t) simulés respectivement à partir de x et y sont indistinguables : $\mathcal{L}(X_t) = \mathcal{L}(Y_t)$, les deux processus ont la même loi, quelle que soit la valeur initiale (i.e. une loi normale dans le cas d'un bruit blanc gaussien).



En revanche, pour un processus ARIMA(1, 1, 0) - c'est à dire un processus AR(1) intégré, la valeur initiale est très importante : pour deux valeurs initiales différentes, les deux processus n'ont pas la même loi $\mathcal{L}(X_t) \neq \mathcal{L}(Y_t)$,



Les deux processus intégrés, à droite, ont sensiblement des lois différentes à la date t .

Remarque 40 Soit Y_t un processus intégré d'ordre d , au sens où il existe (X_t) stationnaire tel que $Y_t = (1-L)^d X_t$ satisfaisant $\Phi(L)Y_t = \mu + \Theta(L)\varepsilon_t$. Alors, (Y_t) n'est pas un processus ARMA car il ne commence pas en $-\infty$. En fait, (Y_t) est asymptotiquement équivalent à un processus stationnaire ARMA.

Propriété 41 Soit (X_t) un processus ARIMA (p, d, q) alors le processus $(\Delta^d X_t)$ converge vers un processus ARMA (p, q) stationnaire.

6.8.1 Processus ARIMA et formes AR ou MA

Propriété 42 Soit (X_t) un processus ARIMA (p, d, q) de valeurs initiales Z_{-1} , alors

(i) (X_t) peut s'écrire sous la forme suivante, fonction du passé du bruit,

$$X_t = \sum_{j=1}^t h_j \varepsilon_{t-j} + h^*(t) Z_{-1},$$

où les h_j sont les coefficients de la division selon les puissances croissantes de Θ par Φ , et $h^*(t)$ est un vecteur (ligne) de fonctions de t

(ii) (X_t) peut s'écrire sous la forme suivante, fonction du passé de X_t

$$X_t = \sum_{j=1}^t \pi_j X_{t-j} + \bar{h}^*(t) Z_{-1} + \varepsilon_t,$$

où les π_j sont les coefficients (pour $j \geq 1$) de la division selon les puissances croissantes de Φ par Θ , et $\bar{h}^*(t)$ est un vecteur (ligne) de fonctions de t quand $t \rightarrow 0$ quand $t \rightarrow \infty$.

Preuve. (ii) La division selon les puissances croissantes de \mathbb{I} par Θ s'écrit, à l'ordre t ,

$$\mathbb{I} = Q_t(Z) \Theta(Z) + Z^{t+1} R_t(Z) \text{ où } \deg(Q_t) = t \text{ et } \deg(R_t) \leq q-1.$$

Posons $\Psi(L) = (1-L)^d \circ \Phi(L)$. Alors l'équation $\Psi(L) X_t = \Theta(L) \varepsilon_t$ peut s'écrire, en multipliant par $Q_t(Z)$,

$$Q_t(Z) \Psi(L) X_t = Q_t(Z) \Theta(L) \varepsilon_t = [\mathbb{I} - L^{t+1} R_t(L)] \varepsilon_t = \varepsilon_t - R_t(L) \varepsilon_{-1}.$$

En posant $\Pi_t(L) = Q_t(Z) \Psi(L)$ (de degré $p+d+t$ fini) on peut écrire $\Pi_t(L) X_t = \varepsilon_t - R_t(L) \varepsilon_{-1}$, soit

$$\sum_{j=0}^{p+d+t} \pi_j X_{t-j} = \varepsilon_t - \sum_{j=0}^{q-1} r_j \varepsilon_{-1-j},$$

c'est à dire, en coupant la première somme,

$$\sum_{j=0}^t \pi_j X_{t-j} = \varepsilon_t - \underbrace{\sum_{j=t+1}^{t+p+d} \pi_j X_{t-j} - \sum_{j=0}^{q-1} r_j \varepsilon_{-1-j}}_{h'(t) Z_{-1}}.$$

■

6.9 Les modèles SARIMA

Les modèles SARIMA peuvent vus comme une généralisation des modèles ARIMA, contenant une partie saisonnière.

Définition 43 De façon générale, soient s_1, \dots, s_n n entiers, alors un processus (X_t) est un processus SARIMA (p, d, q) - autorégressif moyenne mobile intégré saisonnier - s'il vérifie une équation du type

$$\Phi(L) (1 - L^{s_1}) \dots (1 - L^{s_n}) X_t = \Theta(L) \varepsilon_t \text{ pour tout } t \geq 0$$

où $\Phi(L) = \mathbb{I} - \phi_1 L - \phi_2 L^2 + \dots - \phi_p L^p$ où $\phi_p \neq 0$ et $\Theta(L) = \mathbb{I} + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q$ où $\theta_q \neq 0$. sont des polynômes dont les racines sont de module supérieur à 1, et où les conditions initiales

$$Z_{-1} = \{X_{-1}, \dots, X_{-p}, \varepsilon_{-1}, \dots, \varepsilon_{-q}\}$$

sont non-corrélées avec $\varepsilon_0, \dots, \varepsilon_t, \dots$ et où le processus (ε_t) est un bruit blanc de variance σ^2 .

Cette forme inclue les modèles ARIMA puisqu'il suffit de prendre $n = d$ et $s_1 = \dots = s_n = 1$. Toutefois, les deux formes les plus utilisées sont les suivantes,

$$\begin{aligned} \Phi(L) (1 - L^s) X_t &= \Theta(L) \varepsilon_t \text{ pour tout } t \geq 0 \\ \Phi(L) (1 - L^s) (1 - L)^d X_t &= \Theta(L) \varepsilon_t \text{ pour tout } t \geq 0 \end{aligned}$$

où un seul facteur saisonnier s intervient, soit appliqué à un processus ARMA dans le premier cas, soit appliqué à un processus ARIMA dans le second cas.

Exemple 67 Soit $S \in \mathbb{N} \setminus \{0\}$ correspondant à la saisonnalité, et considérons le processus défini par

$$X_t = (1 - \alpha L)(1 - \beta L^S) \varepsilon_t = \varepsilon_t - \alpha \varepsilon_{t-1} - \beta \varepsilon_{t-S} + \alpha \beta \varepsilon_{t-S-1}.$$

Les autocorrélations sont données par

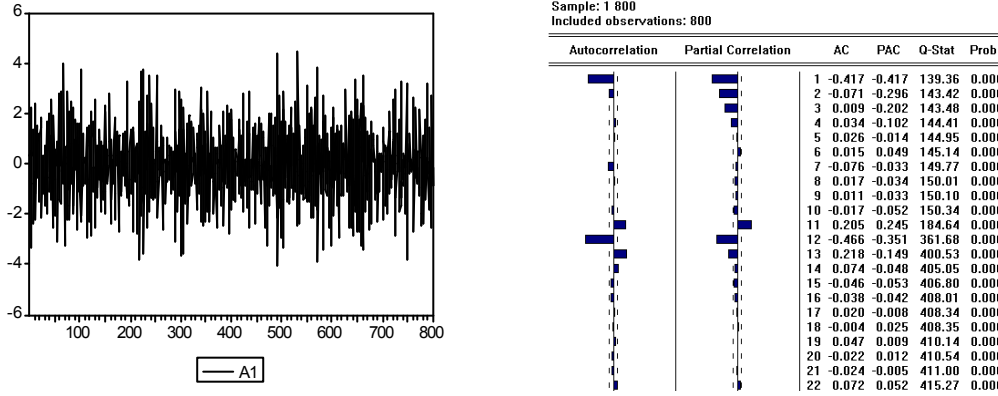
$$\rho(1) = \frac{-\alpha(1 + \beta^2)}{(1 + \alpha^2)(1 + \beta^2)} = \frac{-\alpha}{1 + \alpha^2},$$

$$\rho(S-1) = \frac{\alpha\beta}{(1 + \alpha^2)(1 + \beta^2)},$$

$$\rho(S) = \frac{-\beta(1 + \alpha^2)}{(1 + \alpha^2)(1 + \beta^2)} = \frac{-\beta}{1 + \beta^2},$$

$$\rho(S+1) = \frac{\alpha\beta}{(1 + \alpha^2)(1 + \beta^2)},$$

et $\rho(h) = 0$ ailleurs. On peut noter que $\rho(S-1) = \rho(S+1) = \rho(1) \times \rho(S)$. Le graphique suivant montre l'autocorrélogramme (empirique) d'un tel processus simulé



Pour les autocorrélations partielles, jusqu'en $S-2$ (inclus), la fonction d'autocorrélation partielle est celle d'un MA(1) de paramètre α , puis la fonction est significative en $S-1$, S et $S+1$.

Exemple 68 Soit $S \in \mathbb{N} \setminus \{0\}$ correspondant à la saisonnalité, et considérons le processus défini par

$$(1 - \phi L^S) X_t = (1 - \alpha L)(1 - \beta L^S) \varepsilon_t \text{ ou } X_t - \phi X_{t-S} = \varepsilon_t - \alpha \varepsilon_{t-1} - \beta \varepsilon_{t-S} + \alpha \beta \varepsilon_{t-S-1}.$$

Les autocorrélations sont données par

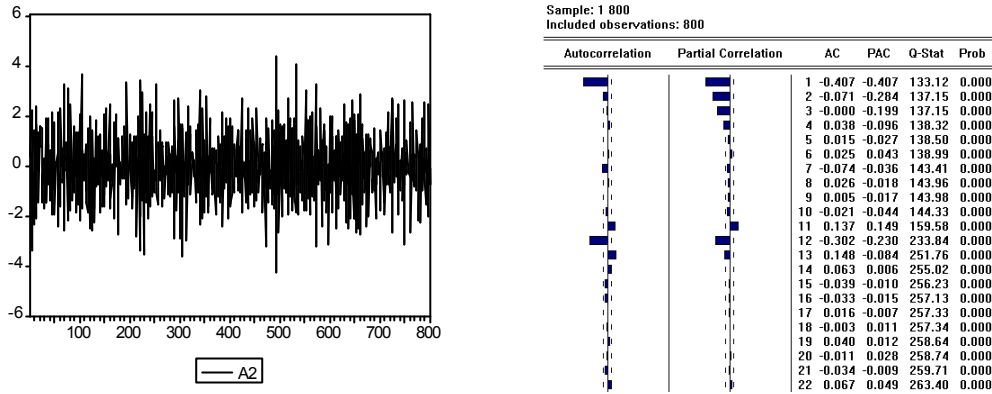
$$\rho(1) = \frac{-\alpha(1 + \beta^2)}{(1 + \alpha^2)(1 + \beta^2)} = \frac{-\alpha}{1 + \alpha^2},$$

$$\rho(S-1) = \frac{\alpha \left[\beta - \phi - \phi(\beta - \phi)^2 / (1 - \phi^2) \right]}{(1 + \alpha^2) \left[1 + (\beta - \phi)^2 / (1 - \phi^2) \right]},$$

$$\rho(S) = \frac{-(1 + \alpha^2)}{\alpha} \rho_{S-1},$$

avec $\rho(h) = 0$ pour $2 \leq h \leq S-2$, puis $\rho(S+1) = \rho(S-1)$ et $\rho(h) = \phi \rho(h-S)$ pour $h \geq S+2$. En particulier

$\rho(kS) = \phi^{k-1}\rho(S)$. Le graphique suivant montre l'autocorrélogramme (empirique) d'un tel processus simulé



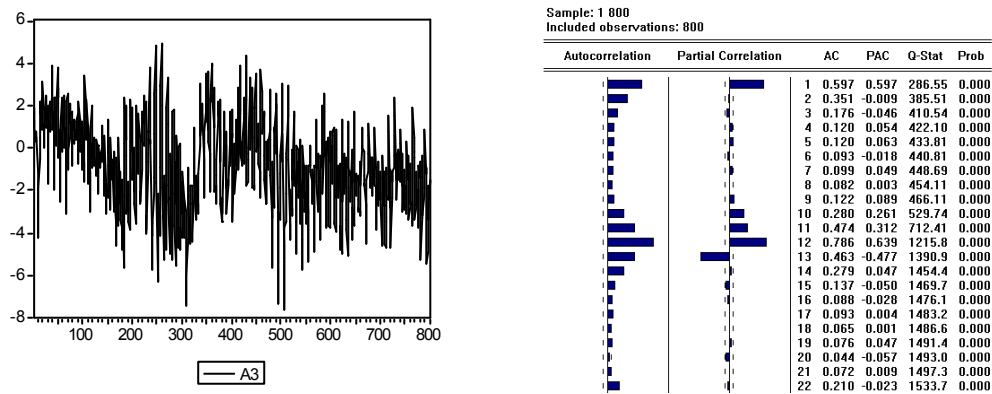
Exemple 69 Soit $S \in \mathbb{N} \setminus \{0\}$ correspondant à la saisonnalité, et considérons le processus défini par

$$(1 - \alpha L)(1 - \beta L^S) X_t = \varepsilon_t \text{ ou } X_t - \alpha X_{t-1} - \beta X_{t-S} + \alpha \beta X_{t-S-1} = \varepsilon_t.$$

Les autocorrélations partielles sont non nulles en 1, S et $S + 1$. De plus la fonction d'autocorrélation vérifie l'équation de récurrence

$$\rho(h) - \alpha \rho(h - 1) - \beta \rho(h - S) + \alpha \beta \rho(h - S - 1) = 0,$$

qui a pour polynôme caractéristique $(z - \alpha)(z^S - \beta)$, qui a pour racines α et les racines S -ièmes de β . Le graphique suivant montre l'autocorrélogramme (empirique) d'un tel processus simulé



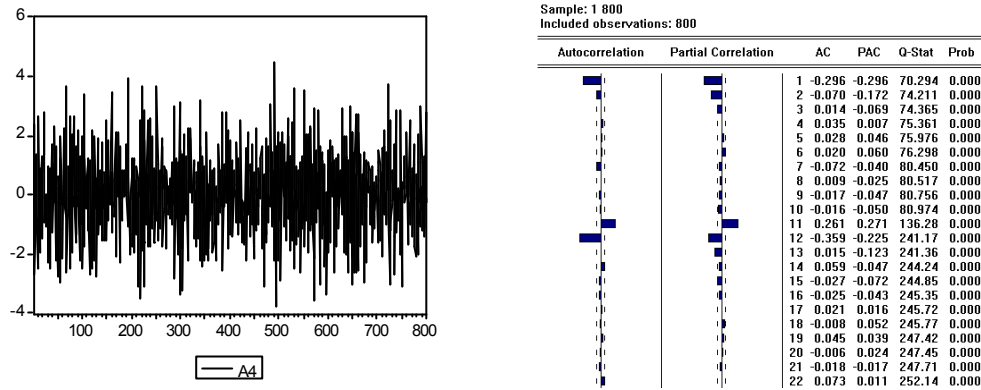
Exemple 70 Soit $S \in \mathbb{N} \setminus \{0\}$ correspondant à la saisonnalité, et considérons le processus défini par

$$X_t = [1 - \alpha L - \beta L^S] \varepsilon_t = \varepsilon_t - \alpha \varepsilon_{t-1} - \beta \varepsilon_{t-S}.$$

On se retrouve dans un cadre assez proche de celui développé dans l'exemple (67), et l'on obtient la fonction d'autocorrélation suivante

$$\rho(1) = \frac{-\alpha}{1 + \alpha^2 + \beta^2}, \rho(S - 1) = \frac{\alpha \beta}{1 + \alpha^2 + \beta^2} \text{ et } \rho(S) = \frac{-\beta}{1 + \alpha^2 + \beta^2}.$$

Le graphique suivant montre l'autocorrélogramme (empirique) d'un tel processus simulé



Exemple 71 Soit $S \in \mathbb{N} \setminus \{0\}$ correspondant à la saisonnalité, et considérons le processus défini par

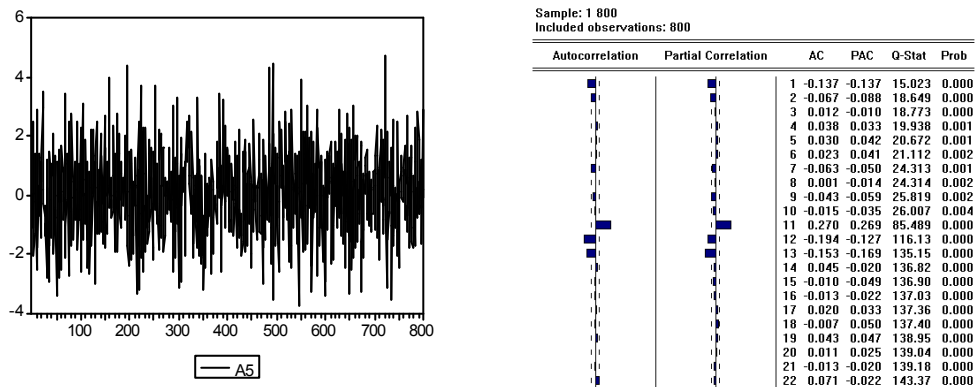
$$X_t = [1 - \alpha L - \beta L^S - \gamma L^{S+1}] \varepsilon_t = \varepsilon_t - \alpha \varepsilon_{t-1} - \beta \varepsilon_{t-S}.$$

On se retrouve dans un cadre assez proche de celui développé dans l'exemple précédent, et l'on obtient la fonction d'autocorrélation suivante

$$\rho(1) = \frac{-\alpha + \beta\gamma}{1 + \alpha^2 + \beta^2 + \gamma^2}, \quad \rho(S-1) = \frac{\alpha\beta}{1 + \alpha^2 + \beta^2 + \gamma^2},$$

$$\rho(S) = \frac{\alpha\gamma - \beta}{1 + \alpha^2 + \beta^2 + \gamma^2} \quad \text{et} \quad \rho(S+1) = \frac{-\gamma}{1 + \alpha^2 + \beta^2 + \gamma^2}.$$

Le graphique suivant montre l'autocorrélogramme (empirique) d'un tel processus simulé



6.10 Théorème de Wold

Théorème 7 Tout processus (X_t) , centré, et stationnaire au second ordre, peut être représenté sous une forme proche de la forme MA

$$X_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} + \eta_t,$$

où

- (1) (ε_t) est l'innovation, au sens où $\varepsilon_t = X_t - EL(X_t | X_{t-1}, X_{t-2}, \dots)$,
- (2) $EL(\varepsilon_t | X_{t-1}, X_{t-2}, \dots) = 0$, $\mathbb{E}(\varepsilon_t X_{t-j}) = 0$, $\mathbb{E}(\varepsilon_t) = 0$, $\mathbb{E}(\varepsilon_t^2) = \sigma^2$ (indépendant de t) et $\mathbb{E}(\varepsilon_t \varepsilon_s) = 0$ pour $t \neq s$,
- (3) toutes les racines de $\Theta(L)$ sont à l'extérieur du cercle unité : le polynôme Θ est inversible,
- (4) $\sum_{j=0}^{\infty} \theta_j^2 < \infty$ et $\theta_0 = 1$,
- (5) les coefficients θ_j et le processus (ε_t) sont uniques,
- (6) (η_t) vérifie $\eta_t = EL(\eta_t | X_{t-1}, X_{t-2}, \dots)$.

La propriété (1) signifie que (ε_t) est l'estimation de X_t comme combinaison linéaire des valeurs passées, et (2) est simplement l'écriture des conditions d'orthogonalité de la projection. La propriété (4) est une conséquence de la stationnarité du processus. Ce théorème se démontre avec les différentes propriétés des espaces de Hilbert. Ce théorème dit juste que (X_t) peut être écrit comme une somme des erreurs de prévision.

Remarque 41 *Ce théorème ne dit pas que les ε_t suivent une loi normale, ou que les ε_t sont i.i.d. (ils ont la même variance et sont non-corrélés).*

Remarque 42 *La propriété (2) dit que $EL(\varepsilon_t|X_{t-1}, X_{t-2}, \dots) = 0$, ce qui ne signifie pas que $\mathbb{E}(\varepsilon_t|X_{t-1}, X_{t-2}, \dots) = 0$. L'écriture $EL(Y|X_{t-1}, X_{t-2}, \dots)$ signifie que l'on recherche la meilleure approximation de Y comme **combinaison linéaire du passé** de X_t , $\alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_h X_{t-h} + \dots$. L'espérance conditionnelle $\mathbb{E}(Y|X_{t-1}, X_{t-2}, \dots)$ est elle la meilleure approximation de Y comme **fonction du passé** de X_t , $g(X_{t-1}, X_{t-2}, \dots, X_{t-h}, \dots)$, où g n'est pas nécessairement linéaire.*

Remarque 43 *Cette représentation n'est unique que parce que l'on a les différentes conditions, en particulier (1) et (3). Par exemple, un processus de la forme $X_t = \eta_t + 2\eta_{t-1}$ où (η_t) est i.i.d. et de variance 1, est stationnaire. Mais sous cette forme, le polynôme MA n'est pas inversible. Pour trouver la représentation de Wold de ce processus, on va chercher θ et ε_t tels que $X_t = \varepsilon_t + \theta\varepsilon_{t-1}$. On peut alors montrer que $V(\varepsilon_t) = 2/\theta$ et que θ est nécessairement soit égal à 2, soit égal à $1/2$. Le cas $\theta = 2$ et $V(\varepsilon_t) = 1$ correspond à l'écriture initiale. Mais $\theta = 1/2$ et $V(\varepsilon_t) = 4$ marche également, et le polynôme MA est alors inversible (comme nous l'avons vu précédemment, il est toujours possible de réécrire un processus MA ou AR de façon inversible, à condition de changer la variance du bruit). Cette représentation est alors la représentation de Wold.*

Remarque 44 *Ce théorème peut s'écrire de façon plus simple si l'on n'est pas intéressé par l'unicité de l'écriture : tout processus (X_t) stationnaire peut se mettre sous forme MA(∞),*

$$X_t = \mu + \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j}.$$

6.11 Théorie spectrale et processus ARIMA

Comme le rappelle Bourbonnais (1998), l'analyse des séries temporelles dans le domaine des fréquences (ou *analyse spectrale*) "est souvent plus riche en terme d'interprétation, mais nécessite un recours à des techniques mathématiques plus complexes". Le principe de base de l'analyse de Fourier est que toute fonction analytique définie sur un intervalle (*fini*) de \mathbb{R} peut s'écrire comme somme pondérée de fonctions sinus et cosinus.

6.11.1 Théorie spectrale et notion de filtre

Théorie spectrale L'analyse spectrale, ou **analyse harmonique**, est une généralisation au cas aléatoire de l'analyse de Fourier. Cette analyse s'appuie sur deux résultats de base : le théorème de Loeve et le théorème de Khintchine. Le premier présente la décomposition harmonique de (X_t) sous la forme

$$X_t = \int_{-\infty}^{+\infty} \exp(i2\pi\lambda t) dU_Z(\lambda),$$

dans laquelle les $dU_Z(\lambda)$ sont des variables aléatoires (*complexes*), alors que le second est équivalent au précédent, mais porte sur la fonction d'autocovariance de (X_t) ,

$$\gamma(h) = \int_{-\infty}^{+\infty} \exp(i2\pi\lambda h) \mathbb{E}(|dU_Z(\lambda)|^2).$$

Théorème 8 (de Khintchine) *La densité spectrale de puissance d'un processus aléatoire stationnaire est égale à la transformée de Fourier de sa fonction d'autocorrélation*

On a alors l'écriture suivante

$$f_X(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} \gamma(h) e^{-i\omega h} \text{ ou } \gamma(h) = \int_0^\pi e^{-i\omega h} f_X(\omega) d\omega, \text{ où } \gamma(h) = \text{cov}(X_t, X_{t-h}),$$

avec $f_X(\omega)$ densité spectrale du processus (X_t) .

Filtre et processus stochastiques Etant donné un processus (X_t) , un **filtre** est une transformation qui associe au processus (X_t) un autre processus $Y_t = F(X_t)$. Par exemple, on dira qu'un filtre est linéaire si $F(\lambda X_t^1 + \mu X_t^2) = \lambda F(X_t^1) + \mu F(X_t^2)$.

De façon générale, on pourra considérer les filtres linéaires de la forme suivante

$$F(X_t) = \sum_{i \in \mathbb{Z}} \phi(i) X_{t+i},$$

où les $\phi(i)$ sont les coefficients de pondération, c'est à dire des filtres moyennes mobiles. Considérons ainsi ϕ une fonction définie sur \mathbb{Z} et à valeurs dans \mathbb{R} (ou \mathbb{C}), appartenant à l'espace des fonctions de carré intégrable sur \mathbb{R} , alors ϕ admet une transformée de Fourier, notée $A(\omega)$ appelée **fonction de réponse en fréquence du filtre** :

$$A(\omega) = \int_{-\infty}^{+\infty} \phi(t) e^{-i\omega t} dt \text{ ou } \phi(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} A(\omega) e^{i\omega t} d\omega.$$

On appellera **gain du filtre** le carré de la norme de la fonction de réponse, $T(\omega) = |A(\omega)|^2$.

6.11.2 Le spectre d'un processus ARMA

D'après le théorème de Wold, un processus stationnaire est une combinaison linéaire infinie des valeurs passées d'un bruit blanc, c'est à dire qu'il peut s'écrire comme un processus MA (∞) :

$$X_t = \Theta(L) \varepsilon_t = \sum_{k=0}^{+\infty} \theta_k \varepsilon_{t-k} \text{ où } \theta_0 = 1.$$

C'est à dire que (X_t) est la réponse à un filtre d'un processus (ε_t) , bruit blanc (*la stabilité étant assurée par la convergence de la somme des carrés de θ_i*). Les θ_i sont alors la fonction de réponse impulsionnelle du filtre. La fonction de gain du filtre s'écrit

$$T(\omega) = |A(\omega)|^2 = \left| \sum_{k=0}^{+\infty} \theta_k e^{-i\omega k} \right|^2,$$

avec $A(\omega)$ correspondant à la fonction de réponse en fréquence au filtre. On a alors la relation suivante entre les spectres des deux processus,

$$f_X(\omega) = T(\omega) f_\varepsilon(\omega).$$

Or, le spectre du bruit blanc vérifie $f_\varepsilon(\omega) = \sigma_\varepsilon^2/2\pi$, et donc

$$f_X(\omega) = \sigma_\varepsilon^2 \left| \sum_{k=0}^{+\infty} \theta_k e^{-i\omega k} \right|^2 = \frac{\sigma_\varepsilon^2}{2\pi} |\Theta(e^{-i\omega k})|^2.$$

De façon générale et analogue, on a le résultat suivant pour les processus ARMA,

Propriété 43 Soit (X_t) un processus ARMA(p, q), vérifiant $\Phi(L)X_t = \Theta(L)\varepsilon_t$, sous forme canonique minimal, avec les racines de Θ et de Φ à l'extérieur du disque unité alors

$$X_t = \frac{\Theta(L)}{\Phi(L)} \varepsilon_t \text{ et } f_X(\omega) = \frac{\sigma_\varepsilon^2 |\Theta(e^{-i\omega k})|^2}{2\pi |\Phi(e^{-i\omega k})|^2}.$$

Cette écriture peut s'écrire sous la forme expensée suivante

$$f_X(\omega) = \frac{\sigma_\varepsilon^2 |1 + \theta_1 e^{-i\omega} + \theta_2 e^{-2i\omega} + \dots + \theta_q e^{-qi\omega}|^2}{2\pi |1 - \phi_1 e^{-i\omega} - \phi_2 e^{-2i\omega} - \dots - \phi_p e^{-pi\omega}|^2}.$$

Compte tenu du lien entre la densité spectrale et la fonction d'autocorrélation, il est possible d'obtenir la densité spectrale dans plusieurs cas simples.

Exemple 72 Considérons le processus $MA(1)$ suivant : $X_t = \varepsilon_t + \theta\varepsilon_{t-1}$ où (ε_t) suit un bruit blanc de variance σ^2 . Pour mémoire, les autocovariances sont données par $\gamma(0) = (1 + \theta^2)\sigma^2$, $\gamma(1) = \theta\sigma^2$ et $\gamma(h) = 0$ pour $h \geq 2$. Ainsi,

$$f(\omega) = \frac{1}{\pi} \left[\gamma(0) + 2 \sum_{k=1}^{+\infty} \gamma(k) \cos(k\omega) \right] = \frac{\sigma^2 (1 + 2\theta \cos(\omega) + \theta^2)}{\pi}.$$

Exemple 73 Considérons le processus $AR(1)$ suivant : $X_t = \phi X_{t-1} + \varepsilon_t$ où (ε_t) suit un bruit blanc de variance σ^2 . Pour mémoire, les autocovariances sont données par $\gamma(0) = \sigma^2 / [1 - \phi^2]$, et $\gamma(h) = \phi\gamma(h-1)$ pour $h \geq 1$. Ainsi, $\gamma(h) = \phi^h \gamma(0)$ pour $h \geq 1$. Cette écriture permet d'obtenir la relation suivante

$$\begin{aligned} f(\omega) &= \frac{1}{\pi} \left[\gamma(0) + 2 \sum_{k=1}^{+\infty} \gamma(k) \cos(k\omega) \right] = \frac{\gamma(0)}{\pi} \left[1 + \sum_{k=1}^{+\infty} \phi^k [e^{i\omega k} + e^{-i\omega k}] \right] \\ &= \frac{\gamma(0)}{\pi} \left[1 + \frac{\phi e^{i\omega}}{1 - \phi e^{i\omega}} + \frac{\phi e^{-i\omega}}{1 - \phi e^{-i\omega}} \right] = \frac{\sigma^2}{\pi [1 - 2\phi \cos(\omega) + \phi^2]}. \end{aligned}$$

ce qui donne une fonction décroissante pour $\phi > 0$ et croissante pour $\phi < 0$.

Exemple 74 Considérons le processus $AR(2)$ suivant : $X_t = \lambda X_{t-1} + \mu X_{t-2} + \varepsilon_t$ où (ε_t) suit un bruit blanc de variance σ^2 . Le polynôme AR s'écrit $\Phi(L) = 1 - \lambda L - \mu L^2$ dont il faut s'assurer que les racines sont à l'extérieur du disque unité. Le spectre du processus (X_t) s'écrit alors

$$f_X(\omega) = \frac{\sigma^2}{2\pi} \frac{1}{1 + \lambda^2 + \mu^2 - 2\lambda(1 - \mu)\cos\omega - 2\mu\cos 2\omega}.$$

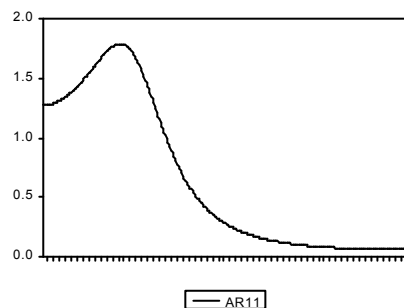
Considérons le cas particulier où $X_t = 0.9X_{t-1} - 0.4X_{t-2} + \varepsilon_t$ avec $\sigma^2 = 1$,

$$f_X(\omega) = \frac{1}{2\pi} \frac{1}{1.97 - 2.52\cos\omega + 0.8\cos 2\omega},$$

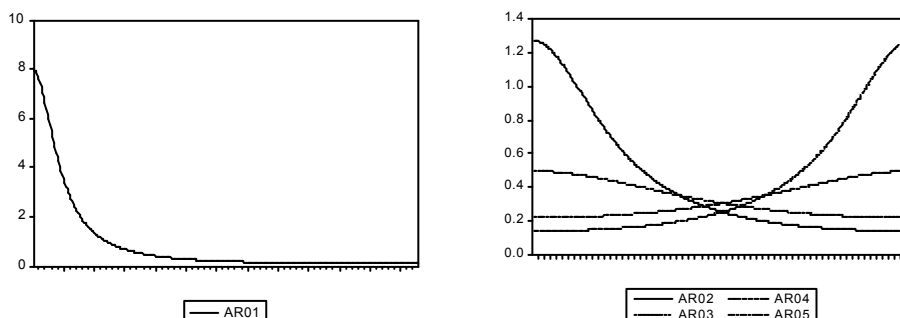
dont le tableau de variation est

$\omega/2\pi$	0	0.212	1
f'	0 +	0	-
f	8	↗ 11.25	↘ 0.38

ce qui donne la représentation suivante



Les graphiques suivants donnent les évolutions de densités spectrales pour différents processus $ARMA$. Les graphiques ci-dessous correspondent à des processus $AR(1)$, avec $\phi = 0.8$ à gauche, puis $\phi = \pm 0.2$ et ± 0.5 à droite, avec des courbes décroissantes quand $\phi \geq 0$ et croissantes quand $\phi \leq 0$,



6.11.3 Estimation de la densité spectrale d'un processus

Supposons que nous ayons T observations d'une série temporelle, avec T impair, soit $T = 2m + 1$. On définit les fréquences (dites parfois "de Fourier") $\omega_j = 2\pi j/T$ pour $j = 1, \dots, m$. Considérons alors le modèle de régression

$$Y_t = \alpha_0 + \sum_{j=1}^m \alpha_j \cos(\omega_j t) + \sum_{j=1}^m \beta_j \sin(\omega_j t),$$

qui peut être écrit sous forme de modèle linéaire $Y = X\theta + \varepsilon$, avec

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_T \end{pmatrix}, X = \begin{pmatrix} 1 & \cos(\omega_1) & \sin(\omega_1) & \cdots & \cos(\omega_m) & \sin(\omega_m) \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & \cos(\omega_1 T) & \sin(\omega_1 T) & \cdots & \cos(\omega_m T) & \sin(\omega_m T) \end{pmatrix}, \theta = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \beta_1 \\ \vdots \\ \alpha_m \\ \beta_m \end{pmatrix} \text{ et } \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{pmatrix}.$$

L'estimation de θ par la méthode des moindres carrés donne $\hat{\theta} = (X'X)^{-1} X'Y$. Compte tenu des propriétés suivantes

$$\sum_{t=1}^T \cos(\omega_j t) = \sum_{t=1}^T \sin(\omega_j t) = 0, \quad \sum_{t=1}^T \cos^2(\omega_j t) = \sum_{t=1}^T \sin^2(\omega_j t) = \frac{T}{2} \text{ pour tout } j, \quad (29)$$

et

$$\sum_{t=1}^T \cos(\omega_j t) \sin(\omega_k t) = \sum_{t=1}^T \cos(\omega_j t) \cos(\omega_k t) = \sum_{t=1}^T \sin(\omega_j t) \sin(\omega_k t) = 0 \text{ pour } j \neq k, \quad (30)$$

on peut montrer aisément que

$$\hat{\theta} = \begin{pmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\alpha}_m \\ \hat{\beta}_m \end{pmatrix} = \begin{pmatrix} T & 0 & \cdots & 0 \\ 0 & T/2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & T/2 \end{pmatrix}^{-1} \begin{pmatrix} \sum Y_t \\ \sum \cos(\omega_1 t) Y_t \\ \vdots \\ \sum \sin(\omega_m t) Y_t \end{pmatrix} = \begin{pmatrix} \bar{Y} = 1/T \cdot \sum Y_t \\ 2/T \cdot \sum \cos(\omega_1 t) Y_t \\ \vdots \\ 2/T \cdot \sum \sin(\omega_m t) Y_t \end{pmatrix},$$

et la variance empirique des observations (Y_t) est donnée (à un facteur $1/T$ près) par

$$\sum_{t=1}^T (Y_t - \bar{Y})^2 = \sum_{j=1}^m \frac{2}{T} \left[\left(\sum_{t=1}^T \cos(\omega_j t) Y_t \right)^2 + \left(\sum_{t=1}^T \sin(\omega_j t) Y_t \right)^2 \right].$$

Tout ceci permet de définir le périodogramme $I(\omega)$ de la façon suivante

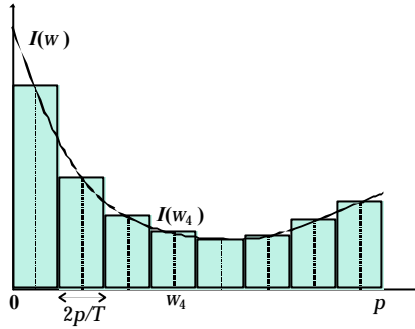
$$I(\omega) = \frac{1}{\pi T} \left[\left(\sum_{t=1}^T \cos(\omega t) Y_t \right)^2 + \left(\sum_{t=1}^T \sin(\omega t) Y_t \right)^2 \right].$$

Dans cette expression, un facteur "1/2π" a été introduit de telle sorte que la variance empirique, donnée par

$$\hat{\gamma}(0) = \frac{1}{T} \sum_{t=1}^T (Y_t - \bar{Y})^2,$$

soit égale à la somme des aires des m rectangles, de hauteur $I(\omega_1), \dots, I(\omega_m)$, et de base $2\pi/T$. La somme des aires

de ces rectangles approxime l'aire sous la courbe $I(\omega)$ pour $0 \leq \omega \leq \pi$,



En utilisant (29) et (30) on peut écrire

$$I(\omega_j) = \frac{1}{\pi} \hat{\gamma}(0) + \frac{2}{\pi} \sum_{k=1}^{T-1} \hat{\gamma}(k) \cos(\omega_j k).$$

Cette fonction $I(\omega)$ est alors la version discrète et empirique de la densité spectrale $f(\omega)$.

Propriété 44 Si le processus est stationnaire, et si la densité spectrale existe, alors $I(\omega)$ est un estimateur sans biais de $f(\omega)$.

Supposons que (Y_t) soit un bruit blanc gaussien, c'est à dire Y_1, \dots, Y_T sont *i.i.d.* et distribués suivant une loi $N(0, \sigma^2)$. Pour toute fréquence de Fourier, $\omega = 2\pi j/T$,

$$I(\omega) = \frac{1}{\pi T} \left[\hat{A}(\omega)^2 + \hat{B}(\omega)^2 \right] \text{ où } \hat{A}(\omega) = \sum_{t=1}^T Y_t \cos(\omega t) \text{ et } \hat{B}(\omega) = \sum_{t=1}^T Y_t \sin(\omega t).$$

Il est facile de montrer que $\hat{A}(\omega)$ et $\hat{B}(\omega)$ sont centrés, et dont les variance sont données par

$$V(\hat{A}(\omega)) = \sigma^2 \sum_{t=1}^T \cos^2(\omega t) = T \frac{\sigma^2}{2} \text{ et } V(\hat{B}(\omega)) = \sigma^2 \sum_{t=1}^T \sin^2(\omega t) = T \frac{\sigma^2}{2},$$

et dont la covariance est nulle

$$\text{cov}(\hat{A}(\omega), \hat{B}(\omega)) = \mathbb{E} \left(\sum_{s,t=1}^T Y_s Y_t \cos(\omega s) \sin(\omega t) \right) = \sigma^2 \left(\sum_{t=1}^T \cos(\omega t) \sin(\omega t) \right) = 0.$$

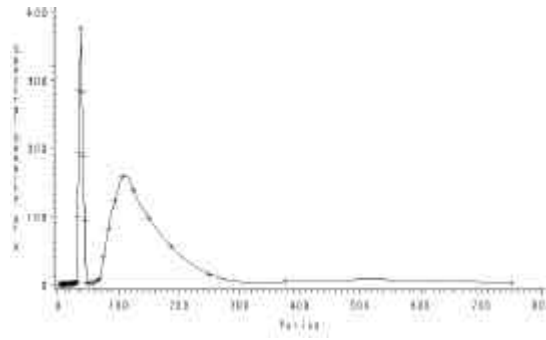
De plus, $\hat{A}(\omega)/\sqrt{2T\sigma^2}$ et $\hat{B}(\omega)/\sqrt{2T\sigma^2}$ sont indépendants, et asymptotiquement distribués suivant une la même loi, normale, centrée et réduite. Et donc, $2 \left[\hat{A}(\omega)^2 + \hat{B}(\omega)^2 \right] / T\sigma^2$ est asymptotiquement distribué suivant une loi du chi-deux, à 2 degrés de liberté. Aussi, $I(\omega) \sim (\sigma^2/\pi) \chi^2(2)/2$, ce qui montre bien que $I(\omega)$ est un estimateur sans biais de $f(\omega) = \sigma^2/\pi$, mais il n'est pas consistant puisque $V(I(\omega)) = \sigma^4/\pi^2 \rightarrow 0$ quand $T \rightarrow \infty$.

Propriété 45 Soit (Y_t) un processus gaussien, de spectre $f(\omega)$ Soit $I(\cdot)$ le périodogramme obtenu à partir de l'échantillon Y_1, \dots, Y_T , posons ω_j les fréquences de Fourier, au sens où $\omega_j = 2\pi j/T$ pour $j < T/2$. Dans ce cas, quand $T \rightarrow \infty$, on a les résultats suivants;

- (i) $I(\omega_j) \sim f(\omega_j) \cdot \chi^2(2)/2$
- (ii) $I(\omega_j)$ et $I(\omega_k)$ sont indépendants pour $j \neq k$.

Remarque 45 La méthode de calcul des $I(\omega_1), \dots, I(\omega_m)$ présentée dans cette partie nécessite de l'ordre de T^2 opérations. La Fast Fourier Transform permet d'augmenter les temps de calcul puisque "seulement" $T \log_2 T$ opérations sont nécessaires (le gain en temps est alors en $T/\log_2 T$: pour 100 observations, les calculs sont alors 15 fois plus rapides).

Remarque 46 Considérons la série (X_t) définie par $X_t = 5 \cos(t/36) + 7 \sin(t/12) + \varepsilon_t$ où $\varepsilon_t \sim N(0, 1)$, représentée ci-dessous à gauche. Sa densité spectrale est représentée ci-dessous



On note deux maximums locaux, aux périodes 110 et 35 (correspondants aux paramètres $1/36$ et $1/12$).

” Pour prévoir l’avenir, il faut connaître le passé, car les événements de ce monde ont en tout temps des liens aux temps qui les ont précédés.”

Machiavel, Le Prince.

7 Estimation des modèles ARIMA : Box-Jenkins

L’ouvrage de Box et Jenkins *’Time series analysis, forecasting and control’*, publié en 1970 a proposé une démarche de prévision pour les séries univariées, fondée sur l’utilisation de processus ARIMA.

Les étapes pour l’estimation des coefficients d’un processus ARIMA sont les suivantes

(1) **identification**

(i) choix de d : combien de fois faut-il différencier pour obtenir une série stationnaire (*autocorrélogrammes, tests statistiques...*)

(ii) choix de p et q : ordres respectifs des composantes AR et MA

(2) **estimation des paramètres**

estimation des ϕ_i et des θ_j : paramètres respectifs des composantes AR et MA

(3) **vérification a posteriori**

(i) significativité des paramètres

(ii) validation de l’hypothèse de bruit blanc des résidus

Remarque 47 Il convient de retenir, comme en économétrie, le modèle le plus parcimonieux, utilisant le moins de paramètres, et ayant le meilleur comportement en prévision.

7.1 Estimation du paramètre d’intégration d

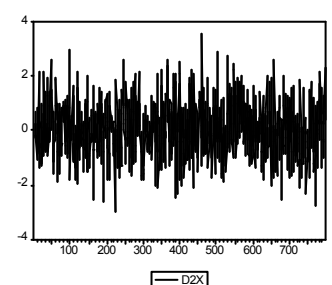
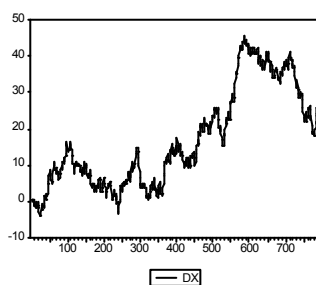
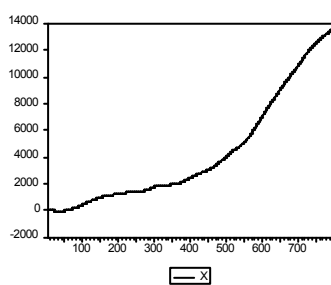
7.1.1 Approche empirique par l’autocorrélogramme

Comme nous l’avons vu dans la partie (2.4.3), les moments empiriques convergent, avec en plus normalité asymptotique (*sous certaines conditions*).

En pratique, si $\hat{\rho}(h)$ est proche de 1 (*pour un grand nombre de retards*), on a une racine unité, et le processus n’est pas stationnaire. On peut également penser à différencier si les premiers $\rho(h)$ sont proches les uns des autres, même si $\rho(1)$ semble assez différent de 1.

Il est à noter que pour des séries économiques, il est assez rare d’avoir $d \geq 3$.

Exemple 75 Les graphiques ci-dessous représentent les séries (en haut) et les autocorrélogrammes (en bas) de X_t , de ΔX_t et de $\Delta^2 X_t$



Sample: 1 798
Included observations: 798

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.996	0.996	794.68	0.000	
2	0.992	-0.003	1594.0	0.000	
3	0.988	-0.003	2388.0	0.000	
4	0.984	-0.003	3182.0	0.000	
5	0.980	-0.003	3976.0	0.000	
6	0.976	-0.003	4682.0	0.000	
7	0.972	-0.003	5451.0	0.000	
8	0.968	-0.003	6200.3	0.000	
9	0.964	-0.003	6960.3	0.000	
10	0.960	-0.003	7707.0	0.000	
11	0.956	-0.003	8440.3	0.000	
12	0.952	-0.003	9184.3	0.000	
13	0.948	-0.003	9914.9	0.000	
14	0.944	-0.003	10640.0	0.000	
15	0.940	-0.004	11360.0	0.000	
16	0.936	-0.004	12074.0	0.000	
17	0.931	-0.004	12784.0	0.000	
18	0.927	-0.004	13487.0	0.000	
19	0.923	-0.004	14186.0	0.000	
20	0.919	-0.004	14879.0	0.000	
21	0.915	-0.004	15566.0	0.000	
22	0.911	-0.004	16248.0	0.000	

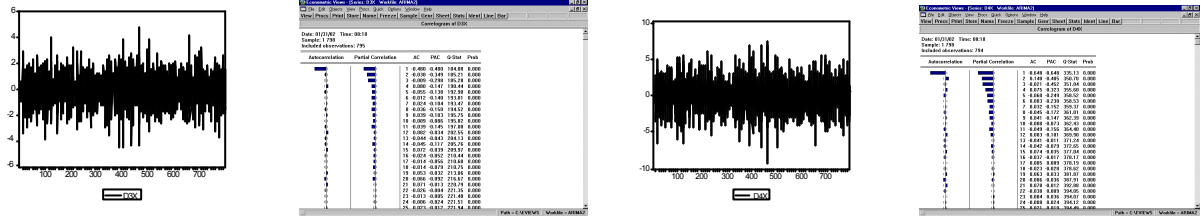
Sample: 1 798
Included observations: 797

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.996	0.996	793.19	0.000	
2	0.991	-0.005	1580.5	0.000	
3	0.987	-0.004	2362.4	0.000	
4	0.983	-0.005	3138.8	0.000	
5	0.979	-0.007	3909.0	0.000	
6	0.974	-0.006	4673.1	0.000	
7	0.970	-0.008	5431.1	0.000	
8	0.965	-0.019	6183.2	0.000	
9	0.961	-0.001	6929.6	0.000	
10	0.956	-0.027	7669.9	0.000	
11	0.952	-0.034	8403.6	0.000	
12	0.947	-0.018	9131.2	0.000	
13	0.942	-0.060	9851.9	0.000	
14	0.937	-0.020	10566.0	0.000	
15	0.932	-0.023	11274.0	0.000	
16	0.927	-0.044	11975.0	0.000	
17	0.922	-0.019	12668.0	0.000	
18	0.917	-0.012	13357.0	0.000	
19	0.913	-0.015	14039.0	0.000	
20	0.907	-0.035	14714.0	0.000	
21	0.902	-0.009	15391.0	0.000	
22	0.896	-0.063	16042.0	0.000	

Sample: 1 798
Included observations: 796

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.022	0.022	0.3911	0.532	
2	-0.017	-0.017	0.6100	0.734	
3	0.018	0.019	0.8813	0.830	
4	0.072	0.071	1.0630	0.281	
5	-0.030	-0.033	1.0910	0.276	
6	-0.024	-0.021	1.2742	0.393	
7	0.002	-0.000	1.2785	0.508	
8	-0.016	-0.021	1.4692	0.593	
9	0.035	0.042	1.5814	0.585	
10	0.012	0.012	1.6147	0.566	
11	-0.003	-0.003	1.6225	0.747	
12	0.073	0.075	1.9880	0.447	
13	-0.018	-0.029	12.245	0.508	
14	0.023	0.020	12.664	0.553	
15	0.060	0.062	15.632	0.407	
16	0.002	-0.012	15.835	0.479	
17	-0.010	-0.003	15.713	0.544	
18	0.007	0.010	15.752	0.610	
19	0.050	0.038	17.836	0.533	
20	-0.010	-0.005	17.912	0.593	
21	0.061	0.060	20.920	0.464	
22	-0.008	-0.017	20.979	0.522	

On peut d'ailleurs noter que si l'on continue à différencier, on a toujours des séries stationnaires



7.1.2 Tests de racine unité

La présentation sera ici inspirée de celle de Hamilton 'Time Series Analysis' (1994).

Le test de Dickey & Fuller simple Ce test permet de tester l'hypothèse H_0 : le processus suit une marche aléatoire contre l'hypothèse alternative H_a : le processus suit un modèle $AR(1)$. Ces tests peuvent être regroupés en 4 cas :

- (1) $Y_t = \rho Y_{t-1} + \varepsilon_t$: on teste H_0 : $\rho = 1$ (marche aléatoire sans dérive)
- (2) $Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t$: on teste H_0 : $\alpha = 0$ et $\rho = 1$ (marche aléatoire sans dérive)
- (3) $Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t$: on teste H_0 : $\alpha \neq 0$ et $\rho = 1$ (marche aléatoire avec dérive)
- (4) $Y_t = \alpha + \beta t + \rho Y_{t-1} + \varepsilon_t$: on teste H_0 : $\alpha = 0, \beta = 0$ et $\rho = 1$ (marche aléatoire sans dérive)

Le test de Dickey & Fuller, dans le cas (1), se construit comme un test de Student de l'hypothèse $\rho = 1$, ou plutôt $\rho - 1 = 0$. Etant donné l'estimateur naturel de ρ , on peut noter que

$$\hat{\rho} - 1 = \frac{\sum \varepsilon_t Y_{t-1}}{\sum Y_{t-1}^2}$$

Le test de Dickey & Fuller augmenté Ce test permet de tester l'hypothèse H_0 : est intégré d'ordre au moins 1 H_a : le processus suit un modèle $AR(p)$. Ces tests peuvent être regroupés en 4 cas :

- (1) $\Phi(L) Y_t = \varepsilon_t$: on teste H_0 : $\Phi(1) = 0$
- (2) $\Phi(L) Y_t = \alpha + \varepsilon_t$: on teste H_0 : $\alpha = 0$ et $\Phi(1) = 0$
- (3) $\Phi(L) Y_t = \alpha + \varepsilon_t$: on teste H_0 : $\alpha \neq 0$ et $\Phi(1) = 0$
- (4) $\Phi(L) Y_t = \alpha + \beta t + \varepsilon_t$: on teste H_0 : $\alpha = 0, \beta = 0$ et $\Phi(1) = 0$

Ces 4 cas peuvent être réécrits en introduisant les notations suivantes,

$$\Phi(L) = \Phi(1) + (1-L)\Phi^*(L) = \Phi(1) - \left[\sum_{i=0}^{p-1} \alpha_i L^i \right] (1-L) \text{ où } \begin{cases} \alpha_0 = \Phi(1) - 1 \\ \alpha_i = \alpha_{i-1} - \phi_i = \phi_{i+1} + \dots + \phi_p \end{cases}$$

pour $i = 1, \dots, p$. En posant $\rho = 1 - \Phi(1)$, on peut réécrire les 4 cas en

- (1) $Y_t = \rho Y_{t-1} + \sum \alpha_i \Delta y_{t-i} + \varepsilon_t$: on teste H_0 : $\rho = 1$
- (2) $Y_t = \alpha + \rho Y_{t-1} + \sum \alpha_i \Delta y_{t-i} + \varepsilon_t$: on teste H_0 : $\alpha = 0$ et $\rho = 1$
- (3) $Y_t = \alpha + \rho Y_{t-1} + \sum \alpha_i \Delta y_{t-i} + \varepsilon_t$: on teste H_0 : $\alpha \neq 0$ et $\rho = 1$
- (4) $Y_t = \alpha + \beta t + \rho Y_{t-1} + \sum \alpha_i \Delta y_{t-i} + \varepsilon_t$: on teste H_0 : $\alpha = 0, \beta = 0$ et $\rho = 1$

Les statistiques de tests et leurs lois Pour simplifier, on écrira

- (1) $\Delta Y_t = \phi Y_{t-1} + \sum \alpha_i \Delta y_{t-i} + \varepsilon_t$, avec $\phi = \rho - 1$ appelé 'Modèle [1]'
- (2 - 3) $\Delta Y_t = \alpha + \phi Y_{t-1} + \sum \alpha_i \Delta y_{t-i} + \varepsilon_t$ appelé 'Modèle [2]'
- (4) $\Delta Y_t = \alpha + \beta t + \phi Y_{t-1} + \sum \alpha_i \Delta y_{t-i} + \varepsilon_t$ appelé 'Modèle [3]'

Les tables ci-après, ont été tabulées par Dickey & Fuller (1979), et sont analogues aux tables du t de Student. Dans le cas 'simple', le paramètre ρ (ou ϕ) est estimé par la méthode des moindres carrés ordinaires. L'estimation des coefficients et des écart-types du modèle fournit un t_ϕ , analogue à la statistique de Student dans les modèles linéaires (rapport du coefficient sur son écart-type). Si t_ϕ est supérieur au t tabulé, on accepte H_0 , hypothèse d'existence d'une racine unité, et le processus n'est alors pas stationnaire.

Il est aussi possible d'effectuer ce test en utilisant $n\hat{\phi}_n$, où $\hat{\phi}_n$ est l'estimateur de ϕ obtenu à partir de n observations. Si cette valeur (empirique) est supérieure à celle tabulée (et donnée dans la deuxième table), on accepte l'hypothèse H_0 .

7. TABLES DE DICKEY-FULLER ¹
 Modèle [1] sans tendance et sans terme constant
 Modèle [2] sans tendance et avec terme constant
 Modèle [3] avec tendance et avec terme constant

Tables de la distribution du t_{β_1}

Nombre observations n	Probabilités							
	0,01	0,025	0,05	0,10	0,90	0,95	0,975	0,99
25	-2,66	-2,26	-1,95	-1,60	0,92	1,33	1,70	2,16
50	-2,62	-2,25	-1,95	-1,61	0,91	1,31	1,66	2,08
100	-2,60	-2,4	-1,95	-1,61	0,91	1,29	1,64	2,03
250	-2,58	-2,23	-1,95	-1,62	0,89	1,29	1,63	2,01
500	-2,58	-2,23	-1,95	-1,62	0,89	1,28	1,62	2,00
∞	-2,58	-2,23	-1,95	-1,62	0,89	1,28	1,62	2,00
25	-3,75	-3,33	-3,00	-2,63	-0,37	0,00	0,34	0,72
50	-3,58	-3,22	-2,93	-2,60	-0,40	-0,03	0,29	0,66
100	-3,51	-3,17	-2,89	-2,58	-0,42	-0,05	0,26	0,63
250	-3,46	-3,14	-2,88	-2,57	-0,42	-0,06	0,24	0,62
500	-3,44	-3,13	-2,87	-2,57	-0,43	-0,07	0,24	0,61
∞	-3,43	-3,12	-2,86	-2,57	-0,44	-0,07	0,23	0,60
25	-4,38	-3,95	-3,60	-3,24	-1,14	-0,80	-0,50	-0,15
50	-4,15	-3,80	-3,50	-3,18	-1,19	-0,87	-0,58	-0,24
100	-4,04	-3,73	-3,45	-3,15	-1,22	-0,90	-0,62	-0,28
250	-3,99	-3,69	-3,43	-3,13	-1,23	-0,92	-0,64	-0,31
500	-3,98	-3,68	-3,42	-3,13	-1,24	-0,93	-0,65	-0,32
∞	-3,96	-3,66	-3,41	-3,12	-1,25	-0,94	-0,66	-0,33

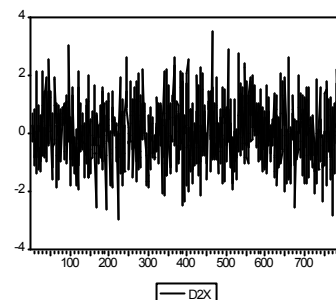
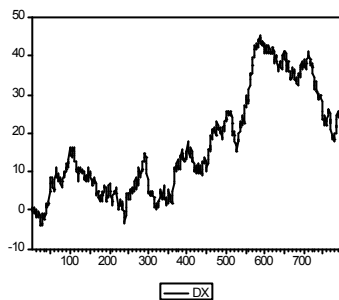
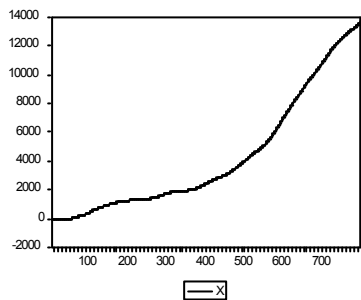
Tables de la distribution du $n(\phi_1 - 1)$

Nombre observations n	Probabilités							
	0,01	0,025	0,05	0,10	0,90	0,95	0,975	0,99
25	-11,9	-9,3	-7,3	-5,3	1,01	1,40	1,79	2,28
50	-12,9	-9,9	-7,7	-5,5	0,97	1,35	1,70	2,16
100	-13,3	-10,2	-7,9	-5,6	0,95	1,31	1,65	2,09
250	-13,6	-10,3	-8,0	-5,7	0,93	1,28	1,62	2,04
500	-13,7	-10,4	-8,0	-5,7	0,93	1,28	1,61	2,04
∞	-13,8	-10,5	-8,1	-5,7	0,93	1,28	1,60	2,03
25	-17,2	-14,6	-12,5	-10,2	-0,76	0,01	0,65	1,40
50	-18,9	-15,7	-13,3	-10,7	-0,81	-0,07	0,53	1,22
100	-19,8	-16,3	-13,7	-11,0	-0,83	-0,10	0,47	1,14
250	-20,3	-16,6	-14,0	-11,2	-0,84	-0,12	0,43	1,09
500	-20,5	-16,8	-14,0	-11,2	-0,84	-0,13	0,42	1,06
∞	-20,7	-16,9	-14,1	-11,3	-0,85	-0,13	0,41	1,04
25	-22,5	-19,9	-17,9	-15,6	-3,66	-2,51	-1,53	-0,43
50	-25,7	-22,4	-19,8	-16,8	-3,71	-2,60	-1,66	-0,65
100	-27,4	-23,6	-20,7	-17,5	-3,74	-2,62	-1,73	-0,75
250	-28,4	-24,4	-21,3	-18,0	-3,75	-2,64	-1,78	-0,82
500	-28,9	-24,8	-21,5	-18,1	-3,76	-2,65	-1,78	-0,84
∞	-29,5	-25,1	-21,8	-18,3	-3,77	-2,66	-1,79	-0,87

1. Source : Fuller W.A. *Introduction to Statistical Times Series*, John Wiley, 1976.

Mise en place pratique des tests On choisit tout d'abord un p suffisamment grand pour que $\Phi(L)X_t$ suive à peu près un bruit blanc. On choisit alors parmi les cas proposés suivant que le graphique de la série présente, ou pas, une tendance linéaire.

Exemple 76 *Considérons la série décrit précédemment,*



Le test (simple) de Dickey & Fuller revient à estimer les 3 modèles suivants,

$$\begin{cases} X_t - X_{t-1} = \phi X_{t-1} \\ X_t - X_{t-1} = \alpha + \phi X_{t-1} \\ X_t - X_{t-1} = \alpha + \beta t + \phi X_{t-1} \end{cases}$$

et dans le cas du test augmenté, avec $p = 2$

$$\begin{cases} X_t - X_{t-1} = \rho X_{t-1} - [\phi_2 \Delta X_{t-1} + \phi_3 \Delta X_{t-2}] \\ X_t - X_{t-1} = \alpha + \rho X_{t-1} - [\phi_2 \Delta X_{t-1} + \phi_3 \Delta X_{t-2}] \\ X_t - X_{t-1} = \alpha + \beta t + \rho X_{t-1} - [\phi_2 \Delta X_{t-1} + \phi_3 \Delta X_{t-2}] \end{cases}$$

Le troisième modèle s'écrit, compte tenu des sorties obtenues ci-dessous,

$$\Delta X_t = -0.048502 + 0.00919 t - 0.000083 X_{t-1} - \begin{bmatrix} 1.01516 \Delta X_{t-1} - 0.022332 \Delta X_{t-2} \\ (0.092874) \quad (0.000466) \quad (0.0000244) \quad (0.035561) \quad (0.035629) \end{bmatrix}$$

avec $n = 794$. Les valeurs du test de Dickey & Fuller sont données par

ADF Test Statistic	-1.555831	1% Critical Value*	-3.9744
		5% Critical Value	-3.4177
		10% Critical Value	-3.1405

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation
LS // Dependent Variable is D(X)
Sample: 4 798
Included observations: 795 after adjusting endpoints

Variable	Coefficient	Std. Error	T-Statistic	Prob.
X(-1)	-3.80E-05	2.44E-05	-1.555831	0.1201
D(X(-1))	1.015160	0.035561	28.54703	0.0000
D(X(-2))	-0.022332	0.035629	-0.626804	0.5310
C	-0.048502	0.092874	-0.522231	0.6017
Trend	0.000919	0.000466	1.971959	0.0490
R-squared	0.994281	Mean dependent var	17.17545	
Adjusted R-squared	0.994252	S. D. dependent var	13.56918	
S.E. of regression	1.028792	Akaike info criterion	0.063041	
Sum squared resid	836.1470	Schwartz criterion	0.092464	
Log likelihood	-1148.115	F-statistic	34333.72	
Durbin-Watson stat	1.998565	Prob(F-statistic)	0.000000	

En répétant ce test en changeant la forme du modèle (ici sans trend $\alpha + \beta t$, et en changeant l'ordre p), on confirme ce rejet de H_0 : la série X_t possède une racine unitaire et n'est pas stationnaire : la statistique de test ADF Test

Statistic est toujours supérieure aux valeurs critiques :

ADF Test Statistic 0.083143 1% Critical Value* -2.5683
 5% Critical Value -1.9398
 10% Critical Value -1.6158

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation
 LS // Dependent Variable is D(X)
 Sample: 3 798
 Included observations: 796 after adjusting endpoints

Variable	Coefficient	Std. Error	T-Statistic	Prob.
X(-1)	1.19E-06	1.44E-05	0.083143	0.9338
D(X(-1))	0.999599	0.003892	256.8238	0.0000
R-squared	0.994239	Mean dependent var	17.15451	
Adjusted R-squared	0.994231	S.D. dependent var	13.57350	
S.E. of regression	1.030928	Akaike info criterion	0.063428	
Sum squared resid	843.8728	Schwartz criterion	0.075186	
Log likelihood	-1152.719	F-statistic	137020.3	
Durbin-Watson stat	1.952055	Prob(F-statistic)	0.000000	

ADF Test Statistic 0.147201 1% Critical Value* -2.5683
 5% Critical Value -1.9398
 10% Critical Value -1.6158

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation
 LS // Dependent Variable is D(X)
 Sample: 4 798
 Included observations: 795 after adjusting endpoints

Variable	Coefficient	Std. Error	T-Statistic	Prob.
X(-1)	2.12E-06	1.44E-05	0.147201	0.8830
D(X(-1))	1.023187	0.035517	28.80845	0.0000
D(X(-2))	-0.023865	0.035721	-0.668114	0.5043
R-squared	0.994234	Mean dependent var	17.17545	
Adjusted R-squared	0.994220	S.D. dependent var	13.56918	
S.E. of regression	1.031640	Akaike info criterion	0.066067	
Sum squared resid	842.9113	Schwartz criterion	0.083721	
Log likelihood	-1151.318	F-statistic	68285.65	
Durbin-Watson stat	1.998543	Prob(F-statistic)	0.000000	

En faisant le test sur la série différenciée une fois (ΔX_t), on observe là aussi que l'ADF Test Statistic est toujours supérieure aux valeurs critiques : H_0 est encore acceptée, et donc la série ΔX_t possède elle aussi une racine unitaire et n'est donc pas stationnaire

ADF Test Statistic -1.301307 1% Critical Value* -3.4412
 5% Critical Value -2.8656
 10% Critical Value -2.5689

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation
 LS // Dependent Variable is D(DX)
 Sample: 5 798
 Included observations: 794 after adjusting endpoints

Variable	Coefficient	Std. Error	T-Statistic	Prob.
DX(-1)	-0.003514	0.002700	-1.301307	0.1935
D(DX(-1))	0.024104	0.035551	0.678025	0.4980
D(DX(-2))	-0.016162	0.035648	-0.453368	0.6504
C	0.094442	0.059015	1.600316	0.1099
R-squared	0.002941	Mean dependent var	0.034441	
Adjusted R-squared	-0.000845	S.D. dependent var	1.030669	
S.E. of regression	1.031104	Akaike info criterion	0.066286	
Sum squared resid	839.9090	Schwartz criterion	0.089848	
Log likelihood	-1148.953	F-statistic	0.776723	
Durbin-Watson stat	1.998939	Prob(F-statistic)	0.507138	

Le test de Dickey & Fuller appliqué cette fois-ci à $\Delta^2 X_t$ donne les résultats suivants,

ADF Test Statistic -15.92501 1% Critical Value* -3.4412
 5% Critical Value -2.8656
 10% Critical Value -2.5689

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation
 LS // Dependent Variable is D(D2X)
 Sample: 6 798
 Included observations: 793 after adjusting endpoints

Variable	Coefficient	Std. Error	T-Statistic	Prob.
D2X(-1)	-0.975906	0.061281	-15.92501	0.0000
D(D2X(-1))	-0.000917	0.049987	-0.018335	0.9854
D(D2X(-2))	-0.018974	0.035701	-0.531468	0.5952
C	0.033307	0.036722	0.907005	0.3647
R-squared	0.489083	Mean dependent var	0.000169	
Adjusted R-squared	0.487141	S.D. dependent var	1.441939	
S.E. of regression	1.032633	Akaike info criterion	0.069255	
Sum squared resid	841.3347	Schwartz criterion	0.092840	
Log likelihood	-1148.678	F-statistic	251.7610	
Durbin-Watson stat	2.002323	Prob(F-statistic)	0.000000	

Cette fois-ci, le test de Dickey & Fuller permet de rejeter H_0 : $\Delta^2 X_t$ n'a pas de racine unitaire, et la série $\Delta^2 X_t$ est donc stationnaire. Ce test valide les résultats graphiques de l'exemple (75) : la série X_t est intégrée d'ordre 2 : $d = 2$.

Remarque 48 Dans le cas de séries financières (par exemple), il convient de faire attention lors de la lecture des résultats des tests de Dickey & Fuller : les processus à mémoire longue, bien que stationnaires, semblent avoir une racine unité. Avant de différencier afin d'obtenir une série stationnaire, il peut être intéressant de tester l'hypothèse de mémoire longue du processus.

A retenir 11 Dans les tests de Dickey Fuller augmenté, trois (ou quatre) alternatives sont proposées : avec ou sans tendance et constante. Il vaut mieux choisir l'alternative permettant de mieux décrire la série : si la série (X_t) n'est pas centrée, et que l'on tente un test de Dickey Fuller sans constante, il est possible qu'il est possible H_0 soit rejetée, non pas parce qu'il n'y a pas de racine unité, mais parce que le modèle testé est mal spécifié (cf exercice 16 de l'examen 2002/2003).

Compléments sur les tests de racine unité Considérons une écriture de la forme

$$\Phi(L)X_t = \Theta(L)\varepsilon_t, \text{ où } (\varepsilon_t) \text{ est un bruit blanc.}$$

L'hypothèse à tester est (H_0) : il existe θ tel que $\Phi(e^{i\theta}) = 0$, c'est à dire qu'une racine est sur le disque unité (*racine unité*) le reste des racines étant à l'extérieur du disque unité : $\Phi(L) = (1-L)\Phi^*(L)$ où $\Phi^*(1) \neq 0$, avec les racines de Φ^* à l'extérieur du disque unité. Alors

$$\Delta X_t = \Phi^*(L)^{-1}\Theta(L)\varepsilon_t = \Psi^*(L)\varepsilon_t = \eta_t \text{ ou } X_t = X_{t-1} + \eta_t.$$

L'hypothèse alternative (H_1) s'écrit alors $\Phi(e^{i\theta}) \neq 0$ pour tout θ : Φ n'a pas de racine unité, et on suppose de plus que toutes les racines sont à l'extérieur du disque unité :

$$X_t = \Phi(L)^{-1}\Theta(L)\varepsilon_t = \Psi(L)\varepsilon_t = \eta_t.$$

Les tests de Dickey-Fuller permet de tester cette hypothèse : le test de régression s'écrit alors

$$X_t = \phi X_{t-1} + \eta_t \text{ dont l'estimation est } X_t = \hat{\phi} X_{t-1} + \hat{\eta}_t.$$

Il est alors possible de montrer que sous l'hypothèse (H_0) : $\phi = 1$, la statistique de test s'écrit

$$\hat{t}_{\phi=1} = \frac{\hat{\phi} - 1}{\hat{\sigma}(\hat{\phi})} \text{ où } \hat{\phi} - 1 = \frac{\sum X_{t-1}\eta_t}{\sum X_{t-1}^2}, \quad s^2 = \frac{1}{T-1} \sum [X_t - \hat{\phi}X_{t-1}]^2 \text{ et } \hat{\sigma}(\hat{\phi})^2 = \frac{s^2}{\sum X_{t-1}^2},$$

avec $\hat{\sigma}(\hat{\phi})$ écart type (*par moindres carrés*) de l'estimateur de ϕ , et sa distribution est donnée par

$$\hat{t}_{\phi=1} = \frac{\hat{\phi} - 1}{\hat{\sigma}(\hat{\phi})} \xrightarrow{\mathcal{L}} \frac{\int_0^1 W_t dW_t}{\left[\int_0^1 W_t^2 dt\right]^{1/2}} \neq \mathcal{N}(0, 1) \text{ où } (W_t) \text{ est un brownien standard sur } [0, 1].$$

Cette distribution n'est pas gaussienne, et des tabulations (*obtenues par des méthodes de type Monte-Carlo*) sont nécessaire pour tabuler la distribution limite²⁰.

Tests de Phillips et Perron Ces tests non paramétriques ont été introduits en 1988. La distribution théorique à la base des tests de Dickey & Fuller repose sur l'hypothèse d'hétéroscédasticité du bruit. La généralisation des tests *DF* aux tests *ADF* se fait en considérant

$$Y_t = D_t + \rho Y_{t-1} + \varepsilon_t \quad \rightarrow \quad Y_t = D_t + \rho Y_{t-1} + \sum \alpha_i \Delta y_{t-i} + \varepsilon_t,$$

où (D_t) est une tendance déterministe. La généralisation des tests *DF* proposée par Phillips et Perron consiste à ne plus supposer que (ε_t) est un bruit blanc, et à autoriser que ce processus soit autocorrélé. La généralisation de ces tests au cas hétéroscédastique a été proposée par Phillips et Perron, les valeurs critiques correspondant à celles des tests *ADF*. Ces tests reposent sur des résultats de la théorie de la convergence faible fonctionnelle (*théorème central limite fonctionnel (FCLT) par exemple*). L'utilisation du FCLT pour des tests de racines unité a été proposé dès 1958 par White.

²⁰Le lien entre les processus intégrés et le mouvement brownien est donné page 25.

Si (X_t) est un processus stationnaire alors les statistiques calculées sur ce processus vérifieront le FCLT. Considérons par exemple le cas $AR(1)$, $X_t = \rho X_{t-1} + \varepsilon_t$ pour $t = 1, \dots, T$, et cherchons à tester $\rho = 1$ (hypothèse H_0). En supposons H_0 vérifiée, et considérons la somme partielle du processus d'innovation,

$$S_t = X_t - X_0 = \sum_{i=1}^t \varepsilon_i.$$

On prendra comme valeur initiale de (S_t) , $S_0 = 0$, mais pour le choix de X_0 trois possibilités sont généralement envisagées : (i) $X_0 = c$ (constante), (ii) X_0 admet une distribution spécifiée a priori, (iii) $X_0 = X_T$. Cette dernière condition, dite hypothèse de circularité, a été proposé par Hotelling. Phillips avait suggéré la seconde possibilité.

En notant $X_T(r) = S_{[Tr]}/\sqrt{T}$, il est possible de montrer (cf partie précédente) que $X_T(r)$ converge faiblement (noté \Rightarrow) vers un mouvement brownien (cf. Billingsley (1968)).

Propriété 46 Si (ε_t) vérifie l'hypothèse (i) et si $\sup \{|\varepsilon_t|^{\beta+\eta}\} < \infty$ pour $\eta > 0$ et $\beta > 0$ alors, quand $T \rightarrow \infty$, sous l'hypothèse $H_0 : \rho = 1$ dans le modèle $X_t = \rho X_{t-1} + \varepsilon_t$ on a les résultats suivants

$$(i) \frac{1}{T^2} \sum_{t=1}^T X_{t-1}^2 \Rightarrow \sigma^2 \int_0^1 W_s^2 ds$$

$$(ii) \frac{1}{T} \sum_{t=1}^T X_{t-1} (X_t - X_{t-1}) \Rightarrow \frac{\sigma^2}{2} \left[W_1^2 - \frac{\sigma_\varepsilon^2}{\sigma^2} \right]$$

$$(iii) T(\hat{\rho} - 1) \Rightarrow \frac{1}{2} \frac{W_1^2 - \sigma_\varepsilon^2/\sigma^2}{\int_0^1 W_s^2 ds}$$

$$(iv) \hat{\rho} \xrightarrow{\mathbb{P}} 1$$

$$(v) t_{\hat{\rho}} = \frac{\hat{\rho} - 1}{\hat{\sigma}} \sqrt{\sum_{t=1}^T (X_t - X_{t-1})^2} \Rightarrow \frac{\sigma}{2\sigma_\varepsilon} \frac{W_1^2 - \sigma_\varepsilon^2/\sigma^2}{\sqrt{\int_0^1 W_s^2 ds}}$$

Preuve. Phillips (1987), *Testing for a unit root in a time series regression.* ■

Le point (iv) montre que les moindres carrés ordinaires conservent la propriété de convergence quand il y a une racine unité.

Exemple 77 En reprenant la série de l'exemple (75), on retrouve que la série (X_t) admet une racine unité, que l'on teste un modèle simple, sans constante ni tendance (à gauche), ou avec tendance et constante (à droite),

Null Hypothesis: X has a unit root			
Exogenous: None			
Bandwidth: 11 (Newey-West using Bartlett kernel)			
	Adj. t-Stat	Prob.*	
Phillips-Perron test statistic	4.053815	1.0000	
Test critical values:		1% level	-2.576634
		5% level	-1.942431
		10% level	-1.615638
*MacKinnon (1996) one-sided p-values.			
Residual variance (no correction)	31.50308		
HAC corrected variance (Bartlett kernel)	354.1617		

Phillips-Perron Test Equation				
Dependent Variable: D(X)				
Method: Least Squares				
Date: 07/11/03 Time: 14:01				
Sample(adjusted): 2 200				
Included observations: 199 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
X(-1)	0.008077	0.000570	14.17192	0.0000
R-squared	-0.347519	Mean dependent var	6.330875	
Adjusted R-squared	-0.347519	S.D. dependent var	4.847334	
S.E. of regression	5.626917	Akaike info criterion	6.299013	
Sum squared resid	8269.114	Schwarz criterion	6.314562	
Log likelihood	-625.6523	Durbin-Watson stat	0.029997	

Null Hypothesis: X has a unit root			
Exogenous: Constant, Linear Trend			
Bandwidth: 11 (Newey-West using Bartlett kernel)			
	Adj. t-Stat	Prob.*	
Phillips-Perron test statistic	-2.876236	0.1725	
Test critical values:		1% level	-4.004836
		5% level	-3.432566
		10% level	-3.140059
*MacKinnon (1996) one-sided p-values.			
Residual variance (no correction)	16.21938		
HAC corrected variance (Bartlett kernel)	170.1988		

Phillips-Perron Test Equation				
Dependent Variable: D(X)				
Method: Least Squares				
Date: 07/11/03 Time: 14:02				
Sample(adjusted): 2 200				
Included observations: 199 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
X(-1)	-0.019037	0.002875	-6.622402	0.0000
C	-2.270652	0.990735	-2.292088	0.0230
@TREND(1)	0.184788	0.023594	7.861856	0.0000
R-squared	0.306229	Mean dependent var	6.330875	
Adjusted R-squared	0.299150	S.D. dependent var	4.847334	
S.E. of regression	4.058033	Akaike info criterion	5.654234	
Sum squared resid	3227.656	Schwarz criterion	5.703882	
Log likelihood	-559.5963	F-statistic	43.25701	
Durbin-Watson stat	0.056810	Prob(F-statistic)	0.000000	

avec les mêmes conclusions pour la série différenciée une fois,

Null Hypothesis: DX has a unit root
Exogenous: None
Bandwidth: 2 (NeweyWest using Bartlett kernel)

	Adj. t-Stat	Prob. *
Phillips-Perron test statistic	-0.638485	0.4397
Test critical values:		
1% level	-2.576693	
5% level	-1.942439	
10% level	-1.615633	

*MacKinnon (1996) one-sided p-values.

Residual variance (no correction)	0.937906
HAC corrected variance (Bartlett kernel)	0.925710

Phillips-Perron Test Equation
Dependent Variable: D(DX)
Method: Least Squares
Date: 07/11/03 Time: 14:04
Sample(adjusted): 3200
Included observations: 198 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DX(-1)	-0.005587	0.008655	-0.645434	0.5194
R-squared	0.000805	Mean dependent var	0.035032	
Adjusted R-squared	0.000805	S.D. dependent var	0.971301	
S.E. of regression	0.970910	Akaike info criterion	2.783872	
Sum squared resid	185.7053	Schwarz criterion	2.800479	
Log likelihood	-274.6033	Durbin-Watson stat	2.089261	

Null Hypothesis: DX has a unit root
Exogenous: Constant, Linear Trend
Bandwidth: 2 (Newey-West using Bartlett kernel)

	Adj. t-Stat	Prob. *
Phillips-Perron test statistic	-1.543621	0.8112
Test critical values:		
1% level	-4.005076	
5% level	-3.432682	
10% level	-3.140127	

*MacKinnon (1996) one-sided p-values.

Residual variance (no correction)	0.924445
HAC corrected variance (Bartlett kernel)	0.920735

Phillips-Perron Test Equation
Dependent Variable: D(DX)
Method: Least Squares
Date: 07/11/03 Time: 14:05
Sample(adjusted): 3200
Included observations: 198 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DX(-1)	-0.023867	0.015432	-1.546605	0.1236
C	0.200901	0.146759	1.368915	0.1726
@TREND(1)	-0.000147	0.001309	-0.112696	0.9104
R-squared	0.015146	Mean dependent var	0.035032	
Adjusted R-squared	0.005045	S.D. dependent var	0.971301	
S.E. of regression	0.968848	Akaike info criterion	2.789618	
Sum squared resid	183.0401	Schwarz criterion	2.839440	
Log likelihood	-273.1722	F-statistic	1.499425	
Durbin-Watson stat	2.081225	Prob(F-statistic)	0.228821	

En revanche, dans le cas de la série différenciée deux fois, tous les tests valident l'hypothèse d'absence de racine unité

Null Hypothesis: D2X has a unit root
Exogenous: None
Bandwidth: 2 (NeweyWest using Bartlett kernel)

	Adj. t-Stat	Prob. *
Phillips-Perron test statistic	-14.71894	0.0000
Test critical values:		
1% level	-2.576753	
5% level	-1.942448	
10% level	-1.615628	

*MacKinnon (1996) one-sided p-values.

Residual variance (no correction)	0.939841
HAC corrected variance (Bartlett kernel)	0.984773

Phillips-Perron Test Equation
Dependent Variable: D(D2X)
Method: Least Squares
Date: 07/11/03 Time: 14:08
Sample(adjusted): 4200
Included observations: 197 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D2X(-1)	-1.049877	0.071265	-14.73193	0.0000
R-squared	0.525457	Mean dependent var	-0.001932	
Adjusted R-squared	0.525457	S.D. dependent var	1.410894	
S.E. of regression	0.971924	Akaike info criterion	2.785985	
Sum squared resid	185.1487	Schwarz criterion	2.802651	
Log likelihood	-273.4195	Durbin-Watson stat	1.990409	

Null Hypothesis: D2X has a unit root
Exogenous: Constant, Linear Trend
Bandwidth: 1 (Newey-West using Bartlett kernel)

	Adj. t-Stat	Prob. *
Phillips-Perron test statistic	-14.71024	0.0000
Test critical values:		
1% level	-4.005318	
5% level	-3.432799	
10% level	-3.140195	

*MacKinnon (1996) one-sided p-values.

Residual variance (no correction)	0.936095
HAC corrected variance (Bartlett kernel)	0.940320

Phillips-Perron Test Equation
Dependent Variable: D(D2X)
Method: Least Squares
Date: 07/11/03 Time: 14:08
Sample(adjusted): 4200
Included observations: 197 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D2X(-1)	-1.054154	0.071654	-14.71176	0.0000
C	0.124773	0.141908	0.879253	0.3804
@TREND(1)	-0.000904	0.001224	-0.738662	0.4610
R-squared	0.527348	Mean dependent var	-0.001932	
Adjusted R-squared	0.522476	S.D. dependent var	1.410894	
S.E. of regression	0.974972	Akaike info criterion	2.802296	
Sum squared resid	184.4107	Schwarz criterion	2.852294	
Log likelihood	-273.0261	F-statistic	108.2252	
Durbin-Watson stat	1.989237	Prob(F-statistic)	0.000000	

Remarques complémentaires

Un certains nombres d'études sur des données simulées ont montré que ces tests rejettent difficilement l'hypothèse H_0 dans le cas de séries désaisonnalisées. Il est alors parfois intéressant d'agréger des données mensuelles en données annuelles, et de tester si la série annuelle présente une racine unité. Nelson et Plosser (1992) ont montré que les racines unitaires caractérisent un grand nombre de séries macroéconomiques.

Bien d'autres tests sont disponibles sous EViews (*KPSS - Kwiatkowski, Philipps, Schmidt et Shin (1992) ou encore Ng-Perron (1993)*). Sous *RATS*, un grand nombre de procédures sont également disponibles sur <http://www.estima.com>, également des implémentations des tests de racine unité saisonnière).

Le test de **Schmidt-Philipps** repose sur l'idée que dans le cas du test *ADF* de type 4 - avec tendance linéaire - l'interprétation des paramètre n'est pas la même : considérons le modèle $Y_t = \alpha + \beta t + \rho Y_{t-1} + \varepsilon_t$ et l'hypothèse $H_0 : \beta = 0$ et $\rho = 1$. Sous H_0 et l'hypothèse alternative H_a , on a respectivement

$$H_0 : Y_t = Y_0 + \alpha t + \sum_{k=0}^t \varepsilon_{t-k} \text{ et } H_a : Y_t = \left[\beta \rho + \frac{\alpha}{1 - \rho} \right] + \beta(1 - \rho) + \sum_{k=0}^{\infty} \rho^k \varepsilon_{t-k}.$$

Autrement dit, sous H_a , (Y_t) est stationnaire autour d'une tendance déterministe dont la pente est $\beta(1 - \rho)$, alors que sous H_0 , (Y_t) est non stationnaire, avec pour tendance α . Aussi, Schmidt et Philipps ont proposé de modéliser (Y_t) sous la forme $Y_t = \alpha + \beta t + X_t$ où (X_t) est non stationnaire sous H_0 et (X_t) est stationnaire sous H_a . On a alors

$$\begin{cases} Y_t = \alpha + \beta t + X_t \\ X_t = \rho X_{t-1} + \varepsilon_t \end{cases} \text{ où } |\rho| \leq 1 \text{ et } (\varepsilon_t) \sim BB(0, \sigma^2) \text{ et on teste } \begin{cases} H_0 & \rho = 1 \\ H_a & \rho < 1 \end{cases}.$$

7.1.3 Tests de racines unitaires saisonnières

Dans le cas d'une modélisation *SARIMA*, avec une saisonnalité d'ordre s , il peut être intéressant de tester l'ordre s . Un certain nombre de tests ont été mis en oeuvre dans les années 80 et 90, en particulier pour tester de la saisonnalité à l'ordre 4 et à l'ordre 12.

Tests de Hasza et Fuller (1982) et de Osborn, Chui, Smith & Birchenhall (OCSB, 1988) Hasza et Fuller ont considéré le modèle

$$Y_t = \phi_1 Y_{t-1} + \phi_s Y_{t-s} + \phi_{s+1} Y_{t-s-1} + \varepsilon_t$$

où (ε_t) est un bruit blanc. L'hypothèse H_0 s'écrit ici $H_0 : \phi_1 = \phi_s = -\phi_{s+1} = 1$.

Osborn, Chui, Smith et Birchenhall ont alors étendu cette approche sous la forme

$$\Theta(L)(1-L)(1-L^s)Y_t = \sum_{i=1}^s \delta_s D_{s,t} + \lambda(1-L^s)Y_{t-1} + \mu(1-L)Y_{t-s} + \varepsilon_t$$

Si l'on accepte l'hypothèse $\mu = 0$, la différence à l'ordre s est appropriée, et si $\lambda = \mu = 0$, alors le filtre $(1-L)(1-L^s)$ est nécessaire.

Test de Hylleberg, Engle, Granger et Yoo (HEGY, 1990) Ce test utilise la décomposition des polynômes $(1-L^4)$ et $(1-L^{12})$, avec respectivement 4 et 12 racines unités : dans le cas d'une saisonnalité à l'ordre $s = 12$, on considère une écriture de la forme

$$\Theta(L)P_8(L)Y_t = \mu_t + \pi_1 P_1(L)Y_{t-1} + \pi_2 P_2(L)Y_{t-2} + \pi_3 P_3(L)Y_{t-1} + \pi_4 P_3(L)Y_{t-2} + \pi_5 P_4(L)Y_{t-1} + \pi_6 P_4(L)Y_{t-2} + \pi_7 P_5(L)Y_{t-1} + \pi_8 P_5(L)Y_{t-2} + \pi_9 P_6(L)Y_{t-1} + \pi_{10} P_6(L)Y_{t-2} + \pi_{11} P_7(L)Y_{t-1} + \pi_{12} P_7(L)Y_{t-2}$$

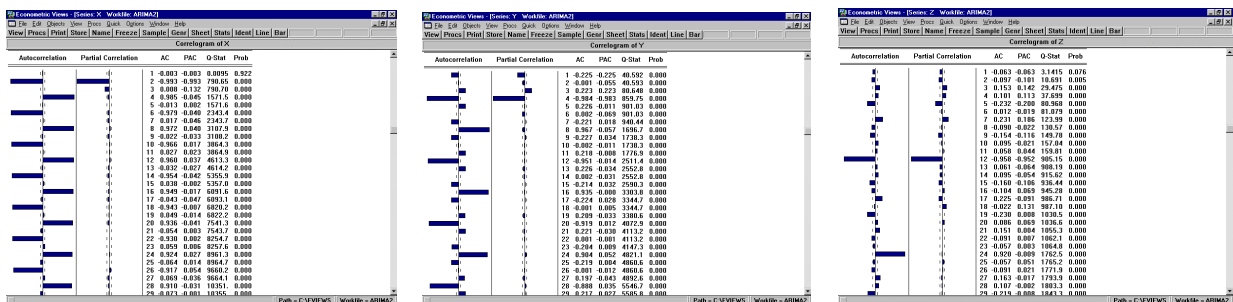
où les polynômes retards P_i sont définis par

$$\begin{cases} P_1(L) = -(1+L)(1+L^2)(1+L^4+L^8) & \text{et } P_2(L) = (1-L)(1+L^2)(1+L^4+L^8), \\ P_3(L) = (1-L^2)(1+L^4+L^8) & \text{et } P_4(L) = (1-L^4)(1-\sqrt{3}L+L^2)(1+L^2+L^4), \\ P_5(L) = (1-L^4)(1+\sqrt{3}L+L^2)(1+L^2+L^4) & \text{et } P_6(L) = (1-L^4)(1-L^2+L^4)(1-L+L^2), \\ P_7(L) = (1-L^4)(1-L^2+L^4)(1+L+L^2) & \text{et } P_8(L) = (1-L^{12}). \end{cases}$$

Les variables $Z_t^{(i)} = P_i(L)Y_t$ sont alors associées aux différentes racines du polynôme. On peut alors considérer les t de Student pour les variables π_1 et π_2 , ainsi que les F de Fisher associés aux couples (π_3, π_4) , (π_5, π_6) , (π_7, π_8) , (π_9, π_{10}) et (π_{11}, π_{12}) .

Test de Franses (1990) Ce test a été mis en place pour tester une saisonnalité à l'ordre 12.

Détection graphique d'une racine unitaire saisonnière Considérons les séries suivantes, (X_t) , (Y_t) et (Z_t) comportant respectivement une racine unitaire saisonnière d'ordre 2, 4 et 12,



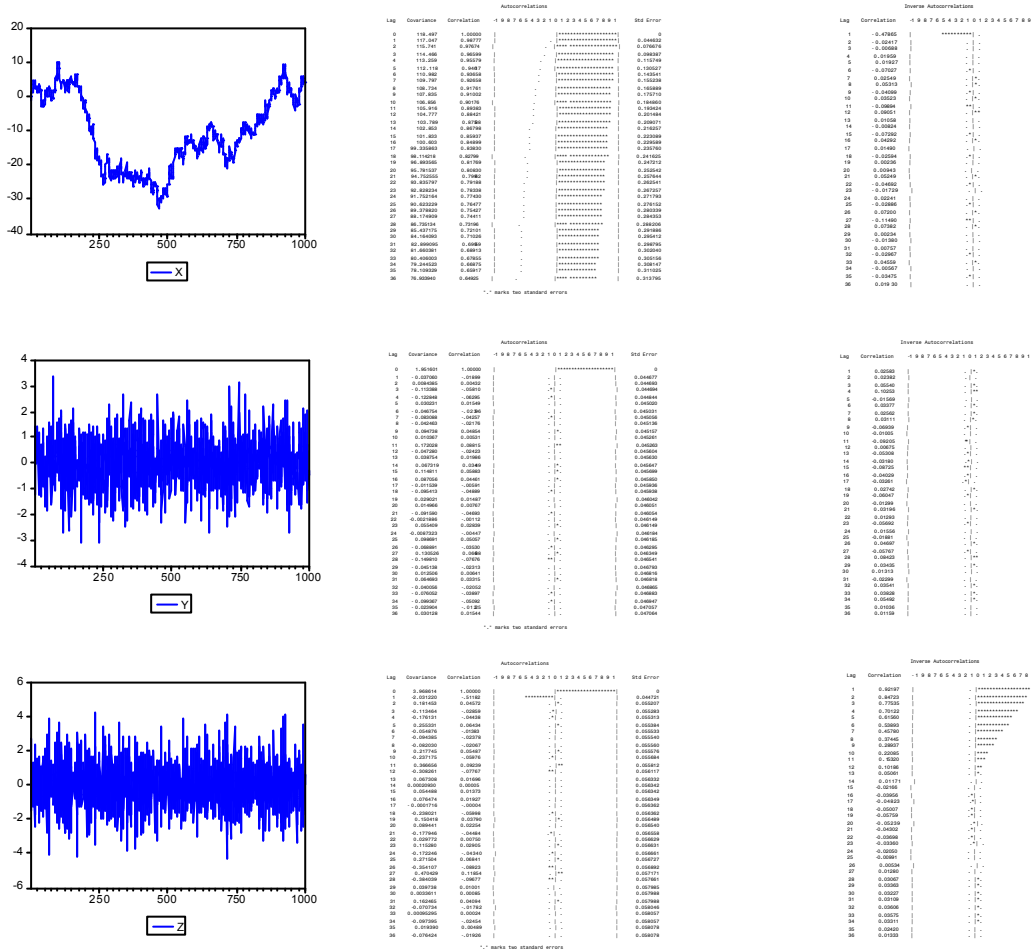
Sur ces trois graphiques, en considérant la série partielle des autocorrélogrammes $r^s(h) = |\rho(sh)|$, on obtient une série 'constante', proche de 1, de même que l'autocorrélogramme d'une série en présence de racine unitaire.

Toutefois, si ce genre de comportement laisse à penser qu'il y a une racine unitaire saisonnière, l'ordre s n'est pas nécessairement celui indiqué par l'autocorrélogramme : une série saisonnière d'ordre 4 peut avoir un autocorrélogramme proche de celui de gauche.

7.1.4 Complément sur la notion de sur-différentiation

Considérons la série suivante, correspondant à une marche aléatoire (X_t) . On notera alors $Y_t = (1-L)X_t$ et $Z_t = (1-L)Y_t$, autrement dit, on différencie respectivement une fois et deux fois la marche aléatoire. On représentera

respectivement les autocorrélations et les autocorrélations inverses, au centre et à droite,



Comme nous l'avons déjà évoqué, l'autocorrélogramme de la série (X_t) permet - a priori - de conclure à la présence d'une racine unité. Le comportement de l'autocorrélogramme inverse de la série (Z_t) présente, de façon moins nette certes, le même genre de comportement.

On peut noter également sur les autocorrélogrammes de (Y_t) , correspondant à un bruit blanc, que les autocorrélations et les autocorrélations inverses sont identiques (*ce qui est une caractérisation des bruits blancs*).

[A COMPLETER]

7.2 Estimation des ordres p et q d'un modèle $ARMA(p, q)$

Pour l'estimation des paramètres p et q , on utilise le fait que si (X_t) suit un $ARIMA(p, d, q)$, alors $(1 - L)^d X_t$ suit asymptotiquement un processus $ARMA(p, q)$. En pratique, l'idée est d'appliquer la règle suivante : si $(X_t) \sim ARIMA(p, d, q)$ alors $(1 - L)^d X_t \sim ARMA(p, q)$.

On appelle processus $ARMA(p, q)$, un processus stationnaire (X_t) vérifiant une relation du type

$$X_t + \sum_{i=1}^p \phi_i X_{t-i} = \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \text{ pour tout } t \in \mathbb{Z}, \tag{31}$$

où les θ_i sont des réels et (ε_t) est un bruit blanc de variance σ^2 . (22) est équivalent à l'écriture

$$\Phi(L) X_t = \Theta(L) \varepsilon_t \text{ où } \begin{cases} \Theta(L) = \mathbb{I} + \theta_1 L + \dots + \theta_q L^q \\ \Phi(L) = \mathbb{I} + \phi_1 L + \dots + \phi_p L^p. \end{cases} \tag{32}$$

On supposera de plus que les polyômes Φ et Θ n'ont pas de racines en module strictement supérieures à 1 (*écriture sous forme canonique*), et n'ont pas de racine commune. On supposera de plus que les degrés de Θ et Φ sont respectivement q et p , au sens où $\theta_q \neq 0$ et $\phi_p \neq 0$.

7.2.1 Problèmes d'unicité de la représentation ARMA

On peut noter que l'écriture ARMA (32) n'est pas unique. En effet, il suffit de multiplier à gauche et à droite de (32) par un même polynôme en L , $\Gamma(L)$. Alors, en posant $\Phi^*(L) = \Gamma(L) \circ \Phi(L)$ et $\Theta^*(L) = \Gamma(L) \circ \Theta(L)$, on peut noter que $\Phi(L)^* X_t = \Theta(L)^* \varepsilon_t$.

Propriété 47 Soit Φ un polynôme dont les racines $z \in \mathbb{C}$ soient toutes à l'extérieur du disque unité. Alors l'équation $\Phi(L) X_t = \Theta(L) \varepsilon_t$ admet une solution stationnaire (X_t) et celle-ci est unique.

Définissons la matrice suivante, à partir des autocorrélations $\rho(h)$ du processus stationnaire (X_t)

$$i, j = \begin{bmatrix} \rho(i) & \rho(i-1) & \rho(i-2) & \dots & \rho(i-j+2) & \rho(i-j+1) \\ \rho(i+1) & \rho(i) & \rho(i-1) & \dots & \rho(i-j+3) & \rho(i-j+2) \\ \rho(i+2) & \rho(i+1) & \rho(i) & \dots & \rho(i-j+4) & \rho(i-j+3) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho(i+j-2) & \rho(i+j-3) & \rho(i+j-4) & \dots & \rho(i) & \rho(i-1) \\ \rho(i+j-1) & \rho(i+j-2) & \rho(i+j-3) & \dots & \rho(i+1) & \rho(i) \end{bmatrix}$$

et soit $\Delta(i, j)$ son déterminant.

Définition 44 Un processus (X_t) est un ARMA (p, q) minimal si $\Phi(L) X_t = \Theta(L) \varepsilon_t$ où (ε_t) est un bruit blanc et où Φ et Θ sont de degré respectif p et q (avec $\phi_p \neq 0$ et $\theta_q \neq 0$) dont les racines sont de module supérieur à 1, et où Φ et Θ n'ont pas de racines communes.

Propriété 48 Le processus (X_t) est un ARMA (p, q) minimal si et seulement si

- (i) $\Delta(i, j) = 0$ pour $i \geq q+1$ et $j \geq p+1$,
- (ii) $\Delta(i, j) \neq 0$ pour $i \geq q$,
- (iii) $\Delta(i, j) \neq 0$ pour $j \geq p$.

Autrement dit, on peut construire le tableau des $\Delta(i, j)$, et il aura la forme suivante pour un processus ARMA (p, q) minimal,

$i \setminus j$	1	2	...	p	$p+1$	$p+2$...
1	$\Delta(1, 1)$	$\Delta(1, 2)$...	$\Delta(1, p)$	$\Delta(1, p+1)$	$\Delta(1, p+2)$...
2	$\Delta(2, 1)$	$\Delta(2, 2)$...	$\Delta(2, p)$	$\Delta(2, p+1)$	$\Delta(2, p+2)$...
\vdots	\vdots	\vdots		\vdots	\vdots		
q	$\Delta(q, 1)$	$\Delta(q, 2)$...	$\Delta(q, p)$	$\Delta(q, p+1)$	$\Delta(q, p+2)$...
$q+1$	$\Delta(q+1, 1)$	$\Delta(q+1, 2)$...	$\Delta(q+1, p)$	0	0	...
$q+2$	$\Delta(q+2, 1)$	$\Delta(q+2, 2)$		$\Delta(q+2, p)$	0	0	...
\vdots	\vdots	\vdots		\vdots	\vdots	\vdots	

soit $\left[\begin{array}{c|c} D_{p,q} & D_p \\ \hline D_q & 0 \end{array} \right]$

où les termes $D_{p,q}$, D_q et D_p sont non-nuls.

Remarque 49 Dans le cas d'un processus MA (q) , le tableau des $\Delta(i, j)$ a la forme suivante

$$\left[\begin{array}{c} D_q \\ \hline 0 \end{array} \right] = \begin{bmatrix} i \setminus j & 1 & 2 & \dots \\ 1 & \Delta(1, 1) & \Delta(1, 2) & \dots \\ 2 & \Delta(2, 1) & \Delta(2, 2) & \dots \\ \vdots & \vdots & \vdots & \\ q & \Delta(q, 1) & \Delta(q, 2) & \dots \\ q+1 & 0 & 0 & \dots \\ q+2 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \end{bmatrix}$$

Remarque 50 Dans le cas d'un processus AR (p) , le tableau des $\Delta(i, j)$ a la forme suivante

$$\left[D_p \mid 0 \right] = \begin{bmatrix} i \setminus j & 1 & 2 & \dots & p & p+1 & p+2 & \dots \\ 1 & \Delta(1, 1) & \Delta(1, 2) & \dots & \Delta(1, p) & 0 & 0 & \dots \\ 2 & \Delta(2, 1) & \Delta(2, 2) & \dots & \Delta(2, p) & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \end{bmatrix}$$

L'autocorrélogramme partiel s'écrit $a(h) = (-1)^{j-1} \Delta(1, j) / \Delta(0, j)$ où $\Delta(0, j)$ est strictement positif (comme déterminant d'une matrice de corrélation) pour un $AR(p)$, et donc $a(h) = 0$ pour $h \geq p + 1$.

7.2.2 Comportement asymptotique des moments empiriques

Nous avons vu dans la partie (2.4.3) que les moments empiriques $(\gamma(h), \rho(h), \rho_i(h) \dots)$ convergent vers les vraies moments.

Propriété 49 Sous l'hypothèse où $(X_t) \sim MA(q)$, et que (ε_t) est stationnaire à l'ordre 4, alors

$$\sqrt{T} \frac{\hat{\rho}_T(h) - \rho(h)}{\sqrt{1 + 2 \sum_{k=1}^q \rho^2(k)}} \xrightarrow{\mathcal{L}} N(0, 1) \text{ pour } h > q.$$

Cette proposition permet en particulier d'avoir l'intervalle de confiance à 95% des autocorrélations,

$$\left[\hat{\rho}_T(h) \pm 1.96 \sqrt{\frac{1 + 2 \sum_{k=1}^q \rho^2(k)}{T}} \right]. \quad (33)$$

Propriété 50 Sous l'hypothèse où $(X_t) \sim AR(p)$, et que (ε_t) est stationnaire à l'ordre 4, alors

$$\sqrt{T} [\hat{a}_T(h) - a(h)] \xrightarrow{\mathcal{L}} N(0, 1) \text{ pour } h > q.$$

Cette proposition permet en particulier d'avoir l'intervalle de confiance à 95% des autocorrélations partielles,

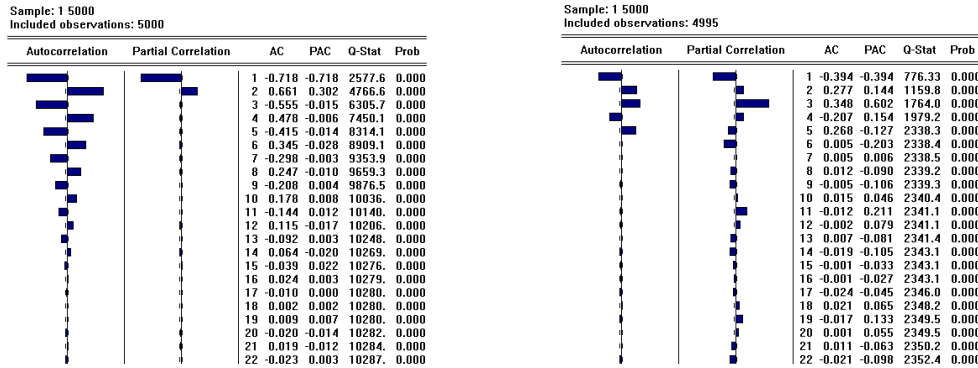
$$\left[\hat{a}_T(h) \pm 1.96 \frac{1}{\sqrt{T}} \right],$$

(résultat de Quenouille, 1949).

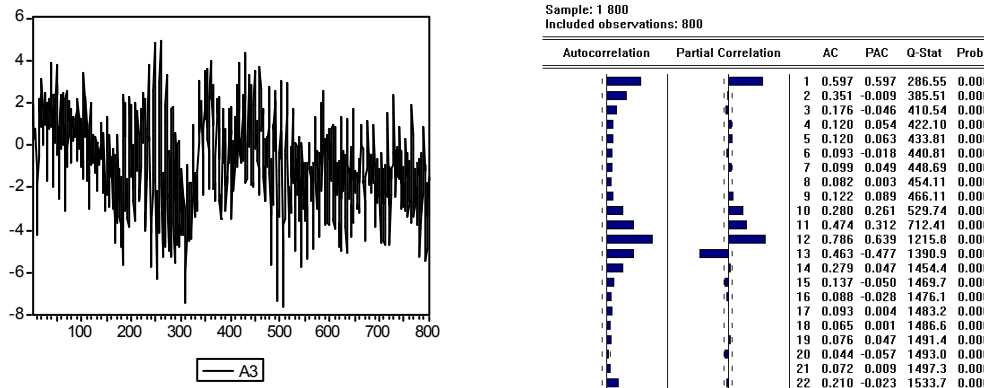
7.2.3 Méthode pratique d'estimation des ordres p et q

Pour estimer les ordres p ou q , on utilise les propriétés vues précédemment sur les formes des autocorrélogrammes $(\rho(h))$ ou des autocorrélogrammes partiels $(a(h))$. En particulier

- (i) pour les processus $AR(p)$ l'autocorrélogramme partiel s'annule à partir de p (à gauche)
- (ii) pour les processus $MA(q)$ l'autocorrélogramme s'annule à partir de q (à droite)



Remarque 51 S'il reste de la saisonnalité, celle-ci apparaîtra également dans les autocorrélogrammes



7.2.4 Cas d'un processus $MA(q)$

Si (X_t) suit un processus $MA(q)$, on peut noter que la variance des autocorrélations empiriques est donnée par la relation

$$V(\hat{\rho}(h)) \approx 1 + 2 \sum_{i=1}^q \rho^2(i), \text{ pour } h > q,$$

et donc, on peut prendre comme estimation de l'écart-type

$$\hat{\sigma}(\hat{\rho}(h)) = \frac{1}{\sqrt{T}} \sqrt{1 + 2 \sum_{i=1}^q \hat{\rho}^2(i)}.$$

En pratique, on identifie q , ordre d'un processus $MA(q)$ comme la première valeur à partir de laquelle les $\rho(h)$ sont dans l'intervalle dont les extrémités sont délimitées par

$$\pm \frac{1.96}{\sqrt{T}} \left[1 + 2 \left(\hat{\rho}^2(1) + \hat{\rho}^2(2) + \dots + \hat{\rho}^2(h-1) \right) \right]^{1/2},$$

puisque sous l'hypothèse où le processus est effectivement un $MA(q)$

$$\sqrt{T}\hat{\rho}(h) \xrightarrow{L} N(0, 1 + 2(\rho^2(1) + \dots + \rho^2(q-1))) \text{ pour } h > q.$$

7.2.5 Cas d'un processus $ARMA(p, q)$

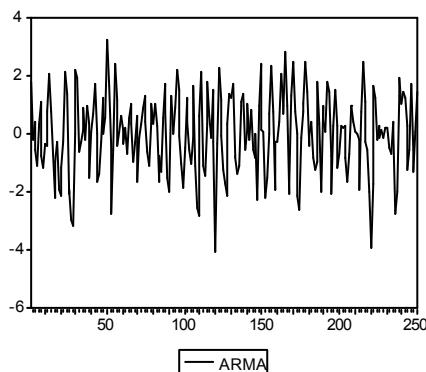
La méthode du coin (Beguïn, Gourieroux, Monfort) La méthode suivante, dite 'méthode du coin' permet d'estimer conjointement p et q lorsque les deux sont non-nuls. Elle est basée sur la propriété (48). Les valeurs de i, j où

$$i, j = \begin{bmatrix} \rho(i) & \rho(i-1) & \rho(i-2) & \dots & \rho(i-j+2) & \rho(i-j+1) \\ \rho(i+1) & \rho(i) & \rho(i-1) & \dots & \rho(i-j+3) & \rho(i-j+2) \\ \rho(i+2) & \rho(i+1) & \rho(i) & \dots & \rho(i-j+4) & \rho(i-j+3) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho(i+j-2) & \rho(i+j-3) & \rho(i+j-4) & \dots & \rho(i) & \rho(i-1) \\ \rho(i+j-1) & \rho(i+j-2) & \rho(i+j-3) & \dots & \rho(i+1) & \rho(i) \end{bmatrix}$$

sont inconnues mais peuvent être estimée par les $\hat{\rho}(h)$. On pose alors $\Delta(i, j) = \det i, j$, qui sera, de la même façon, estimé par $\hat{\Delta}(i, j) = \det \hat{i}, j$. Les $\hat{\Delta}(i, j)$ sont alors des estimateurs convergents des $\Delta(i, j)$ (par continuité du déterminant). Les coefficients p et q sont alors les valeurs pour lesquels s'observent une rupture. La variance asymptotique de $\hat{\Delta}(i, j)$ est une fonction différentiable du vecteur des autocorrélations $\hat{\rho}(h)$, avec une loi normale asymptotique.

Un test de nullité est basé sur l'utilisation de la statistique de Student $\hat{\Delta}(i, j) / \sqrt{\hat{V}(\hat{\Delta}(i, j))}$, qui doit être comparée à 1.96 pour un seuil de 5%.

Exemple 78 Considérons le processus simulé (sur 250 valeurs) $(1 - 0.5L)X_t = (1 + 0.1L - 0.7L^2)\varepsilon_t$ où (ε_t) est un bruit blanc gaussien de variance 1



Sample: 1 250
Included observations: 250

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.352	0.352	31.430	0.000	
2	-0.296	-0.480	53.727	0.000	
3	-0.316	0.011	79.237	0.000	
4	-0.179	-0.230	87.450	0.000	
5	0.036	-0.011	87.737	0.000	
6	0.048	-0.100	88.398	0.000	
7	0.184	0.185	97.136	0.000	
8	0.133	-0.100	101.75	0.000	
9	-0.047	0.050	102.31	0.000	
10	-0.072	0.012	103.69	0.000	
11	-0.061	-0.025	104.66	0.000	
12	-0.038	-0.004	105.95	0.000	
13	0.009	0.007	105.08	0.000	
14	0.029	-0.029	105.30	0.000	
15	0.048	0.031	105.93	0.000	
16	0.052	0.042	106.65	0.000	
17	0.022	0.009	106.78	0.000	
18	-0.043	-0.026	107.29	0.000	
19	0.005	0.135	107.30	0.000	
20	0.036	-0.056	107.65	0.000	
21	-0.046	-0.012	108.24	0.000	
22	-0.055	-0.006	109.08	0.000	

Le tableau des Δ_{ij} est donné par

$i \setminus j$	1	2	3	4	5	$i \setminus j$	1	2	3	4	5
1	0.352	0.420	0.006	0.095	-0.003	1	0.352	0.420	0.006	0.095	-0.003
2	-0.296	0.199	-0.067	0.022	-0.006	2	-0.296	0.199	-0.067	0.022	-0.006
3	-0.316	0.047	0.006	0.001	0.003	3	-0.316	0.047	0.000	0.000	0.000
4	-0.179	0.021	0.000	-0.001	-0.001	4	-0.179	0.021	0.000	0.000	0.000
5	-0.036	0.010	0.002	0.001	0.000	5	-0.036	0.010	0.000	0.000	0.000

En effet, par exemple, le terme $\Delta_{1,2}$ est donné par

$$\Delta_{1,2} = \begin{vmatrix} \rho(1) & \rho(0) \\ \rho(2) & \rho(1) \end{vmatrix} = \begin{vmatrix} 0.352 & 1 \\ -0.296 & 0.352 \end{vmatrix} = 0.352^2 + 0.296 = 0.420.$$

L'approximation indiquée ci-dessous semble valider l'hypothèse de modélisation $ARMA(1,2)$. Cette intuition est confirmée en étudiant le tableau des 'Student'.

Utilisation de la fonction d'autocorrélation étendue (Tsay, & Ciao) Cette méthode est appelée **ESCAF** (*Extended Sample Autocorrelation Function*) sous *SAS*, et est décrite dans la document *SAS - ETS* (pages 236-237). Pour cela, on effectue des regressions linéaires, de façon itérative pour calculer les paramètres *AR* d'un *ARMA* (*stationnaire ou pas*). Ensuite, à partir de cette estimation, la série observée est modélisée sous forme *MA*.

Soit (X_t) un processus *ARMA* (p, q) dont on observe n réalisations, suivant le modèle

$$\Phi(L) X_t = (1 - L)^d \phi(L) X_t = \Theta(L) \varepsilon_t \text{ où } (\varepsilon_t) \text{ suit un bruit blanc de variance } \sigma^2.$$

Sur la partie autorégressive du processus, on utilise une régression linéaire pour obtenir des estimateurs (*par moindres carrés*) des paramètres autorégressifs de la composante *AR*. On définit alors la première régression

$$X_t = \underbrace{\sum_{i=0}^p \phi_i^{p,1} X_{t-i} + \theta_i^{p,1} \eta_{t-1}^{p,0}}_{\text{Forme autorégressive}} + u_t^{p,1}.$$

Ce modèle est estimé par les *mco*. On définit alors la k -ème régression itérative d'un *AR* (m) quelconque

$$X_t = \sum_{i=0}^m \phi_i^{m,k} X_{t-i} + \sum_{j=0}^k \theta_j^{m,k} \eta_{t-j}^{j,k} + u_t^{m,k},$$

où les $\eta_t^{j,k}$ sont les erreurs du processus *AR* de la k -ième régression, et les $u_t^{m,k}$ les résidus de la régression. Comme on ignore l'ordre p de la partie autorégressive, on choisit m variant de 1 à p' , et on effectue q' régressions itératives : on choisira a priori p' et q' suffisamment grands. Les paramètres peuvent alors être estimés récursivement par

$$\hat{\phi}_i^{m,j} = \hat{\phi}_i^{m+1,j-1} - \hat{\phi}_{i-1}^{m,j-1} \frac{\hat{\phi}_{m+1}^{m+1,j-1}}{\hat{\phi}_m^{m,j-1}}.$$

Ces paramètres sont alors utilisés pour définir la *ESACF*, fonction d'autocorrélation étendue, telle que l'a définie Tsay et Tia (1984),

Définition 45 On appelle fonction d'autocorrélation étendue la fonction $r_j(m)$, fonction d'autocorrélation du processus

$$\omega_j^m = X_t - \sum_{i=1}^m \hat{\phi}_i^{m,j} X_{t-i} \text{ pour } j = 1, 2, \dots$$

Si le processus suit un *ARMA* $(p + d, q)$ la série ω_j^m suit un processus *MA* (q) pour $j \geq q$, c'est à dire

$$\begin{cases} r_j(p + d) \approx 0 \text{ pour } j > q \\ r_j(p + q) \neq 0 \text{ pour } j = q. \end{cases} \quad (34)$$

La table *ESACF* est alors définie par

AR\MA	0	1	2	3	...
0	$r_1(0)$	$r_2(0)$	$r_3(0)$	$r_4(0)$...
1	$r_1(1)$	$r_2(1)$	$r_3(1)$	$r_4(1)$...
2	$r_1(2)$	$r_2(2)$	$r_3(2)$	$r_4(2)$...
3	$r_1(3)$	$r_2(3)$	$r_3(3)$	$r_4(3)$...
⋮	⋮	⋮	⋮	⋮	⋮

La nullité théorique de (34) est interprétée statistiquement par une valeur inférieure à $1.96/\sqrt{n}$.

Méthode 'SCAN' Cette méthode vise à utiliser la plus petite corrélation canonique (*smallest canonical correlation*) pour identifier les ordres p et q . Elle a été implémentée sous *SAS*, et est décrite dans la document *SAS – ETS (pages 239-241)*. La syntaxe pour utiliser cette procédure ou la procédure *ESACF* est la suivante

```
proc arima data = base;
identify var = x esacf scan;
run;
```

Considérons une série X_t que nous allons centrer, $Z_t = X_t - \mu$, dont on observe n réalisations, suivant un processus *ARIMA* (p, d, q). Cette méthode analyse les valeurs propres de la matrice de corrélation du processus.

[A COMPLETER]

Exemple 79 Dans le cas d'un processus *ARMA*(2, 1), les tables *ESACF* et *SCAN* théoriques seront de la forme

Table ESACF								Table SCAN							
AR\MA	0	1	2	3	4	5	6	AR\MA	0	1	2	3	4	5	6
0	*	*	*	*	*	*	*	0	*	*	*	*	*	*	*
1	*	*	*	*	*	*	*	1	*	*	*	*	*	*	*
2	*	0	0	0	0	0	0	2	*	0	0	0	0	0	0
3	*	*	0	0	0	0	0	3	*	0	0	0	0	0	0
4	*	*	*	0	0	0	0	4	*	0	0	0	0	0	0

où l'ordre de l'AR se lit à gauche, et l'ordre du MA se lit en haut. Dans l'exemple ci-dessous, nous avons simulé 1500 réalisations d'un tel processus.

<p>ARIMA Procedure</p> <p>Extended Sample Autocorrelation Function</p> <table border="1"> <thead> <tr> <th>Lags</th> <th>MA 0</th> <th>MA 1</th> <th>MA 2</th> <th>MA 3</th> <th>MA 4</th> <th>MA 5</th> </tr> </thead> <tbody> <tr><td>AR 0</td><td>0.5312</td><td>0.0134</td><td>-0.3184</td><td>-0.4086</td><td>-0.3562</td><td>-0.2131</td></tr> <tr><td>AR 1</td><td>0.5206</td><td>0.0203</td><td>-0.2599</td><td>-0.1589</td><td>-0.1552</td><td>-0.1731</td></tr> <tr><td>AR 2</td><td>-0.4404</td><td>-0.0326</td><td>-0.0757</td><td>0.0609</td><td>0.0045</td><td>-0.0305</td></tr> <tr><td>AR 3</td><td>-0.4803</td><td>0.1110</td><td>-0.1018</td><td>0.0609</td><td>0.0114</td><td>-0.0188</td></tr> <tr><td>AR 4</td><td>-0.4981</td><td>-0.5046</td><td>-0.3928</td><td>0.0213</td><td>0.0523</td><td>-0.0252</td></tr> <tr><td>AR 5</td><td>-0.4721</td><td>-0.0174</td><td>-0.3009</td><td>0.1085</td><td>0.0143</td><td>-0.0257</td></tr> </tbody> </table> <p>ESACF Probability Values</p> <table border="1"> <thead> <tr> <th>Lags</th> <th>MA 0</th> <th>MA 1</th> <th>MA 2</th> <th>MA 3</th> <th>MA 4</th> <th>MA 5</th> </tr> </thead> <tbody> <tr><td>AR 0</td><td>0.0001</td><td>0.6786</td><td>0.0001</td><td>0.0001</td><td>0.0001</td><td>0.0001</td></tr> <tr><td>AR 1</td><td>0.0001</td><td>0.5352</td><td>0.0001</td><td>0.0001</td><td>0.0001</td><td>0.0001</td></tr> <tr><td>AR 2</td><td>0.0001</td><td>0.2836</td><td>0.0094</td><td>0.0496</td><td>0.8859</td><td>0.3192</td></tr> <tr><td>AR 3</td><td>0.0001</td><td>0.0010</td><td>0.0012</td><td>0.0544</td><td>0.7021</td><td>0.5632</td></tr> <tr><td>AR 4</td><td>0.0001</td><td>0.0001</td><td>0.0001</td><td>0.4967</td><td>0.1254</td><td>0.4283</td></tr> <tr><td>AR 5</td><td>0.0001</td><td>0.5715</td><td>0.0001</td><td>0.0021</td><td>0.6058</td><td>0.4184</td></tr> </tbody> </table> <p>ARMA(p+d,q) Tentative Order Selection Tests (5% Significance Level)</p> <table border="1"> <thead> <tr> <th>ESACF</th> <th>p+d</th> <th>q</th> </tr> </thead> <tbody> <tr><td>3</td><td>3</td><td></td></tr> <tr><td>4</td><td>3</td><td></td></tr> <tr><td>2</td><td>4</td><td></td></tr> </tbody> </table>	Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5	AR 0	0.5312	0.0134	-0.3184	-0.4086	-0.3562	-0.2131	AR 1	0.5206	0.0203	-0.2599	-0.1589	-0.1552	-0.1731	AR 2	-0.4404	-0.0326	-0.0757	0.0609	0.0045	-0.0305	AR 3	-0.4803	0.1110	-0.1018	0.0609	0.0114	-0.0188	AR 4	-0.4981	-0.5046	-0.3928	0.0213	0.0523	-0.0252	AR 5	-0.4721	-0.0174	-0.3009	0.1085	0.0143	-0.0257	Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5	AR 0	0.0001	0.6786	0.0001	0.0001	0.0001	0.0001	AR 1	0.0001	0.5352	0.0001	0.0001	0.0001	0.0001	AR 2	0.0001	0.2836	0.0094	0.0496	0.8859	0.3192	AR 3	0.0001	0.0010	0.0012	0.0544	0.7021	0.5632	AR 4	0.0001	0.0001	0.0001	0.4967	0.1254	0.4283	AR 5	0.0001	0.5715	0.0001	0.0021	0.6058	0.4184	ESACF	p+d	q	3	3		4	3		2	4		<p>ARIMA Procedure</p> <p>Squared Canonical Correlation Estimates</p> <table border="1"> <thead> <tr> <th>Lags</th> <th>MA 0</th> <th>MA 1</th> <th>MA 2</th> <th>MA 3</th> <th>MA 4</th> <th>MA 5</th> </tr> </thead> <tbody> <tr><td>AR 0</td><td>0.2828</td><td>0.0002</td><td>0.1020</td><td>0.1685</td><td>0.1282</td><td>0.0459</td></tr> <tr><td>AR 1</td><td>0.1413</td><td>0.1353</td><td>0.1208</td><td>0.0242</td><td>0.0148</td><td>0.0145</td></tr> <tr><td>AR 2</td><td>0.0441</td><td>0.0005</td><td>0.0023</td><td>0.0013</td><td>0.0001</td><td>0.0004</td></tr> <tr><td>AR 3</td><td>0.0225</td><td>0.0017</td><td>0.0024</td><td>0.0017</td><td>0.0004</td><td>0.0003</td></tr> <tr><td>AR 4</td><td>0.0323</td><td>0.0021</td><td>0.0025</td><td>0.0001</td><td>0.0005</td><td>0.0003</td></tr> <tr><td>AR 5</td><td>0.0154</td><td>0.0001</td><td>0.0005</td><td>0.0005</td><td>0.0002</td><td>0.0006</td></tr> </tbody> </table> <p>SCAN Chi-Square[1] Probability Values</p> <table border="1"> <thead> <tr> <th>Lags</th> <th>MA 0</th> <th>MA 1</th> <th>MA 2</th> <th>MA 3</th> <th>MA 4</th> <th>MA 5</th> </tr> </thead> <tbody> <tr><td>AR 0</td><td>0.0001</td><td>0.6782</td><td>0.0001</td><td>0.0001</td><td>0.0001</td><td>0.0001</td></tr> <tr><td>AR 1</td><td>0.0001</td><td>0.0001</td><td>0.0001</td><td>0.0001</td><td>0.0001</td><td>0.0001</td></tr> <tr><td>AR 2</td><td>0.0001</td><td>0.4834</td><td>0.1152</td><td>0.2162</td><td>0.9077</td><td>0.5113</td></tr> <tr><td>AR 3</td><td>0.0001</td><td>0.1990</td><td>0.1480</td><td>0.1700</td><td>0.4922</td><td>0.6210</td></tr> <tr><td>AR 4</td><td>0.0001</td><td>0.1510</td><td>0.1198</td><td>0.8188</td><td>0.4879</td><td>0.5618</td></tr> <tr><td>AR 5</td><td>0.0001</td><td>0.9782</td><td>0.4712</td><td>0.4376</td><td>0.6296</td><td>0.4168</td></tr> </tbody> </table> <p>ARMA(p+d,q) Tentative Order Selection Tests (5% Significance Level)</p> <table border="1"> <thead> <tr> <th>SCAN</th> <th>p+d</th> <th>q</th> </tr> </thead> <tbody> <tr><td>2</td><td>1</td><td></td></tr> </tbody> </table>	Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5	AR 0	0.2828	0.0002	0.1020	0.1685	0.1282	0.0459	AR 1	0.1413	0.1353	0.1208	0.0242	0.0148	0.0145	AR 2	0.0441	0.0005	0.0023	0.0013	0.0001	0.0004	AR 3	0.0225	0.0017	0.0024	0.0017	0.0004	0.0003	AR 4	0.0323	0.0021	0.0025	0.0001	0.0005	0.0003	AR 5	0.0154	0.0001	0.0005	0.0005	0.0002	0.0006	Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5	AR 0	0.0001	0.6782	0.0001	0.0001	0.0001	0.0001	AR 1	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	AR 2	0.0001	0.4834	0.1152	0.2162	0.9077	0.5113	AR 3	0.0001	0.1990	0.1480	0.1700	0.4922	0.6210	AR 4	0.0001	0.1510	0.1198	0.8188	0.4879	0.5618	AR 5	0.0001	0.9782	0.4712	0.4376	0.6296	0.4168	SCAN	p+d	q	2	1	
Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5																																																																																																																																																																																																																	
AR 0	0.5312	0.0134	-0.3184	-0.4086	-0.3562	-0.2131																																																																																																																																																																																																																	
AR 1	0.5206	0.0203	-0.2599	-0.1589	-0.1552	-0.1731																																																																																																																																																																																																																	
AR 2	-0.4404	-0.0326	-0.0757	0.0609	0.0045	-0.0305																																																																																																																																																																																																																	
AR 3	-0.4803	0.1110	-0.1018	0.0609	0.0114	-0.0188																																																																																																																																																																																																																	
AR 4	-0.4981	-0.5046	-0.3928	0.0213	0.0523	-0.0252																																																																																																																																																																																																																	
AR 5	-0.4721	-0.0174	-0.3009	0.1085	0.0143	-0.0257																																																																																																																																																																																																																	
Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5																																																																																																																																																																																																																	
AR 0	0.0001	0.6786	0.0001	0.0001	0.0001	0.0001																																																																																																																																																																																																																	
AR 1	0.0001	0.5352	0.0001	0.0001	0.0001	0.0001																																																																																																																																																																																																																	
AR 2	0.0001	0.2836	0.0094	0.0496	0.8859	0.3192																																																																																																																																																																																																																	
AR 3	0.0001	0.0010	0.0012	0.0544	0.7021	0.5632																																																																																																																																																																																																																	
AR 4	0.0001	0.0001	0.0001	0.4967	0.1254	0.4283																																																																																																																																																																																																																	
AR 5	0.0001	0.5715	0.0001	0.0021	0.6058	0.4184																																																																																																																																																																																																																	
ESACF	p+d	q																																																																																																																																																																																																																					
3	3																																																																																																																																																																																																																						
4	3																																																																																																																																																																																																																						
2	4																																																																																																																																																																																																																						
Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5																																																																																																																																																																																																																	
AR 0	0.2828	0.0002	0.1020	0.1685	0.1282	0.0459																																																																																																																																																																																																																	
AR 1	0.1413	0.1353	0.1208	0.0242	0.0148	0.0145																																																																																																																																																																																																																	
AR 2	0.0441	0.0005	0.0023	0.0013	0.0001	0.0004																																																																																																																																																																																																																	
AR 3	0.0225	0.0017	0.0024	0.0017	0.0004	0.0003																																																																																																																																																																																																																	
AR 4	0.0323	0.0021	0.0025	0.0001	0.0005	0.0003																																																																																																																																																																																																																	
AR 5	0.0154	0.0001	0.0005	0.0005	0.0002	0.0006																																																																																																																																																																																																																	
Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5																																																																																																																																																																																																																	
AR 0	0.0001	0.6782	0.0001	0.0001	0.0001	0.0001																																																																																																																																																																																																																	
AR 1	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001																																																																																																																																																																																																																	
AR 2	0.0001	0.4834	0.1152	0.2162	0.9077	0.5113																																																																																																																																																																																																																	
AR 3	0.0001	0.1990	0.1480	0.1700	0.4922	0.6210																																																																																																																																																																																																																	
AR 4	0.0001	0.1510	0.1198	0.8188	0.4879	0.5618																																																																																																																																																																																																																	
AR 5	0.0001	0.9782	0.4712	0.4376	0.6296	0.4168																																																																																																																																																																																																																	
SCAN	p+d	q																																																																																																																																																																																																																					
2	1																																																																																																																																																																																																																						

Les sorties *ESACF*, et *SCAN* peuvent se réécrire

Table ESACF							Table SCAN						
AR\MA	0	1	2	3	4	5	AR\MA	0	1	2	3	4	5
0	0.53	0.01	-0.32	-0.41	-0.36	-0.21	0	0.28	0.00	0.10	0.17	0.13	0.05
1	0.52	0.02	-0.25	-0.16	-0.16	-0.17	1	0.14	0.14	0.12	0.02	0.01	0.01
2	-0.44	-0.03	-0.08	-0.06	0.00	-0.03	2	0.04	0.00	0.00	0.00	0.00	0.00
3	-0.48	0.11	-0.10	-0.06	0.01	-0.02	3	0.02	0.00	0.00	0.00	0.00	0.00
4	-0.50	-0.51	-0.40	0.02	0.05	-0.02	4	0.03	0.00	0.00	0.00	0.00	0.00

Comme on peut le noter, la méthode SCAN donne de très bon résultats, et permet d'identifier les ordres 2 et 1.

Remarque 52 SAS propose une méthode supplémentaire pour estimer l'ordre des processus ARMA : la méthode MINIC.

7.2.6 Propriété des estimateurs

En notant $\omega = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)'$, on a le résultat suivant

Propriété 51 L'estimateur du maximum de vraisemblance est convergent, et asymptotiquement normal,

$$\begin{bmatrix} \sqrt{T}(\hat{\omega}_T - \omega) \\ \sqrt{T}(\hat{\sigma}_T - \sigma) \end{bmatrix} \xrightarrow{\mathcal{L}} \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} & 0 \\ 0 & a \end{bmatrix} \right).$$

Cette propriété permet de mettre en place des tests sur les paramètres.

7.3 Test de bruit blanc et de stationnarité

L'hypothèse $(X_t) \sim ARIMA(p, d, q)$ peut s'écrire $(1 - L)^d \Phi(L) X_t = \Theta(L) \varepsilon_t$, ou encore

$$\varepsilon_t = \Theta(L)^{-1} (1 - L)^d \Phi(L) X_t.$$

Une fois estimés les paramètres d, p, q et l'ensemble des ϕ_i et θ_j , on obtient des polynômes estimés $\hat{\Theta}(L)$ et $\hat{\Phi}(L)$, qui permettent d'obtenir les résidus estimés,

$$\hat{\varepsilon}_t = \hat{\Theta}(L)^{-1} (1 - L)^d \hat{\Phi}(L) X_t.$$

Pour que les modèles obtenus précédemment soient valides, il convient de vérifier que les résidus estimés suivent bien un bruit blanc $H_0 : (\varepsilon_t) \sim BB$.

7.3.1 Analyse des fonctions d'autocorrélation

L'autocorrélation peut être estimée par

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} \text{ où } \hat{\gamma}(h) = \frac{1}{n-h} \sum_{t=1}^{n-h} (X_t - \bar{X})(X_{t+h} - \bar{X}) \text{ et } \bar{X} = \frac{1}{n} \sum_{t=1}^n X_t.$$

L'intervalle de confiance de $\hat{\rho}(h)$ est, dans le cas d'un bruit blanc gaussien $[-t_{\alpha/2}/\sqrt{T}; t_{\alpha/2}/\sqrt{T}]$ où $t_{\alpha/2}$ est le quantile d'ordre $\alpha/2$ de la loi de Student (1.96 pour $\alpha = 5\%$). Pour avoir un bruit blanc, il est nécessaire qu'aucune valeur de l'autocorrélogramme ne soit significativement non-nulle.

Exemple 80 Pour la série (1) à gauche, aucune valeur n'est significativement non-nulle alors que pour la série (2), à droite, certaines le sont, en particulier pour $h = 8$ ou $h = 16$.

Sample: 1 1000
Included observations: 1000

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.000	0.000	0.0002	0.990
		2 0.010	0.010	0.1020	0.950
		3 -0.027	-0.027	0.0191	0.845
		4 0.057	0.057	4.0955	0.393
		5 0.019	0.020	4.4761	0.483
		6 0.049	0.047	6.8518	0.335
		7 0.047	0.050	9.0869	0.246
		8 -0.040	-0.043	10.676	0.221
		9 -0.025	-0.026	11.310	0.255
		10 0.009	0.006	11.388	0.328
		11 -0.013	-0.022	11.556	0.398
		12 -0.008	-0.009	11.629	0.476
		13 -0.034	-0.034	12.803	0.463
		14 0.024	0.025	13.373	0.497
		15 0.017	0.025	13.650	0.552
		16 0.016	0.016	13.914	0.605
		17 -0.009	-0.005	14.005	0.667
		18 0.013	0.016	14.190	0.717
		19 -0.014	-0.013	14.383	0.761
		20 0.027	0.023	15.106	0.770
		21 0.009	0.003	15.187	0.813
		22 -0.009	-0.016	15.277	0.850

Sample: 1 1000
Included observations: 1000

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.046	-0.046	2.0882	0.148
		2 -0.011	-0.013	2.2062	0.332
		3 -0.043	-0.044	4.0590	0.255
		4 -0.025	-0.029	4.6726	0.323
		5 0.051	0.047	7.2646	0.202
		6 0.037	0.039	8.6435	0.195
		7 -0.041	-0.039	10.341	0.170
		8 -0.094	-0.094	19.234	0.014
		9 -0.007	-0.011	19.281	0.023
		10 0.001	-0.006	19.281	0.037
		11 0.011	-0.003	19.403	0.054
		12 0.025	0.023	20.035	0.066
		13 0.024	0.039	20.620	0.081
		14 -0.052	-0.043	23.372	0.054
		15 -0.007	-0.017	23.423	0.076
		16 0.088	0.082	31.374	0.012
		17 -0.021	-0.019	31.806	0.016
		18 -0.009	-0.018	31.871	0.023
		19 -0.026	-0.016	32.560	0.027
		20 0.014	0.028	32.783	0.036
		21 -0.024	-0.031	33.350	0.042
		22 0.031	0.014	34.356	0.045

7.3.2 Statistique de Box-Pierce, ou test de "portmanteau"

Le test de Box-Pierce permet d'identifier les processus de bruit blanc (*i.e.* les processus aléatoires de moyenne nulle, de variance constante et non autocorrélés). Cette statistique permet de tester $cov(\varepsilon_t, \varepsilon_{t-h}) = 0$ pour tout h , soit $\rho(h) = 0$ pour tout h . Ce test s'écrit

$$\begin{cases} H_0 : \rho(1) = \rho(2) = \dots = \rho(h) = 0 \\ H_a : \text{il existe } i \text{ tel que } \rho(i) \neq 0. \end{cases}$$

Pour effectuer ce test, on utilise la statistique de Box et Pierce (1970) Q , donnée par

$$Q_h = T \sum_{k=1}^h \hat{\rho}_k^2,$$

où h est le nombre de retards, T est le nombre d'observations et $\hat{\rho}_k$ l'autocorrélation empirique. Asymptotiquement, sous H_0 , Q_h suit un χ^2 à h degrés de liberté. Nous rejetons l'hypothèse de bruit blanc au seuil h si Q est supérieure au quantile d'ordre $(1 - \alpha)$ de la loi du χ^2 à h degrés de liberté.

Une statistique ayant de meilleures propriétés asymptotiques peut être utilisée :

$$Q'_h = T(T+2) \sum_{k=1}^h \frac{\hat{\rho}_k}{T-k},$$

qui suit asymptotiquement, sous H_0 une loi du χ^2 à h degrés de liberté. Ces tests sont appelés par les anglo-saxons 'portmanteau tests', soit littéralement tests 'fourre-tout'.

Exemple 81 Cette statistique est généralement fournie avec l'autocorrélogramme (*Q-stat*). Les deux sorties ci-dessous correspondent aux valeurs pour 2 séries de résidus

Sample: 1 1000 Included observations: 1000							Sample: 1 1000 Included observations: 1000						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob		Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.000	0.000	0.0002	0.990			1	-0.046	-0.046	2.0882	0.148
		2	0.010	0.010	0.1020	0.950			2	-0.011	-0.013	2.2062	0.332
		3	-0.027	-0.027	0.8191	0.845			3	-0.043	-0.044	4.0590	0.255
		4	0.057	0.057	4.0955	0.393			4	-0.025	-0.029	4.6726	0.323
		5	0.019	0.020	4.4761	0.483			5	0.051	0.047	7.2646	0.202
		6	0.049	0.047	6.8518	0.335			6	0.037	0.039	8.6435	0.195
		7	0.047	0.050	9.0869	0.246			7	-0.041	-0.039	10.341	0.170
		8	-0.040	-0.043	10.676	0.221			8	-0.094	-0.094	19.234	0.014
		9	-0.025	-0.026	11.310	0.255			9	-0.007	-0.011	19.281	0.023
		10	0.009	0.006	11.388	0.328			10	0.001	-0.006	19.281	0.037
		11	-0.013	-0.022	11.556	0.398			11	0.011	-0.003	19.403	0.054
		12	-0.008	-0.009	11.629	0.476			12	0.025	0.023	20.035	0.066
		13	-0.034	-0.034	12.803	0.463			13	0.024	0.039	20.620	0.081
		14	0.024	0.025	13.373	0.497			14	-0.052	-0.043	23.372	0.054
		15	0.017	0.025	13.650	0.552			15	-0.007	-0.017	23.423	0.076
		16	0.016	0.016	13.914	0.605			16	0.088	0.082	31.374	0.012
		17	-0.009	-0.005	14.005	0.667			17	-0.021	-0.019	31.806	0.016
		18	0.013	0.016	14.190	0.717			18	-0.008	-0.018	31.871	0.023
		19	-0.014	-0.013	14.383	0.761			19	-0.026	-0.016	32.580	0.027
		20	0.027	0.023	15.106	0.770			20	0.014	0.028	32.783	0.036
		21	0.009	0.003	15.187	0.813			21	-0.024	-0.031	33.350	0.042
		22	-0.009	-0.016	15.277	0.850			22	0.031	0.014	34.356	0.045

La table du χ^2 est donnée ci-dessous. A titre comparatif, nous obtenons le tableau suivant

h	1	2	3	4	5	6	7	8	9	10
Série (1)	0.000	0.102	0.819	4.095	4.476	6.852	9.087	10.676	11.310	11.388
Série (2)	2.088	2.206	4.059	4.673	7.2646	8.643	10.341	19.234	19.281	19.281
$\chi_{10\%}(h)$	2.706	4.605	6.251	7.779	9.236	10.645	12.017	13.362	14.684	15.987
$\chi_{5\%}(h)$	3.841	5.991	7.815	9.488	11.070	12.592	14.067	15.507	16.919	18.307

Si la série (1) est statistiquement un bruit blanc, il ne semble pas en être de même pour la seconde série, pour laquelle Q_h est parfois trop élevée (en particulier à partir de $h = 8$ - ce qui était confirmé par l'analyse graphique des autocorrélogrammes, avec cette valeur $\rho(8)$ significativement non nulle). Le seuil apparaît d'ailleurs en pointillé sous EViews, et en '.' sous SAS.

7.3.3 Compléments : les tests de normalité

Dans le cadre de la prévision, ou lors des tests de Student sur les paramètres, il convient de vérifier la normalité des résidus. Un test possible est celui de Bera & Jarque (1984), basé sur le skewness (coefficient d'asymétrie de la distribution) et la kurtosis (aplatissement - épaisseur des queues).

En notant μ_k le moment d'ordre k de la distribution, $\mu_k = \mathbb{E} \left([X - \mathbb{E}(X)]^k \right)$, on appelle skewness le coefficient $s = \mu_3 / \mu_2^{3/2}$ et kurtosis $k = \mu_4 / \mu_2^2$. Sous des hypothèses de normalité, on a normalité des estimateurs du skewness et de la kurtosis,

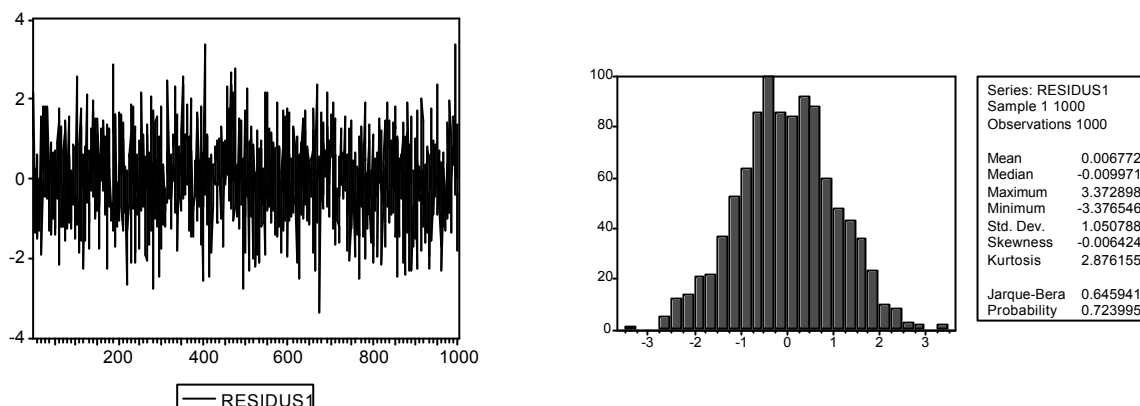
$$s \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \sqrt{6/T} \right) \text{ et } k \xrightarrow{\mathcal{L}} \mathcal{N} \left(3, \sqrt{24/T} \right) \text{ quand } T \rightarrow \infty.$$

Le test de Bera & Jarque repose sur le fait que, si la distribution suit une loi normale, alors la quantité

$$BJ = \frac{T}{6} s^2 + \frac{T}{24} [k - 3]^2,$$

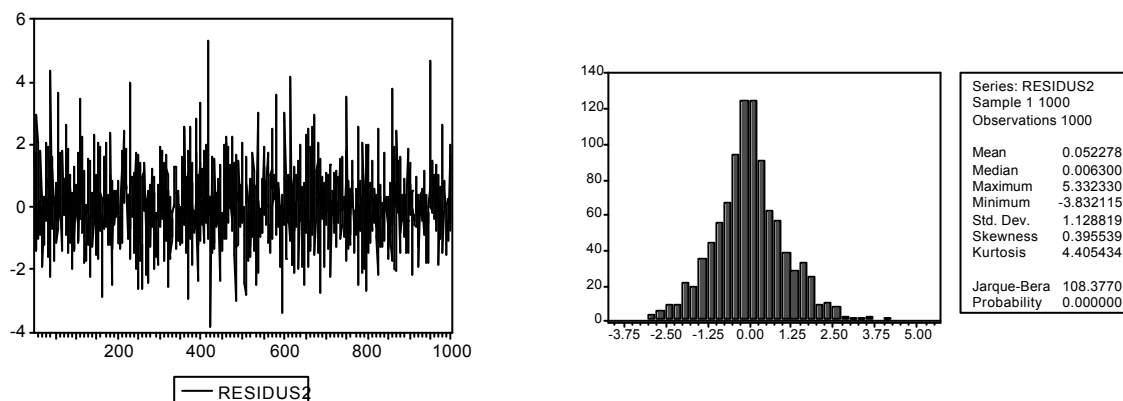
suit asymptotiquement une loi du χ^2 à 2 degrés de liberté. Aussi, si $BJ \geq \chi_{1-\alpha}^2(2)$ on rejette l'hypothèse H_0 de normalité des résidus au seuil α .

Exemple 82 Dans l'exemple ci-dessous, les résidus suivent effectivement une loi normale $\mathcal{N}(0, 1)$



La kurtosis vaut 2.876 (3 pour une loi normale) et le skewness -0.006 (0 pour un loi normale). La statistique de Bera & Jarque vaut alors 0.6459, qui est inférieur le centile du χ^2 à 2 degrés de liberté, au seuil $\alpha = 5\%$, soit 5.991 (table ci-dessous).

Exemple 83 Dans l'exemple ci-dessous, les résidus ne suivent pas une loi normale $\mathcal{N}(0, 1)$



La distribution est beaucoup trop aplatie pour suivre une loi normale (kurtosis valant 4.4), avec un asymétrie également trop forte. La statistique de Bera & Jarque vaut alors $108 \gg 5.991$.

Remarque 53 La table ci-dessous est la table de la loi du chi-deux, où P est le seuil α et ν le nombre de degrés de

liberté

ν	$P=0,90$	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01
1	0,0158	0,0642	0,148	0,455	1,074	1,642	2,706	3,841	5,412	6,635
2	0,211	0,446	0,713	1,386	2,408	3,219	4,605	5,991	7,824	9,210
3	0,584	1,005	1,424	2,366	3,665	4,642	6,251	7,815	9,837	11,345
4	1,064	1,649	2,195	3,357	4,878	5,989	7,779	9,488	11,668	13,277
5	1,610	2,343	3,000	4,351	6,064	7,289	9,236	11,070	13,388	15,086
6	2,204	3,070	3,828	5,348	7,231	8,558	10,645	12,592	15,033	16,812
7	2,833	3,822	4,671	6,346	8,383	9,803	12,017	14,067	16,662	18,475
8	3,490	4,594	5,527	7,344	9,524	11,030	13,362	15,507	18,168	20,090
9	4,168	5,380	6,393	8,343	10,656	12,242	14,684	16,919	19,679	21,666
10	4,865	6,179	7,267	9,342	11,781	13,442	15,987	18,307	21,161	23,209
11	5,578	6,989	8,148	10,341	12,899	14,631	17,275	19,675	22,618	24,725
12	6,304	7,807	9,034	11,340	14,011	15,812	18,549	21,026	24,054	26,217
13	7,042	8,634	9,926	12,340	15,119	16,985	19,812	22,362	25,472	27,688
14	7,790	9,467	10,821	13,339	16,222	18,151	21,064	23,685	26,873	29,141
15	8,547	10,307	11,721	14,339	17,322	19,311	22,307	24,996	28,259	30,578
16	9,312	11,152	12,624	15,338	18,418	20,465	23,542	26,296	29,633	32,000
17	10,085	12,002	13,531	16,338	19,511	21,615	24,769	27,587	30,995	33,409
18	10,865	12,857	14,440	17,338	20,601	22,760	25,989	28,869	32,346	34,805
19	11,651	13,716	15,352	18,338	21,689	23,900	27,204	30,144	33,687	36,191
20	12,443	14,578	16,266	19,337	22,775	25,038	28,412	31,410	35,020	37,566
21	13,240	15,445	17,182	20,337	23,858	26,171	29,615	32,671	36,343	38,932
22	14,041	16,314	18,101	21,337	24,939	27,301	30,813	33,924	37,659	40,289
23	14,848	17,187	19,021	22,337	26,018	28,429	32,007	35,172	38,968	41,638
24	15,659	18,062	19,943	23,337	27,096	29,553	33,196	36,415	40,270	42,980
25	16,473	18,940	20,867	24,337	28,172	30,675	34,382	37,652	41,566	44,314
26	17,292	19,820	21,792	25,336	29,246	31,795	35,563	38,885	42,856	45,642
27	18,114	20,703	22,719	26,336	30,319	32,912	36,741	40,113	44,140	46,963
28	18,939	21,588	23,647	27,336	31,391	34,027	37,916	41,337	45,419	48,278
29	19,768	22,475	24,577	28,336	32,461	35,139	39,087	42,557	46,693	49,588
30	20,599	23,364	25,508	29,336	33,530	36,250	40,256	43,773	47,962	50,892

7.3.4 Complément : Test de rupture et de changement de tendance

Perron a proposé d'introduire, dès 1989, dans la régression de Dickey & Fuller une variable indicatrice spécifiant l'existence d'une rupture. La date de rupture peut d'ailleurs être connue ou inconnue. Dans le cas où elle est inconnue, une procédure séquentielle permet de la localiser. Ce lien entre les tests de racine unité et les changements de structure ont donné lieu à de nombreuses publications depuis une dizaine d'années. Différentes formes de changement de structure ont d'ailleurs été étudiées : changement de niveau, changement du coefficient de tendance linéaire, changement sur les coefficients des variables de la modélisation...etc.

Les tests de racine unité Plusieurs tests ont été implémentés afin de tester l'hypothèse nulle que la série stationnaire (Y_t) possède une racine unité et une constante, éventuellement nulle, avec une rupture au temps τ où $1 < \tau < T$, contre l'hypothèse alternative que la série soit stationnaire autour d'une tendance linéaire avec rupture en τ sur cette tendance. Une distinction est alors généralement apportée entre deux cas :

- *AO* - "additive outliers" - effet instantané
- *IO* - "innovational outlier" - effet avec transition

Pour chacun des effets, trois modèles sont alors considérés : dans la version *AO*

$$\left\{ \begin{array}{ll} (1) & X_t = \mu + \beta t + \gamma DU_t(\tau) + Y_t \quad \text{pour } t = 1, \dots, T, \\ (2) & X_t = \mu + \beta t + \theta DT_t^*(\tau) + Y_t \quad \text{pour } t = 1, \dots, T, \\ (3) & X_t = \mu + \beta t + \gamma DU_t(\tau) + \theta DT_t^*(\tau) + Y_t \quad \text{pour } t = 1, \dots, T, \end{array} \right.$$

où (Y_t) est la série (X_t) à laquelle on a retiré la tendance déterministe, avec $DU_t(\tau) = 1$ si $t > \tau$ et 0 sinon ($DU_t(\tau) = \mathbb{I}(t > \tau)$) et $DT_t^*(\tau) = (t - \tau)$ si $t > \tau$, 0 sinon ($DT_t^*(\tau) = [t - \tau] \cdot \mathbb{I}(t > \tau)$). La mise en oeuvre du test se fait en deux étapes :

- estimation (par une méthode de type moindres carrés) de la tendance avec les modèles de régression (1), (2) et (3), et calcul de la série résiduelle obtenue en retranchant à la série observée la tendance estimée
- pour les modèles (1) et (3), le test est basé sur la valeur de la t -statistique relative à $\alpha = 0$, noté $t_{\hat{\alpha}}(\tau)$ et correspond au test de racine unité dans la régression *ADF*

$$\Delta Y_t = \alpha Y_{t-1} + \sum_{j=0}^k d_j DTB_{t-j}(\tau) + \sum_{i=1}^k \alpha_i \Delta Y_{t-i} + \varepsilon_t \quad \text{où } DTB_{t-j}(\tau) = \mathbb{I}(t = \tau + 1).$$

Pour le modèle (2), la seconde étape consiste à effectuer la régression

$$\Delta Y_t = \alpha Y_{t-1} + \sum_{i=1}^k \alpha_i \Delta Y_{t-i} + \varepsilon_t,$$

et à utiliser la t -statistique $t_{\hat{\alpha}}(\tau)$ pour effectuer les tests classiques *ADF*.

Dans la version *IO*, les équations de régression s'écrivent

$$\begin{cases} (1) & \Delta X_t = \mu + \beta t + \theta DU_t(\tau) + \delta DTB_t(\tau) + \left[\alpha X_{t-1} + \sum_{i=1}^k c_i \Delta X_{t-i} + \varepsilon_t \right] & \text{pour } t = 1, \dots, T, \\ (2) & \Delta X_t = \mu + \beta t + \gamma DT_t^*(\tau) + \left[\alpha X_{t-1} + \sum_{i=1}^k c_i \Delta X_{t-i} + \varepsilon_t \right] & \text{pour } t = 1, \dots, T, \\ (3) & \Delta X_t = \mu + \beta t + \theta DU_t(\tau) + \delta DTB_t(\tau) + \gamma DT_t^*(\tau) + \left[\alpha X_{t-1} + \sum_{i=1}^k c_i \Delta X_{t-i} + \varepsilon_t \right] & \text{pour } t = 1, \dots, T, \end{cases}$$

où, encore une fois, $DTB_t(\tau) = \mathbb{I}(t = \tau + 1)$. Le test de Zivot et Andrews (1992) considère seulement les modèles de type *IO*, sans introduire l'indicatrice $DTB_{t-j}(\tau)$ (la "justification" étant que ce coefficient est asymptotiquement négligeable)²¹.

Toutefois, dans le cas général, la date de rupture τ est inconnue, ainsi que le paramètre k , permettant d'approcher le processus *ARMA* (p, q) par un processus *AR* $(k + 1)$. Différentes méthodes pour sélectionner k ont été proposées par Perron (1989 et 1993) pour les trois modèles et les deux types d'effets, *AO* et *IO*. Les procédures séquentielles d'identification de Zivot et Andrews (1992) et de Perron (1993) permettent de déterminer τ , ou plutôt $\lambda = \tau/T$. Cette méthode consiste à estimer les modèles de régressions (A), (B) et (C) dans les deux cas *AO* et *IO*, et à retenir le cas où $t_{\hat{\alpha}}(\tau) = t_{\hat{\alpha}}(\lambda)$ est minimal. Les auteurs ont étudié la distribution asymptotique de $\inf t_{\hat{\alpha}}(\lambda)$ quand λ appartient à un intervalle fermé de $]0, 1[$, $\Lambda = [3/20, 17/20]$ dans l'étude de Zivot et Andrews. On rejette alors l'hypothèse nulle de présence de racine unité si $\inf \{t_{\hat{\alpha}}(\lambda), \lambda \in \Lambda\}$ est plus petit que le fractile correspondant à une probabilité fixée de la distribution asymptotique de $\inf \{t_{\hat{\alpha}}(\lambda), \lambda \in \Lambda\}$.

Les tests de Gregory et Hansen (1996) Ces tests sont une généralisation des tests de Zivot et Andrews dans le cas multivarié, où $X_t = (X_t^1, X_t^2)$.

Les tests du "CUSUM" Ce test permet d'étudier la stabilité d'un modèle économétrique estimé au cours du temps. Il existe deux versions de ce test : le *CUSUM* fondé sur la somme cumulée des résidus récurrents, et le *CUSUMSQ* (*SQ pour 'square'*) fondé sur la somme cumulée des carrés des résidus récurrents. Pour cela, on note $(\tilde{\varepsilon}_t)$ le résidu normalisé par rapport à l'écart-type, c'est à dire $\tilde{\varepsilon}_t = \hat{\varepsilon}_t / \hat{\sigma}_\varepsilon$, et on note k le nombre de paramètres à estimer dans le modèles. Les statistiques S_t du *CUSUM* et S'_t du *CUSUMSQ* sont définies par

$$S_t = (T - k) \frac{\sum_{i=k+1}^t \tilde{\varepsilon}_i}{\sum_{i=k+1}^t \tilde{\varepsilon}_i^2} \text{ pour } t = k + 1, \dots, T,$$

et

$$S'_t = \frac{\sum_{i=k+1}^t \tilde{\varepsilon}_i^2}{\sum_{i=k+1}^T \tilde{\varepsilon}_i^2} \text{ pour } t = k + 1, \dots, T.$$

Si les coefficients sont variables au cours du temps, alors les résidus récurrents S_t doivent rester dans l'intervalle défini par

$$S_t \in \left[-\frac{\alpha(2t + T - 3k)}{\sqrt{T - k}}, +\frac{\alpha(2t + T - 3k)}{\sqrt{T - k}} \right],$$

où $\alpha = 1.143, 0.918$ ou 0.850 suivant que le seuil est 1%, 5% ou 10%. De la même façon, les résidus S'_t doivent appartenir à l'intervalle

$$S'_t \in \left[\frac{t - T}{T - k} - C, \frac{t - T}{T - k} + C \right],$$

où C est la constante du Durbin. En fait, on peut montrer que sous l'hypothèse de stabilité, l'espérance de S'_t est $\mathbb{E}(S'_t) = (t - T) / (T - k)$ allant de 0 à 1 quand t varie entre k et T . Plus précisément, la variable S'_t suit une loi Bêta.

Le test de Chow ou test d'homoscédasticité Puisque les bruits blancs doivent être homoscédastiques, le test de Chow, visant à comparer les variances des résidus sur des sous-périodes, peuvent être utilisés²².

²¹Ce test est évoqué ici car il existe des codes téléchargeables sur internet, en *EViews*, *SAS* ou *Gauss*.

²²Ce test n'est pas détaillé ici puisqu'il se trouve dans tous les cours d'économétrie. Pour mémoire, ce test est un test de Fisher : on considère un premier modèle $Y = X\beta_m + \varepsilon_m$ obtenu sur m observations, et un second modèle $Y = X\beta_n + \varepsilon_n$ obtenu sur n observations. Le test de Chow permet de tester l'égalité des coefficients : $\beta_m = \beta_n$, ainsi que $V(\varepsilon_m) = V(\varepsilon_n)$.

7.4 Estimation des paramètres d'un modèle $ARMA(p, q)$

A cette étape, les coefficients d, p et q ont été fixés. Il convient alors d'estimer les paramètres ϕ_i et θ_j du processus $ARIMA(p, d, q)$, ainsi que la volatilité σ^2 du bruit blanc. Sous l'hypothèse $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, on peut utiliser des méthodes du type maximum de vraisemblance. On supposera ici que le processus (X_t) est centré.

7.4.1 Estimation pour les modèles $AR(p)$ par la méthode des moindres carrés

Un modèle $AR(p)$ s'écrit

$$\begin{aligned} X_t &= c + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t \text{ où } (\varepsilon_t) \text{ est un bruit blanc,} \\ &= Z_t' \beta + \varepsilon_t \text{ où } Z_t' = (1, X_{t-1}, X_{t-2}, \dots, X_{t-p}) \text{ et } \beta' = (c, \phi_1, \phi_2, \dots, \phi_p). \end{aligned}$$

L'estimation des paramètres du modèle $X = Z' \beta + \varepsilon$ par la méthode des moindres carrés donne

$$\hat{\beta} = (ZZ')^{-1} ZX \text{ et } \hat{\sigma}^2 = \frac{1}{T - (p + 1)} \sum (X_t - Z_t' \hat{\beta})^2.$$

Toutefois, les résultats usuels d'économétries ne sont pas vérifiés ici, en particulier $\mathbb{E}(\hat{\beta}) \neq \beta$. Il est toutefois possible de montrer le résultat suivant,

Propriété 52 *Si les racines du polynôme caractéristique (racines de $\Phi(z) = 0$) sont à l'extérieur du disque unité alors*

$$\hat{\beta} \xrightarrow{\mathbb{P}} \beta \text{ et } \hat{\sigma}^2 \xrightarrow{\mathbb{P}} \sigma^2,$$

et de plus

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} N(0, \sigma^2 V) \text{ où } V = p \lim_{T \rightarrow \infty} \frac{1}{T} ZZ'.$$

Remarque 54 *Si la méthode des moindres carrés peut être utilisée pour estimer les paramètres d'un modèle $AR(p)$, elle ne marche plus dès lors que l'on a des termes autorégressifs sur les résidus.*

7.4.2 Vraisemblance d'un processus $ARMA(p, q)$

Pour déterminer la vraisemblance, il est nécessaire de supposer connue la loi des erreurs : nous supposons les erreurs normalement distribuées. Les erreurs étant normalement distribuées et indépendantes (*le processus (ε_t) est, par hypothèse un bruit blanc*), le vecteur $(\varepsilon_1, \dots, \varepsilon_n)$ est un vecteur gaussien. Les composantes du vecteur (X_1, \dots, X_n) étant obtenues par combinaisons linéaires des composantes du vecteur $(\varepsilon_1, \dots, \varepsilon_n)$, (X_1, \dots, X_n) sera un vecteur gaussien :

$$\mathcal{L}(X = (X_1, \dots, X_n)', \phi, \theta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \frac{1}{[\det \Sigma]^{1/2}} \exp\left(-\frac{1}{2\sigma^2} X' \Sigma^{-1} X\right),$$

où Σ est la matrice $(n \times n)$ des covariances du vecteur $X = (X_1, \dots, X_n)'$.

La maximisation, et même le calcul de cette vraisemblance étaient relativement difficile il y a quelques années, en particulier à cause du calcul de l'inverse Σ^{-1} , et du déterminant, de Σ , surtout lorsque n devenait relativement grand. Newbold a proposé une autre expression de cette vraisemblance, plus facile à calculer. Soit H la matrice triangulaire inférieure, à éléments positifs sur la diagonale telle que $HH' = \Sigma$ (*décomposition de Cholesky*). Soit alors e le vecteur tel que $e = H^{-1}X$. La log-vraisemblance du modèle s'écrit alors

$$\begin{aligned} \log \mathcal{L} &= -\frac{n}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 - \frac{1}{2} \log |\det \Sigma| - \frac{1}{2\sigma^2} X' \Sigma^{-1} X, \\ &= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log(e'e) - \log |\det H| = -\frac{n}{2} \log \left(|\det H|^{1/n} e'e |\det H|^{1/n} \right). \end{aligned}$$

La méthode du maximum de vraisemblance revient alors à chercher le minimum de $\ell = \left(|\det H|^{1/n} e'e |\det H|^{1/n} \right)$.

Une autre écriture, relativement proche est possible dans le cas des processus $MA(q)$. Soit ε^* le vecteur d'initialisation des erreurs,

$$\varepsilon^* = (\varepsilon_{1-q}, \dots, \varepsilon_{-1}, \varepsilon_0)'$$

permettant d'engendrer la série x_1, \dots, x_n . Considérons alors les vecteurs $\varepsilon = (\varepsilon_{1-q}, \dots, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \dots, \varepsilon_n)'$ et X . On peut alors écrire

$$\varepsilon = NX + M\varepsilon^*,$$

où M est une matrice $(n+q) \times q$ et N $(n+q) \times n$. L'initialisation des erreurs s'estimant par $\hat{\varepsilon}^* = -(M'M)^{-1} M'NX$, et en notant

$$S(\theta) = (NX + M\hat{\varepsilon}^*)'(NX + M\hat{\varepsilon}^*),$$

on peut alors montrer que la log-vraisemblance peut s'écrire

$$\log \mathcal{L} = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2} \log (\det (M'M)) - \frac{S(\theta)}{2\sigma^2}.$$

Et finalement, puisqu'on peut écrire $\sigma^2 = S(\theta)/n$, la fonction à minimiser s'écrit

$$\ell^* = n \log S(\theta) + \log (\det (M'M)).$$

Exemple 84 Dans le cas d'un modèle $AR(1)$, de la forme $X_t = c + \phi X_{t-1} + \varepsilon_t$ où ε_t est i.i.d. et distribué suivant une loi $\mathcal{N}(0, \sigma^2)$, avec $|\phi| < 1$, alors

$$X_t | X_{t-1} \sim \mathcal{N}(c + \phi X_{t-1}, \sigma^2).$$

Aussi, la loi conditionnelle de X_t est donnée par

$$f(x_t | x_{t-1}, (c, \phi, \sigma^2)) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} (x_t - c - \phi x_{t-1})^2 \right],$$

c'est à dire que

$$X_t \sim \mathcal{N}(\mathbb{E}(X_t), V(X_t)) \text{ soit } X_t \sim \mathcal{N}\left(\frac{c}{1-\phi}, \frac{\sigma^2}{1-\phi^2}\right).$$

En posant $\beta = (c, \theta, \sigma^2)$, la vraisemblance conditionnelle du modèle est alors donnée par

$$\mathcal{L}(\beta | X_1, \dots, X_T) = \prod_{t=2}^T \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} (X_t - c - \phi X_{t-1})^2 \right],$$

$$\log \mathcal{L}(\beta | X_1, \dots, X_T) = -\frac{T-1}{2} \ln(2\pi) - \frac{T-1}{2} \ln \sigma^2 + \frac{1}{2\sigma^2} \sum_{t=2}^T (X_t - c - \phi X_{t-1})^2.$$

La vraisemblance marginale s'écrit

$$\mathcal{L}(\beta, X_1) = \sqrt{\frac{1-\phi^2}{2\pi\sigma^2}} \exp \left[-\frac{(1-\phi^2)}{2\sigma^2} \left(X_1 - \frac{c}{1-\phi} \right)^2 \right],$$

on en déduit la forme de la log-vraisemblance (exacte, et non plus conditionnelle),

$$\begin{aligned} \log \mathcal{L}(\beta, X_1, \dots, X_T) &= -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln \left(\frac{\sigma^2}{1-\phi^2} \right) - \frac{(1-\phi^2)}{2\sigma^2} \left(X_1 - \frac{c}{1-\phi} \right)^2 \\ &\quad - \frac{T-1}{2} \ln(2\pi) - \frac{T-1}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=2}^T (X_t - c - \phi X_{t-1})^2. \end{aligned}$$

On peut noter que la maximisation de la vraisemblance exacte est un problème d'optimisation non-linéaire.

7.4.3 Résolution du programme d'optimisation

Une fois écrite la vraisemblance, deux méthodes sont alors possibles

- (1) des méthodes exactes, visant à minimiser effectivement la log-vraisemblance $\log \mathcal{L}$, de façon numérique
- (2) des méthodes de type moindres carrés, visant à minimiser la fonction $S(\theta)$ dans le cas MA , le second terme dans $\log \mathcal{L}_n$ devenant négligeable quand n augmente (*méthode utilisée sous EViews*).

Pour les modèles $ARMA$ stationnaires, les méthodes de maximisation de la vraisemblance conditionnelle, et de maximisation de la vraisemblance (*exacte*), sont asymptotiquement équivalentes. L'explication heuristique est que pour les modèles stationnaires, l'effet des valeurs initiales devient asymptotiquement négligeable, alors que dans le cas où des racines du polynôme caractéristique sont sur le cercle unité, les valeurs initiales influencent les échantillons finis.

Critère des moindres carrés conditionnel (MCC)

Exemple 85 *Considérons ici un modèle de la forme $MA(1)$, $X_t = \varepsilon_t - \theta\varepsilon_{t-1}$. Cette équation peut s'écrire $\varepsilon_t = x_t + \theta\varepsilon_{t-1}$, et donc, en supposant $\varepsilon_0 = 0$,*

$$\varepsilon_t = \sum_{i=0}^{t-1} \theta^i x_{t-i}, \text{ pour } t \geq 2, \quad (35)$$

et donc, la somme des carrés, conditionnelle à $\varepsilon_0 = 0$ s'écrit

$$S(\theta|\varepsilon_0 = 0) = \sum_{t=1}^T \varepsilon_t^2 = \sum_{t=1}^T \left[\sum_{i=0}^{t-1} \theta^i x_{t-i} \right]^2.$$

L'équation permettant d'estimer θ n'est pas linéaire.

Dans le cas général, pour un processus $ARMA(p, q)$, on suppose que $x_1 = \dots = x_p$ sont fixés et connus, et que $\varepsilon_p = \varepsilon_{p+1} = \dots = \varepsilon_{p+q} = 0$. Alors, par récurrence

$$\varepsilon_t = x_t - \sum_{i=1}^p \phi_i x_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j}.$$

La somme des carrés conditionnelle aux valeurs initiales s'écrit

$$S(\theta|\varepsilon_0 = 0) = \sum_{t=1}^T \varepsilon_t^2 = \sum_{t=1}^T \left[x_t - \sum_{i=1}^p \phi_i x_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \right]^2,$$

où les ε_{t-j} peuvent être écrits en fonction des $x_{t-j}, \dots, x_{t-j-p}$ et des $\varepsilon_{t-j-1}, \dots, \varepsilon_{t-q}$.

Critère des moindres carrés non conditionnel (MCN)

Exemple 86 *Considérons ici un modèle de la forme $MA(1)$, $X_t = \varepsilon_t - \theta\varepsilon_{t-1}$, que l'on notera, en considérant les innovations en temps inversé $\bar{\varepsilon}_t$, $X_t = \bar{\varepsilon}_t - \theta\bar{\varepsilon}_{t+1}$. On supposant $\bar{\varepsilon}_{T+1} = 0$, on détermine récursivement $\bar{\varepsilon}_T = x_T, \bar{\varepsilon}_{T-1} = x_T + \theta\bar{\varepsilon}_T \dots$ etc. De façon rétrospective, on peut ainsi définir $\hat{x}_0 = -\theta\bar{\varepsilon}_1$. De façon analogue à (35), on peut écrire*

$$\hat{x}_0 = - \sum_{t=1}^T \theta^t x_T.$$

En posant alors $\varepsilon_0 = \hat{x}_0$, on peut obtenir les ε_t en utilisant (35). On obtient alors une expression (non conditionnelle) de la somme des carrés des résidus

$$S(\theta) = \sum_{t=1}^T \varepsilon_t^2 = \sum_{t=1}^T \left[\sum_{i=0}^{t-1} \theta^i x_{t-i} - \theta^t \sum_{i=1}^T \theta^i x_i \right]^2.$$

Là encore, l'équation permettant d'estimer θ n'est pas linéaire.

Un des problèmes de cette méthode est que, dans le cas de processus comprenant une part autorégressive, les valeurs initiales doivent être obtenues, théoriquement, en $-\infty$. Ceci impose de faire une approximation sur la base d'un critère d'arrêt portant sur la convergence numérique de la récurrence.

Critère du maximum de vraisemblance conditionnelle (MV) Pour utiliser la méthode du maximum de vraisemblance, il est nécessaire de faire des hypothèses sur la loi des ε_t : ce sont des variables indépendantes, et de même loi $\mathcal{N}(0, \sigma^2)$. La vraisemblance conditionnelle est obtenue de la façon suivante :

La densité de $\varepsilon = (\varepsilon_1, \dots, \varepsilon_T)'$ est donnée par

$$f(\varepsilon_1, \dots, \varepsilon_T) = \frac{1}{(2\pi\sigma^2)^{T/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{t=1}^T \varepsilon_t^2\right).$$

On supposera connues les valeurs initiales x^* et ε^* . La densité de x peut s'exprimer conditionnellement à ε^* et x^* .

Programme d'optimisation Nous allons ici nous limiter à un cas simple, d'un modèle $MA(1)$, avec un critère de type MCC . On part d'une valeur initiale θ_0 , et on va mettre en place un algorithme convergent vers la vraie valeur θ . A la $i + 1$ -ème étape, on estime θ_{i+1} en fonction de θ_i en utilisant

$$S(\theta_{i+1}) = S(\theta_i) + [\theta_{i+1} - \theta_i]g(\theta^*) \text{ où } g(\theta^*) = \left. \frac{\partial S(\theta)}{\partial \theta} \right|_{\theta=\theta^*},$$

où θ^* est compris entre θ_i et θ_{i+1} . Aussi, on minimise la fonction $S(\theta)$ en choisant θ_{i+1} de telle sorte que θ_i soit de signe opposé au gradient de $S(\theta)$ en θ^* . Mais comme θ^* est inconnu, on choisit

$$\theta_{i+1} = \theta_i - \lambda g(\theta_i) \text{ avec } \lambda > 0$$

et ainsi, $S(\theta_{i+1}) < S(\theta_i)$. Le gradient, s'il est difficile à évaluer peut être remplacé par une différence de la forme $[S(\theta_i + \mu) - S(\theta_i)]/\mu$ avec μ petit. Ces deux constantes λ et μ , propres à l'algorithme, peuvent être fixées initialement, par exemple $\lambda = 0.001$ et $\mu = 0.01$.

Exemple 87 Considérons un cas relativement simple avec 6 observations (5, 6, 3, 2, 7, 6), et cherchons θ tel que $X_t = \varepsilon_t + \theta\varepsilon_{t-1}$.

$\theta_0 = 0$ et $\theta_0 + \mu = 0.01$. Alors $S(\theta_0) = 5^2 + 6^2 + \dots + 7^2 + 6^2 = 159$. Alors $X_1^{(i)} = 5$, $X_2^{(i)} = X_2 + (\theta_0 + \mu) \times X_1^{(i)} = 6 + 5 \times 0.01 = 6.05$, $X_3^{(i)} = X_3 + (\theta_0 + \mu) \times X_2^{(i)} = 3 + 6.05 \times 0.01 = 3.06$, ...etc. D'où la somme $S(\theta_0 + \mu) = 161.225$. Aussi, on obtient $g(\theta_0) = 222.458$ d'où finalement $\theta_1 = -0.222$.

Cet algorithme se répète à l'étape suivante, et les résultats sont alors

itération i	θ_i	1	2	3	4	5	6	$S(\theta_i)$	$g(\theta_i)$	θ_{i+1}	
0	θ_i	0.000	5.000	6.000	3.000	2.000	7.000	6.000	159.00	222.46	-0.222
	$\theta_i + \mu$	0.010	5.000	6.050	3.061	2.031	7.020	6.070	161.22		
1	θ_i	-0.222	5.000	4.888	1.913	1.575	6.650	4.521	119.68	302.02	-0.524
	$\theta_i + \mu$	-0.212	5.000	4.988	2.001	1.606	6.679	4.651	122.70		
2	θ_i	-0.524	5.000	3.378	1.229	1.356	6.289	2.702	86.61	181.06	-0.706
	$\theta_i + \mu$	-0.514	5.000	3.478	1.271	1.377	6.312	2.823	88.42		
3	θ_i	-0.706	5.000	2.472	1.256	1.114	6.214	1.616	75.16	115.23	-0.821
	θ_i	-0.696	5.000	2.572	1.271	1.146	6.223	1.742	76.31		
4	θ_i	-0.821	5.000	1.896	1.444	0.815	6.331	0.804	72.07	59.32	-0.880
	θ_i	-0.811	5.000	1.996	1.442	0.861	6.322	0.945	72.66		
5	θ_i	-0.880	5.000	1.600	1.592	0.599	6.473	0.303	72.44	19.73	-0.900
	θ_i	-0.870	5.000	1.700	1.582	0.654	6.451	0.457	72.64		
6	θ_i	-0.900	5.000	1.501	1.649	0.516	6.536	0.119	72.97	4.01	-0.905
	θ_i	-0.890	5.000	1.601	1.636	0.575	6.509	0.279	73.01		

En allant jusqu'à l'étape 10, on obtient $\theta = -0.905$.

Dans le cas des modèles moyennes mobiles (MA), l'algorithme du filtre de Kalman peut être utilisé, en considérant que ε_t (ou ε_{t-1}) est inobservable. La méthode d'estimation est alors la suivante : (Y_t) suit un modèle de la forme $Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$ où ε_t est *i.i.d.* et suit une loi $\mathcal{N}(0, \sigma^2)$, avec $|\theta| < 1$. La fonction de vraisemblance conditionnelle est $Y_t | \varepsilon_{t-1} \sim N(\mu + \theta\varepsilon_{t-1}, \sigma^2)$ et

$$f(y_t | \varepsilon_{t-1}, \theta, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2} [Y_t - \mu - \theta\varepsilon_{t-1}]^2 \right]$$

Le problème est que ε_{t-1} est "inobservable". Le raisonnement est alors le suivant :

- on suppose que $\varepsilon_0 = 0$, alors $Y_1 | \varepsilon_0 \sim \mathcal{N}(\mu, \sigma^2)$
- $Y_1 = \mu + \varepsilon_1 + \theta\varepsilon_0 = \mu + \varepsilon_1$ donc $\varepsilon_1 = Y_1 - \mu$
- $Y_2 = \mu + \varepsilon_2 + \theta\varepsilon_1$ donc $\varepsilon_2 = Y_2 - \mu - \theta(Y_1 - \mu)$
- ...
- $Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$ et donc

$$\varepsilon_t = (Y_t - \mu) - \theta(Y_{t-1} - \mu) + \dots + (-\theta)^{t-1} (Y_1 - \mu) \quad (36)$$

(on peut reconnaître la version tronquée de la représentation $AR(\infty)$ du processus $MA(1)$) La log vraisemblance conditionnelle est

$$-\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^T \varepsilon_t^2 \text{ où } \varepsilon_t \text{ est donnée par (36)}$$

Compléments : introduction au filtre de Kalman Un modèle **espace-état** est défini par le système d'équation

$$\begin{cases} Z_{t+1} = A_t Z_t + \varepsilon_t : \text{équation d'état} \\ Y_t = C_t Z_t + \eta_t : \text{équation de mesure} \end{cases} \quad \text{où } (\varepsilon_t, \eta_t)' \text{ est un bruit blanc normal}$$

où A_t et C_t sont déterministes, où $Z_0 \sim \mathcal{N}(m, p)$ est indépendant des $(\varepsilon_t, \eta_t)'$. Les variables peuvent ici être de dimension supérieure à 1. La matrice de variance-covariance

$$V \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} = \begin{bmatrix} V(\varepsilon_t) & \text{cov}(\varepsilon_t, \eta_t) \\ \text{cov}(\varepsilon_t, \eta_t) & V(\eta_t) \end{bmatrix}$$

On dira alors

$$\begin{cases} (Z_t) : \text{état du système à la date } t : \text{inobservable} \\ (Y_t) : \text{observations du système à la date } t : \text{observable} \\ (\varepsilon_t) : \text{innovations du système à la date } t : \text{inobservable} \\ (\eta_t) : \text{erreurs de mesure (ou bruit) en } t : \text{inobservable} \\ (A_t) : \text{matrice de transition} \\ (C_t) : \text{matrice de mesure} \\ (C_t Z_t) : \text{signal à la date } t \end{cases}$$

Le filtre de Kalman permet de calculer ${}_t \hat{Z}_t = \mathbb{E}(Z_t | Y_0, \dots, Y_t)$ la prévision de Z_t . On notera

$$\begin{cases} {}_t \Sigma_t = \mathbb{E} \left[(Z_t - {}_t \hat{Z}_t) (Z_t - {}_t \hat{Z}_t)' \right] : \text{erreur quadratique du filtre sur } Z_t \text{ en } t \\ {}_{t-1} \hat{Z}_t = \mathbb{E}(Z_t | Y_0, \dots, Y_{t-1}) : \text{prévision de } Z_t \text{ faite en } t-1 \\ {}_{t-1} \Sigma_t = \mathbb{E} \left[(Z_t - {}_{t-1} \hat{Z}_t) (Z_t - {}_{t-1} \hat{Z}_t)' \right] : \text{erreur quadratique moyenne de prévision} \end{cases}$$

Dans le cas où $\text{cov}(\varepsilon_t, \eta_t) = 0$, alors, pour tout $t \geq 0$, le **filtre de covariance**,

$$\begin{cases} (a) \quad {}_t \hat{Z}_t = {}_{t-1} \hat{Z}_t + K_t [Y_t - C_t \cdot {}_{t-1} \hat{Z}_t] \\ (b) \quad {}_t \hat{Z}_{t+1} = A_t \cdot {}_t \hat{Z}_t \end{cases} \quad \text{et} \quad \begin{cases} (a') \quad {}_t \Sigma_t = [I - K_t C_t] {}_{t-1} \Sigma_t \\ (b') \quad {}_t \Sigma_{t+1} = A_t \cdot {}_t \Sigma_t \cdot A_t' + Q \end{cases}$$

où

$$K_t = {}_{t-1} \Sigma_t \cdot C_t' (C_t \cdot {}_{t-1} \Sigma_t \cdot C_t' + R)^{-1} : \text{matrice de gain du filtre à la date } t$$

On peut alors en déduire directement les formules de calcul de prévisions de la variable observée : soit ${}_{t-1} \hat{Y}_t = \mathbb{E}(Y_t | Y_0, \dots, Y_{t-1})$ et ${}_{t-1} M_t = V(Y_t - {}_{t-1} \hat{Y}_t)$ alors

$$\begin{cases} (c) \quad {}_t \hat{Y}_{t+1} = C_{t+1} \cdot {}_t \hat{Z}_{t+1} \\ (c') \quad {}_t M_{t+1} = C_{t+1} \cdot {}_t \Sigma_{t+1} \cdot C_{t+1}' + R \end{cases}$$

Dans le cas dit "stationnaire", c'est à dire quand $A_t = A$ et $C_t = C$ alors le modèle se réécrit

$$\begin{cases} Z_{t+1} = AZ_t + \varepsilon_t \\ Y_t = CZ_t + \eta_t \end{cases}$$

Le filtre doit être initialisé, et on prend généralement ${}_{-1} \hat{Z}_0 = \mathbb{E}(Z_0) = m$ et ${}_{-1} \Sigma_0 = V(Z_0) = P$. De façon récursive, on peut alors calculer les ${}_t \hat{Z}_t$ à l'aide de (a) et (a'), puis ${}_t \Sigma_t$ et ${}_t \hat{Y}_{t+1}$ à l'aide de (b) et (b'), ainsi que de (c) et (c').

Remarque 55 Dans le cas où les bruits sont corrélés, des méthodes similaires peuvent être utilisées, en introduisant le résidu de la régression de (ε_t) sur (η_t) .

Pour une prévision à l'ordre h , on introduit une seconde itération : on cherche

$$\begin{cases} {}_t \hat{Y}_{t+h} = \mathbb{E}(Y_{t+h} | Y_0, \dots, Y_t) \\ {}_t \hat{Z}_{t+h} = \mathbb{E}(Z_{t+h} | Y_0, \dots, Y_t) \end{cases} \quad \text{et} \quad \begin{cases} {}_t M_{t+h} = V \left({}_t \hat{Y}_{t+h} - Y_{t+h} \right) \\ {}_t \Sigma_{t+h} = V \left({}_t \hat{Z}_{t+h} - Z_{t+h} \right) \end{cases}$$

(pour $h = 1$ on retrouve le filtre de covariance). Dans le cas où $\text{cov}(\varepsilon_t, \eta_t) = 0$, on a les formules de récurrence

$$\begin{cases} {}_t\widehat{Y}_{t+h} = C_{t+h} \cdot {}_t\widehat{Z}_{t+h} \\ {}_t\widehat{Z}_{t+h} = A_{t+h-1} \cdot {}_t\widehat{Z}_{t+h-1} \end{cases} \quad \text{et} \quad \begin{cases} {}_tM_{t+h} = C_{t+h} \cdot {}_t\Sigma_{t+h} \cdot C'_{t+h} + R \\ {}_t\Sigma_{t+h} = A_{t+h+1} \cdot {}_t\Sigma_{t+h-1} \cdot A_{t+h-1} + Q \end{cases}$$

La procédure itératif prend alors la forme suivante :

- (i) initialisation : $t = 0$, $h = 1$, on pose ${}_{-1}\widehat{Z}_0 = \mathbb{E}(Z_0) = m$ et ${}_{-1}\Sigma_0 = V(Z_0) = P$
- (ii) formules (a) et (a')
- (iii) formules (b) et (b'), et (c) et (c')
- (iv) si $h < H$ (horizon de prévision), alors $h = h + 1$ et (iii), sinon (v)
- (v) si $t < T$ alors $t = t + 1$ et $h = 1$, observation de Y_{t+1} , et (ii), sinon fin

Remarque 56 De façon analogue, on peut utiliser le filtre d'information, basé sur ${}_t\widehat{U}_t = {}_t\Sigma_t^{-1} \cdot {}_t\widehat{Z}_t$ et ${}_t\widehat{U}_{t+1} = {}_t\Sigma_{t+1}^{-1} \cdot {}_t\widehat{Z}_{t+1}$, et on alors les relations

$$\begin{cases} (\alpha) & {}_t\widehat{U}_t = {}_{t-1}\widehat{U}_t + C'_t R^{-1} Y_t \\ (\beta) & {}_t\widehat{U}_{t+1} = [\mathbb{I} - M_t] A_t'^{-1} \cdot {}_t\widehat{U}_t \end{cases} \quad \text{et} \quad \begin{cases} (\alpha') & {}_t\Sigma_t^{-1} = {}_{t-1}\Sigma_t^{-1} + C'_t R^{-1} C_t \\ (\beta') & {}_t\Sigma_{t+1}^{-1} = [\mathbb{I} - M_t] N_t \end{cases}$$

où $M_t = N_t (N_t + Q^{-1})^{-1}$ et $N_t = A_t'^{-1} \cdot {}_t\Sigma_t^{-1} \cdot A_t^{-1}$.

7.4.4 Tests statistiques de validation du modèle

Après avoir estimé les paramètres p et q d'un modèle *ARMA*, il convient de vérifier que les polynômes *AR* et *MA* ne possèdent pas de racine commune. Lorsque c'est le cas, il y a redondance, ce qui peut conduire à des erreurs lors des prévisions. Il convient alors d'estimer les paramètres processus *ARMA* avec moins de retards (ou d'autres types de retards).

Comme lors d'une regression linéaire, un certain nombre d'indicateurs sont intéressants. Par exemple le test de Student des paramètres permet de vérifier que les paramètres sont bien significatifs.

Il convient ensuite de vérifier que le processus ε_t est effectivement un bruit blanc. Par exemple, pour vérifier que la moyenne est nulle, on compare la moyenne $\bar{\varepsilon}$ à $t_{\alpha/2} \widehat{\sigma} / \sqrt{n - p - q}$ dans le cas d'un processus $p + q$. Pour tester l'absence d'autocorrélation de ε_t , il est possible d'utiliser la statistique de Box & Pierce (Q) ou la statistique de Ljung & Box (Q') définies par

$$Q(k) = n \sum_{i=1}^k r_i^2 \quad \text{et} \quad Q'(k) = n(n+2) \sum_{i=1}^k \frac{r_i^2}{n-i},$$

qui sont à comparer aux quantiles du chi-deux à $k - (p + q)$ degrés de liberté (l'hypothèse H_0 testée étant $\rho(1) = \dots = \rho(h) = 0$).

7.5 Choix d'un modèle

7.5.1 Critère de pouvoir prédictif

Comme nous le verrons par la suite, dans un modèle *ARMA*, l'erreur de prévision à horizon 1 dépend de la variance du résidu. On peut alors choisir le modèle conduisant à la plus petite erreur de prévision. Plusieurs indicateurs sont alors possibles :

- (i) la variance du résidu σ^2 , ou la somme des carrés des résidus *SCR*
 - (ii) le coefficient de détermination R^2 , correspondant à une normalisation de la variance
 - (iii) le coefficient de détermination modifié \overline{R}^2
 - (iv) la statistique de Fisher (comme dans le cas du modèle linéaire)
- Le but est alors de minimiser (i), ou de maximiser (ii), (iii) ou (iv).

Exemple 88 Dans l'exemple ci-dessous, considérons les 2 modèles suivants : un modèle *ARMA* (1,1) à gauche, ou

un modèle AR(4), à droite

LS // Dependent Variable is X Sample: 2 5000 Included observations: 4999 after adjusting endpoints Convergence achieved after 4 iterations					LS // Dependent Variable is X Sample: 5 5000 Included observations: 4996 after adjusting endpoints Convergence achieved after 2 iterations				
Variable	Coefficient	Std. Error	T-Statistic	Prob.	Variable	Coefficient	Std. Error	T-Statistic	Prob.
AR(1)	0.767134	0.019055	40.25967	0.0000	AR(1)	0.302722	0.014141	21.40765	0.0000
MA(1)	-0.463423	0.026398	-17.55536	0.0000	AR(2)	0.143773	0.014739	9.754472	0.0000
					AR(3)	0.066880	0.014740	4.537428	0.0000
					AR(4)	0.041655	0.014135	2.946946	0.0032
R-squared	0.183040	Mean dependent var	0.014388		R-squared	0.183303	Mean dependent var	0.014946	
Adjusted R-squared	0.182877	S.D. dependent var	1.115828		Adjusted R-squared	0.182812	S.D. dependent var	1.115693	
S.E. of regression	1.008651	Akaike infocriterion	0.017628		S.E. of regression	1.008569	Akaike infocriterion	0.017865	
Sum squared resid	5083.836	Schwartz criterion	0.020236		Sum squared resid	5077.916	Schwartz criterion	0.023082	
Log likelihood	-7135.336	F-statistic	1119.579		Log likelihood	-7129.642	F-statistic	373.4759	
Durbin-Watson stat	2.002189	Prob(F-statistic)	0.000000		Durbin-Watson stat	1.999789	Prob(F-statistic)	0.000000	
Inverted AR Roots	.77				Inverted AR Roots	.73	-.02+.39i	-.02-.39i	-.38
Inverted MA Roots	.46								

soit

$$\begin{cases} [1] : X_t = 0.767 X_{t-1} + \varepsilon_t - 0.463 \varepsilon_{t-1} \\ \quad \quad \quad (0.019) \quad \quad \quad (0.026) \\ [2] : X_t = 0.303 X_{t-1} + 0.226 X_{t-2} + 0.162 X_{t-3} + 0.116 X_{t-4} \\ \quad \quad \quad (0.014) \quad \quad \quad (0.014) \quad \quad \quad (0.014) \quad \quad \quad (0.014) \end{cases}$$

Nous obtenons les indicateurs suivants

	σ^2	R^2	\overline{R}^2	F-stat
[1]	1.01737	0.18304	0.18287	1119.579
[2]	1.02751	0.17505	0.17455	353.3722

Le modèle [1] semble meilleur que le modèle [2] : la variance du résidu est plus faible, mais de plus, les trois autres indicateurs sont plus élevés dans le premier cas que dans le second.

7.5.2 Critère d'information

Cette approche a été introduite par Akaike en 1969. Cette mesure de l'écart entre le modèle proposé et la vraie loi peut être obtenue à l'aide de la quantité d'information de Kullback.

Définition 46 Soit f_0 la densité inconnue d'observations, et $\{f(\cdot), f \in \mathcal{F}\}$ la famille des densités parmi lesquelles ont fait l'estimation. L'écart entre la vraie loi et le modèle est donné par

$$I(f_0, \mathcal{F}) = \min_{f \in \mathcal{F}} \int \log \frac{f_0(x)}{f(x)} \cdot f_0(x) dx$$

Cette quantité est toujours positive, et ne s'annule que si f_0 appartient à \mathcal{F} . Cette mesure étant inconnue puisque f_0 est inconnue, on essaiera de minimiser un estimateur de I, \hat{I} . Plusieurs estimateur de la quantité d'information ont été proposé, dans le cas de modèles ARMA(p, q), à partir de T observations,

(i) **Aikaïke** (1969) :

$$AIC(p, q) = \log \hat{\sigma}^2 + 2 \frac{p+q}{T}$$

(ii) **Schwarz** (1977) :

$$BIC(p, q) = \log \hat{\sigma}^2 + [p+q] \frac{\log T}{T}$$

(iii) **Hanna-Quinn** (1979) :

$$\phi(p, q) = \log \hat{\sigma}^2 + [p+q] c \frac{\log(\log T)}{T} \text{ avec } c > 2$$

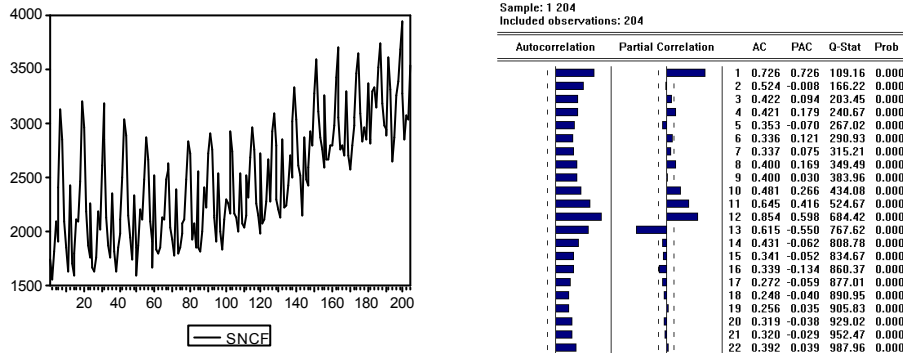
Exemple 89 En reprenant l'exemple précédant un critère d'Aikaïke (AIC sous EViews) de 0.017628 pour le modèle ARMA(1,1) contre 0.027968 pour le modèle AR(4). Ici encore, le modèle ARMA est préféré au modèle AR.

7.6 Application

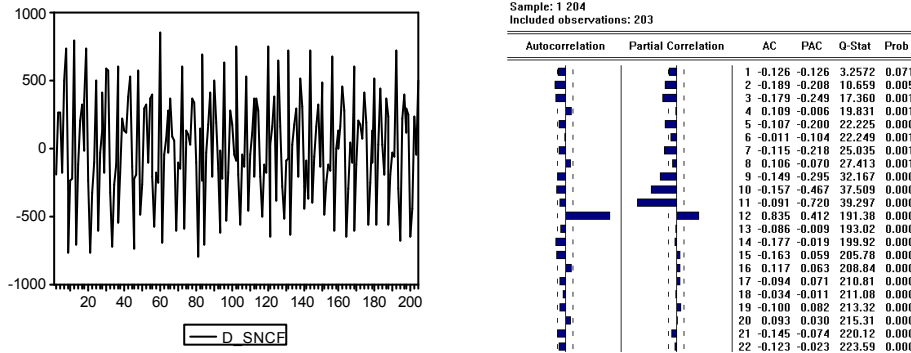
Nous allons essayer ici de modéliser la série mensuelle du nombre de voyageurs SNCF.

7.6.1 Identification du modèle : recherche des paramètres d , p et q

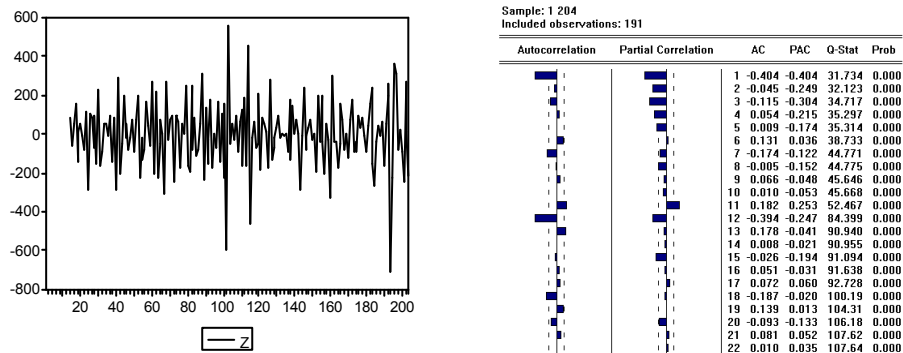
La série, composée de 204 observations peut être représentée par



Compte tenu de la significativité des premières autocorrélations (ou tout du moins le fait qu'elles sont significativement non-nulles pour les 40 premiers retards) suggère de différencier au moins un fois la série,



La série $Y_t = (1 - L)X_t$ présente alors de fortes corrélations pour les retards multiples de 12 (nous retrouvons ici la saisonnalité annuelle longuement développée dans les premières parties)



La série $Z_t = (1 - L^2)Y_t = (1 - L)(1 - L^2)X_t$ semble cette fois-ci stationnaire. Néanmoins, la présence de fortes valeurs pour $\rho(1)$ et $\rho(12)$ suggère d'introduire une moyenne mobile de la forme $(1 - \theta_1 L)(1 - \theta_2 L^{12})$. Ce type de représentation est confirmé par la forme de l'autocorrélogramme partiel : une modélisation de type AR nécessiterait d'introduire un trop grand nombre de termes (les 5 premières valeurs de l'autocorrélogramme partiel étant significativement non-nulles). De plus, la moyenne mobile $\Theta(L) = (1 - \theta_1 L)(1 - \theta_2 L^{12})$ s'écrit

$$\Theta(L)\varepsilon_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-12} + \theta_1\theta_2\varepsilon_{t-13}$$

admettant des autocorrélations $\rho(h)$ non nulles pour $h = 1, 11, 12, 13$ (ce qui est confirmé par le graphique des autocorrélations).

Les trois paramètres à estimer sont θ_1, θ_2 et σ^2 . Une méthode basée sur les moindres carrés permet d'estimer les 3 paramètres de $\varepsilon_t - \lambda\varepsilon_{t-1} - \mu\varepsilon_{t-2} + \nu\varepsilon_{t-3}$:

LS // Dependent Variableis Z Sample: 14 204 Included observations: 191 after adjusting endpoints Convergence achieved after 8 iterations					Conditional Least Squares Estimation											
					Approx.											
					Parameter	Estimate	Std Error	T Ratio	Lag							
					MA1,1	0.83438	0.04023	20.74	1							
					MA2,1	0.49261	0.06872	7.17	12							
					Variance Estimate = 15407.0492											
					Std Error Estimate = 124.125135											
					AIC = 2385.75684*											
					SBC = 2392.26138*											
					Number of Residuals = 191											
					* Does not include log determinant.											
					Correlations of the Estimates											
							Parameter	MA1,1	MA2,1							
							MA1,1	1.000	-0.030							
							MA2,1	-0.030	1.000							
					Autocorrelation Check of Residuals											
					To	Chi	Autocorrelations									
					Lag	Square	DF	Prob								
					6	7.49	4	0.112	0.148	-0.036	-0.118	0.022	-0.017	0.024		
					12	13.95	10	0.175	-0.157	-0.062	0.029	0.039	0.022	-0.029		
					18	19.61	16	0.238	0.075	0.054	-0.005	0.042	-0.003	-0.128		
					24	22.54	22	0.428	-0.040	-0.085	-0.022	-0.049	-0.023	0.035		
					30	26.98	28	0.520	0.088	0.033	0.023	-0.037	0.006	0.094		
					36	34.07	34	0.465	0.031	-0.004	-0.074	-0.006	0.150	0.034		

Variable	Coefficient	Std. Error	T-Statistic	Prob.
MA(1)	-0.768737	0.046718	-16.45476	0.0000
MA(12)	-0.475989	0.062662	-7.596094	0.0000
MA(13)	0.415363	0.060472	6.868676	0.0000
R-squared	0.471378	Mean dependent var	-0.157068	
Adjusted R-squared	0.465755	S.D. dependent var	169.2716	
S.E. of regression	123.7241	Akaike info criterion	9.651690	
Sum squared resid	2877839.	Schwartz criterion	9.702773	
Log likelihood	-1189.754	F-statistic	83.82096	
Durbin-Watson stat	1.792060	Prob(F-statistic)	0.000000	

Inverted MA Roots	.88 -.06i	.88+.06i	.81+.48i	.81 -.48i
	.46 -.82i	.46+.82i	-.00+.94i	-.00 -.94i
	-.47+.82i	-.47 -.82i	-.82 -.47i	-.82+.47i
	-.94			

Toutefois, cette estimation (à gauche, sous *EViews*) ne permet pas d'intégrer la contrainte $\nu = \lambda\mu$. Toutefois, la procédure ARIMA sous *SAS* permet de prendre en compte ce genre de modèle (à droite).

Remarque 57 La procédure ARIMA intègre une déclaration ESTIMATE permettant d'estimer des paramètres d'un modèle ARIMA(p, d, q). Parmi les options de cette déclaration, il est possible de choisir les options suivantes, P et Q. Par exemple, en demandant P=(1,2,5) (6,12) on peut estimer les paramètres $\alpha, \beta, \gamma, \lambda$ et μ pour un modèle AR de la forme

$$(1 - \alpha L - \beta L^2 - \gamma L^5) (1 - \lambda L^6 - \mu L^{12}) = \varepsilon_t$$

Remarque 58 Il est nécessaire d'utiliser sous SAS l'option *noconstant* de la déclaration ESTIMATE afin de ne pas avoir de constante dans la régression.

D'où finalement le modèle,

$$(1 - L) (1 - L^{12}) X_t = \begin{pmatrix} 1 - 0.8344L \\ (0.0402) \end{pmatrix} \begin{pmatrix} 1 - 0.4926L^{12} \\ (0.0687) \end{pmatrix} \varepsilon_t$$

$$X_t = X_{t-1} + X_{t-12} - X_{t-13} + \varepsilon_t - 0.8344\varepsilon_{t-1} - 0.4926\varepsilon_{t-12} + 0.4110\varepsilon_{t-13}$$

7.6.3 Vérification

On peut tout d'abord noter que les rapports de Student des 2 paramètres θ_1 et θ_2 sont respectivement 21 et 7 (≥ 1.96) : ce deux coefficients sont significatifs.

Comme le montre la sortie SAS présentée auparavant, pour les seuils $h = 12, 24$ et 36 , les statistiques $Q(h)$ valaient respectivement 13.95, 22.54 et 34.07, alors que les quantiles de la loi du chi-deux à 5% sont respectivement (pour 10, 22 et 34 degrés de liberté) 18, 34 et 46. Puisque $Q_h \leq \chi_{1-\alpha}$, on accepte l'hypothèse de test de bruit blanc sur les résidus.

Toutefois, il serait bien sûr possible d'améliorer le modèle. En particulier, on peut noter que les résidus présentent des pics au niveau de l'autocorrélogramme pour les h multiples de 6.



(Source: Les Formidables Aventures de Lapinot, Blacktown, de Lewis Trondheim, Dargaud, 2000)

” Dans toute statistique, l'inexactitude du nombre est compensée par la précision des décimales.”
(Alfred Sauvy)

8 Prévisions à l'aide des modèles *ARIMA* : Box-Jenkins

Etant donnée une série stationnaire (X_t) , observée entre 1 et T , on cherche à faire de la prévision à horizon h , et donc à prévoir X_{T+1}, \dots, X_{T+h} . Tous les processus *AR*, *MA* et *ARMA* seront supposés mis sous forme canonique, et n'avoir aucune racine unité. Aussi, toutes les racines des polynômes autorégressifs Φ et des polynômes moyennes-mobiles Θ auront leurs racines à l'extérieur du disque unité. Ainsi, pour tous les processus X_t tels que $\Phi(L) X_t = \Theta(L) \varepsilon_t$, ε_t sera l'innovation du processus X_t .

8.1 Prévisions à l'aide d'un modèle *AR*(p)

Le modèle s'écrit, quite à recentrer le processus,

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t \text{ ou } \Phi(L) X_t = \varepsilon_t$$

La prévision optimale à la date $T+1$, faite à la date T est ${}_T X_{T+1}^* = EL(X_{T+1}|X_T, X_{T-1}, \dots)$. Aussi,

$${}_T X_{T+1}^* = \phi_1 X_T + \dots + \phi_p X_{T-p}$$

car (ε_t) est l'innovation. De façon analogue, $X_{T+h} = \phi_1 X_{T+h-1} + \dots + \phi_p X_{T+h-p} + \varepsilon_{T+h}$, et donc ${}_T X_{T+h}^* = EL(X_{T+h}|X_T, X_{T-1}, \dots)$ est donné, de façon récursive par

$${}_T X_{T+h}^* = \begin{cases} \phi_1 \cdot {}_T X_{T+h-1}^* + \dots + \phi_{h-1} \cdot {}_T X_{T+1}^* + \phi_h X_T + \dots + \phi_p X_{T+h-p} & \text{pour } h \leq p \\ \phi_1 \cdot {}_T X_{T+h-1}^* + \dots + \phi_p \cdot {}_T X_{T+h-p}^* & \text{pour } h > p \end{cases}$$

Exemple 90 Dans le cas d'un processus *AR*(1) tel que $X_t = \mu + \phi X_{t-1} + \varepsilon_t$ alors

$$(i) \quad {}_T X_{T+1}^* = \mu + \phi X_T,$$

$$(ii) \quad {}_T X_{T+2}^* = \mu + \phi \cdot {}_T X_{T+1}^* = \mu + \phi [\mu + \phi X_T] = \mu [1 + \phi] + \phi^2 X_T,$$

$$(iii) \quad {}_T X_{T+3}^* = \mu + \phi \cdot {}_T X_{T+2}^* = \mu + \phi [\mu + \phi [\mu + \phi X_T]] = \mu [1 + \phi + \phi^2] + \phi^3 X_T,$$

et récursivement, on peut obtenir ${}_T X_{T+h}^*$ de la forme

$${}_T X_{T+h}^* = \mu + \phi \cdot {}_T X_{T+h-1}^* = \mu [1 + \phi + \phi^2 + \dots + \phi^{h-1}] + \phi^h X_T.$$

Exemple 91 Une méthode alternative est de considérer le processus centré $Y_t = X_t - \mu/\phi$, alors $Y_t = \phi Y_{t-1} + \varepsilon_t$. Alors de façon récursive ${}_T Y_{T+h}^* = \phi \cdot {}_T Y_{T+h-1}^*$, et donc ${}_T Y_{T+h}^* = \phi^h Y_T$. Aussi, on peut écrire

$${}_T X_{T+h}^* = \frac{\mu}{\phi} + \phi^h \left[X_T - \frac{\mu}{\phi} \right] = \mu \underbrace{\frac{1 - \phi^h}{1 - \phi}}_{1 + \phi + \phi^2 + \dots + \phi^{h-1}} + \phi^h X_T.$$

8.2 Prévisions à l'aide d'un modèle *MA*(q)

On supposera là aussi que l'on s'est ramené à un processus centré (X_t) , satisfaisant

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} = \Theta(L) \varepsilon_t.$$

La prévision optimale à la date $T+1$, faite à la date T est ${}_T X_{T+1}^* = EL(X_{T+1}|X_T, X_{T-1}, \dots) = EL(X_{T+1}|\varepsilon_T, \varepsilon_{T-1}, \dots)$ car (ε_t) est le processus d'innovation. Aussi,

$${}_T X_{T+1}^* = 0 + \theta_1 \varepsilon_T + \dots + \theta_q \varepsilon_{T+1-q}$$

De façon analogue, X_{T+h} est estimé par ${}_T X_{T+h}^* = EL(X_{T+h}|X_T, X_{T-1}, \dots) = EL(X_{T+h}|\varepsilon_T, \varepsilon_{T-1}, \dots)$, et donc

$${}_T X_{T+h}^* = \begin{cases} \theta_h \cdot {}_T X_T^* + \dots + \theta_q X_{T+h-q} & \text{pour } h \leq q \\ 0 & \text{pour } h > q. \end{cases} \quad (37)$$

Toutefois, cette méthode présente le désavantage d'estimer X_{T+h} à partir des résidus passés, a priori non observables, et non pas du passé de la variable.

8.2.1 Utilisation de l'écriture $AR(\infty)$ du processus $MA(q)$

L'équation $X_t = \Theta(L) \varepsilon_t$ peut se réécrire $\Theta^{-1}(L) X_t = \varepsilon_t$, soit

$$X_t = \sum_{k=1}^{\infty} a_k X_{t-k} + \varepsilon_t \text{ et donc } X_{t+h} = \sum_{k=1}^{\infty} a_k X_{t+h-k} + \varepsilon_{t+h} \text{ pour tout } h \geq 0$$

Aussi, ${}_T X_{T+h}^*$ peut être écrit de façon itérative

$${}_T X_{T+h}^* = \sum_{k=1}^{h-1} a_k \cdot {}_T X_{T+h-k}^* + \sum_{k=h}^{\infty} a_k X_{T+h-k}$$

Toutefois, un des problèmes est que les (X_t) ne sont pas observés, en pratique, pour $t < 0$. On utilise alors l'écriture suivante

$$X_{T+h} = \sum_{k=1}^{\infty} a_k X_{T+h-k} + \varepsilon_{T+h} = \sum_{k=1}^{h-1} a_k X_{T+h-k} + \underbrace{\sum_{k=h}^{\infty} a_k X_{T+h-k}}_{\text{Reste d'une série ACV}} + \varepsilon_{T+h},$$

où le reste de la série absolument convergente tend (*au sens de L^2*) vers 0 quand $T \rightarrow \infty$. On peut alors considérer, quand T est suffisamment grand que

$${}_T X_{T+h}^* = \sum_{k=1}^{h-1} a_k \cdot {}_T X_{T+h-k}^* + \sum_{k=h}^{T+h} a_k X_{T+h-k} + \underbrace{\sum_{k=T+h+1}^{\infty} a_k X_{T+h-k}}_{\text{Négligeable (hyp.)}}$$

et on approxime ${}_T X_{T+h}^*$ par ${}_T \widehat{X}_{T+h}^*$

$${}_T \widehat{X}_{T+h}^* = \sum_{k=1}^{h-1} a_k \cdot {}_T \widehat{X}_{T+h-k}^* + \sum_{k=h}^{T+h} a_k X_{T+h-k}.$$

8.2.2 Utilisation de la formule de mise à jour des résultats

L'idée est ici de comparer l'estimation faite en T de X_{T+1} , ${}_T X_{T+1}^*$ avec la vraie valeur :

$$X_{T+1} - {}_T X_{T+1}^* = \varepsilon_{T+1}.$$

La relation (37) permet d'obtenir une estimation, à la date T de X_{T+h} . En se plaçant à la date $T+1$, on peut noter que ${}_{T+1} X_{T+h}^* = {}_{T+1} X_{(T+1)+(h-1)}^*$ et donc

$${}_{T+1} X_{T+h}^* = \begin{cases} \theta_{h-1} \varepsilon_{T+1} + {}_T X_{T+h}^* & \text{pour } h \leq q+1 \\ 0 & \text{pour } h > q+1, \end{cases}$$

et donc, pour $h \leq q+1$

$${}_{T+1} X_{T+h}^* = {}_T X_{T+h}^* + \theta_{h-1} [X_{T+1} - {}_T X_{T+1}^*] : \text{Formule de mise à jour}$$

8.3 Prévisions à l'aide d'un modèle $ARMA(p, q)$

On supposera là aussi que l'on s'est ramené à un processus centré (X_t) , satisfaisant

$$\Phi(L) X_t = \Theta(L) \varepsilon_t$$

Remarque 59 Dans le cas d'un processus non centré, $\Phi(L) X_t = \mu + \Theta(L) \varepsilon_t$, on peut noter que $\mathbb{E}X_t = \mu/\Phi(1) = m$, et que $\Phi(L)(X_t - m) = \Theta(L) \varepsilon_t$. Il est donc toujours possible, en translatant le processus, de se ramener à un processus centré (pour des processus $ARMA$ seulement).

Sous cette forme *ARMA*, alors

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \text{ et donc } X_{t+h} = \sum_{i=1}^p \phi_i X_{t+h-i} + \varepsilon_{t+h} + \sum_{j=1}^q \theta_j \varepsilon_{t+h-j}.$$

On a alors ${}_T X_{T+h}^* = EL(X_{T+h} | X_T, X_{T-1}, \dots) = EL(X_{T+h} | \varepsilon_T, \varepsilon_{T-1}, \dots)$ car ε_t est le processus d'innovation. On peut noter que pour $h > q$

$${}_{T+1} X_{T+h}^* = \begin{cases} \phi_1 \cdot {}_T X_{T+h-1}^* + \dots + \phi_{h-1} \cdot {}_T X_{T+1}^* + \phi_h X_T + \dots + \phi_p X_{T+h-p} \text{ pour } h \leq p \\ \phi_1 \cdot {}_T X_{T+h-1}^* + \dots + \phi_p \cdot {}_T X_{T+h-p}^* \text{ pour } h > p. \end{cases}$$

La forme générale des solutions est connue (*comme dans le cas des AR(p)*). Toutefois, il y a ici un problème d'initialisation des calculs.

8.3.1 Utilisation de la forme $AR(\infty)$ pu processus $ARMA(p, q)$

L'équation $\Phi(L) X_t = \Theta(L) \varepsilon_t$ peut se réécrire $\Theta^{-1}(L) \Phi(L) X_t = \varepsilon_t$, c'est à dire, comme dans le cas des processus $MA(q)$,

$$X_t = \sum_{k=1}^{\infty} a_k X_{t-k} + \varepsilon_t \text{ et donc } X_{t+h} = \sum_{k=1}^{\infty} a_k X_{t+h-k} + \varepsilon_{t+h} \text{ pour tout } h \geq 0,$$

et de la même façon que pour la modélisation $AR(\infty)$ des processus $MA(q)$, on peut réécrire

$$X_{T+h} = \sum_{k=1}^{\infty} a_k X_{T+h-k} + \varepsilon_{T+h} = \sum_{k=1}^{h-1} a_k X_{T+h-k} + \underbrace{\sum_{k=h}^{\infty} a_k X_{T+h-k}}_{\text{négligeable dans } L^2} + \varepsilon_{T+h},$$

d'où la forme itérative, obtenue par approximation, en ne tenant pas compte du second terme, négligeable dans L^2 ,

$${}_T \widehat{X}_{T+h}^* = \sum_{k=1}^{h-1} a_k \cdot {}_T \widehat{X}_{T+h-k}^* + \sum_{k=h}^{T+h} a_k X_{T+h-k}.$$

8.3.2 Utilisation de la forme $MA(\infty)$ pu processus $ARMA(p, q)$ et des formules de mise à jour

L'équation $\Phi(L) X_t = \Theta(L) \varepsilon_t$ peut se réécrire $X_t = \Phi(L)^{-1} \Theta(L) \varepsilon_t$ soit encore

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \dots = \varepsilon_t + \sum_{j=1}^{\infty} b_j \varepsilon_{t-j}.$$

Puisque²³ $\overline{L^2}(X_t, X_{t-1}, \dots) = \overline{L^2}(\varepsilon_t, \varepsilon_{t-1}, \dots)$, on peut écrire ${}_T X_{T+h}^* = EL(X_{T+h} | X_T, X_{T-1}, \dots) = EL(X_{T+h} | \varepsilon_T, \varepsilon_{T-1}, \dots)$ soit

$${}_T X_{T+h}^* = \sum_{j=h}^{\infty} b_j \varepsilon_{t+h-j}.$$

De plus, puisque ${}_{T+1} X_{T+h}^* = {}_{T+1} X_{(T+1)+(h-1)}^*$, on peut écrire

$${}_{T+1} X_{T+h}^* = {}_T X_{T+h}^* + b_{h-1} [X_{T+1} - {}_T X_{T+1}^*] : \text{Formule de mise à jour.}$$

8.4 Prévisions dans le cas d'un processus $ARIMA(p, d, q)$

On considérons ici (X_t) satisfaisant une équation de la forme $\Phi(L)(1-L)^d X_t = \Theta(L) \varepsilon_t$ avec les conditions initiales

$$Z = (X_{-1}, \dots, X_{-p-d}, \varepsilon_{-1}, \dots, \varepsilon_{-q})'.$$

Posons alors $\Psi(L) = \Phi(L)(1-L)^d$. La forme $ARIMA(p, d, q)$ peut s'écrire

$$X_t = \sum_{i=1}^{p+d} \psi_i X_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \text{ et donc } X_{t+h} = \sum_{i=1}^{p+d} \psi_i X_{t+h-i} + \varepsilon_{t+h} + \sum_{j=1}^q \theta_j \varepsilon_{t+h-j}.$$

²³Cette propriété $\overline{L^2}(X_t, X_{t-1}, \dots) = \overline{L^2}(\varepsilon_t, \varepsilon_{t-1}, \dots)$ est une caractérisation du fait que ε_t est l'innovation du processus X_t .

Notons ${}_T X_{T+h}^*$ la prévision faite à la date T , ${}_T X_{T+h}^* = EL(X_{T+h} | X_T, X_{T-1}, \dots, X_0, Z)$. Alors

$${}_T X_{T+h}^* = \sum_{i=1}^{p+d} \psi_{i \cdot T} X_{T+h-i}^* + 0 + \sum_{j=1}^q \theta_{j \cdot T} \varepsilon_{T+h-j}^* \quad \text{où} \quad \begin{cases} {}_T X_{T+h-i}^* = X_{T+h-i} \text{ pour } i \geq h \\ {}_T \varepsilon_{T+h-j}^* = \begin{cases} 0 & \text{pour } j < h \\ \varepsilon_{T+h-j} & \text{pour } j \geq h \end{cases} \end{cases}$$

En particulier, pour $h \geq q$, on obtient une relation de récurrence de la forme

$${}_T X_{T+h}^* = \sum_{i=1}^{p+d} \psi_{i \cdot T} X_{T+h-i}^*.$$

8.4.1 Utilisation de l'approximation AR

Cette approximation est basée sur la propriété (42), rappelée ci-dessous,

Propriété 53 Soit (X_t) un processus ARIMA(p, d, q) de valeurs initiales Z , alors (X_t) peut s'écrire sous la forme AR,

$$X_t = \sum_{j=1}^t a_j X_{t-j} + f^*(t) Z + \varepsilon_t,$$

où les a_j sont les coefficients (pour $j \geq 1$) de la division selon les puissances croissantes de Φ par Θ , et $f^*(t)$ est un vecteur (ligne) de fonctions de t qui tend vers 0 quand $t \rightarrow \infty$.

On peut alors écrire

$$X_{t+h} = \sum_{j=1}^{t+h} a_j X_{t+h-j} + f'(t+h) Z + \varepsilon_{t+h},$$

et ainsi,

$${}_T X_{T+h}^* = EL(X_{t+h} | X_T, X_{T-1}, \dots, X_0, Z) = \sum_{j=1}^{t+h} a_j \cdot {}_T X_{T+h-j}^* + f'(T+h) Z + 0,$$

avec la convention ${}_T X_{T+h-j}^* = X_{T+h-j}$ pour $j \geq h$. Or la limite de $f(t)$ est 0 quand $t \rightarrow \infty$, d'où l'approximation

$${}_T \widehat{X}_{T+h}^* = \sum_{k=1}^{h-1} a_k \cdot {}_T \widehat{X}_{T+h-k}^* + \sum_{k=h}^{T+h} a_k X_{T+h-k}.$$

8.4.2 Utilisation de l'approximation MA

De la même façon, un processus ARIMA peut être approximé par un processus MA,

Propriété 54 Soit (X_t) un processus ARIMA(p, d, q) de valeurs initiales Z , alors (X_t) peut s'écrire sous la forme MA,

$$X_t = \sum_{j=1}^t b_j \varepsilon_{t-j} + g'(t) Z, \quad (38)$$

où les b_j sont les coefficients de la division selon les puissances croissantes de Θ par Φ , et $g'(t)$ est un vecteur (ligne) de fonctions de t .

La relation (38) permet alors d'écrire

$$X_{t+h} = \sum_{j=1}^{t+h} b_j \varepsilon_{t+h-j} + g'(t+h) Z \quad \text{avec} \quad \lim_{u \rightarrow \infty} g(u) = 0.$$

Puisque $\overline{L^2}(X_t, X_{t-1}, \dots) = \overline{L^2}(\varepsilon_t, \varepsilon_{t-1}, \dots)$, on peut écrire ${}_T X_{T+h}^* = EL(X_{T+h} | X_T, X_{T-1}, \dots, X_0, Z) = EL(X_{T+h} | \varepsilon_T, \varepsilon_{T-1}, \dots)$ soit

$${}_T X_{T+h}^* = \sum_{j=h}^{T+h} b_j \varepsilon_{T+h-j} + g'(T+h) Z,$$

d'où les approximations, compte tenu de la limite de $g(\cdot)$,

$${}_T\widehat{X}_{T+h}^* = \sum_{j=h}^{T+h} b_j \varepsilon_{T+h-j} \text{ et } {}_{T+1}\widehat{X}_{T+h}^* = \sum_{j=h-1}^{T+h} b_j \varepsilon_{T+h-j}.$$

D'où la relation

$${}_{T+1}\widehat{X}_{T+h}^* = {}_T\widehat{X}_{T+h}^* + b_{h-1} \left[X_{T+1} - {}_T\widehat{X}_{T+1}^* \right] : \text{Formule de mise à jour}$$

8.5 Intervalle de confiance de la prévision

Cet intervalle est obtenu à partir de la forme $MA(\infty)$ dans le cas où (X_t) est stationnaire, ou de son approximation MA dans le cas non-stationnaire ($ARIMA$).

(i) dans le cas stationnaire,

$$X_{T+h} = \sum_{i=0}^{\infty} b_i \varepsilon_{T+h-i} = \sum_{i=0}^{T+h} b_i \varepsilon_{T+h-i} + \sum_{i=T+h+1}^{\infty} b_i \varepsilon_{T+h-i},$$

et donc

$${}_T\Delta_h = X_{t+h} - {}_T\widehat{X}_{T+h}^* \approx \sum_{i=0}^h b_i \varepsilon_{T+h-i}.$$

(i) dans le cas non-stationnaire ($ARIMA$),

$$X_{T+h} = \sum_{i=0}^{\infty} b_i \varepsilon_{T+h-i} + g'(T+h)Z = \sum_{i=0}^{T+h} b_i \varepsilon_{T+h-i} + \sum_{i=T+h+1}^{\infty} b_i \varepsilon_{T+h-i} + g'(T+h)Z,$$

et donc

$${}_T\Delta_h = X_{t+h} - {}_T\widehat{X}_{T+h}^* \approx \sum_{i=0}^h b_i \varepsilon_{T+h-i}.$$

Sous l'hypothèse de normalité des résidus (ε_t) , $H_0 : \varepsilon_t$ i.i.d., $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, alors

$${}_T\Delta_h = X_{t+h} - {}_T\widehat{X}_{T+h}^* \sim \mathcal{N}\left(0, \sigma^2 \sum_{i=0}^h b_i^2\right),$$

d'où l'intervalle de confiance pour X_{T+h} au niveau $1 - \alpha$

$$\left[{}_T\widehat{X}_{T+h}^* \pm u_{1-\alpha/2} \cdot s \cdot \sqrt{\sum_{i=0}^h \widehat{b}_i^2} \right],$$

où les \widehat{b}_i sont des estimateurs des coefficients de la forme moyenne mobile, et s est un estimateur de la variance du résidu.

8.6 Prévision pour certains processus AR et MA

8.6.1 Prévision pour un processus $AR(1)$

Considérons le processus stationnaire (X_t) , sous la forme générale $X_t = \phi_1 X_{t-1} + \mu + \varepsilon_t$. La prévision à horizon 1, fait à la date T , s'écrit

$${}_T X_{T+1}^* = \mathbb{E}(X_{T+1} | X_T, X_{T-1}, \dots, X_1) = \phi_1 X_T + \mu,$$

et de façon similaire

$${}_T X_{T+2}^* = \phi_1 {}_T X_{T+1}^* + \mu = \phi_1^2 X_T + [\phi_1 + 1] \mu.$$

De façon plus générale, on obtient récursivement la prévision à horizon h ,

$${}_T X_{T+h}^* = \phi_1^h X_T + \left[\phi_1^{h-1} + \dots + \phi_1 + 1 \right] \mu. \quad (39)$$

On peut noter que quand $h \rightarrow \infty$, ${}_T X_{T+h}^*$ tend vers $\delta / (1 - \phi_1)$, la moyenne du processus X_t . L'erreur de prévision à horizon h est donnée par

$$\begin{aligned} {}_T \Delta_h &= {}_T X_{T+h}^* - X_{T+h} = {}_T X_{T+h}^* - [\phi_1 X_{T+h-1} + \mu + \varepsilon_{T+h}] \\ &= \dots = {}_T X_{T+h}^* - \left[\phi_1^h X_T + \left(\phi_1^{h-1} + \dots + \phi_1 + 1 \right) \mu + \varepsilon_{T+h} + \phi_1 \varepsilon_{T+h-1} + \dots + \phi_1^{h-1} \varepsilon_{T+1} \right], \end{aligned}$$

d'où, en substituant (39), on obtient

$${}_T \Delta_h = \varepsilon_{T+h} + \phi_1 \varepsilon_{T+h-1} + \dots + \phi_1^{h-1} \varepsilon_{T+1},$$

qui possède la variance

$$\widehat{V} = \left[1 + \phi_1^2 + \phi_1^4 + \dots + \phi_1^{2h-2} \right] \sigma^2, \text{ où } V(\varepsilon_t) = \sigma^2.$$

La variance de la prévision croît avec l'horizon.

Exemple 92 *Considérons le processus suivant,*

$$X_t = 5 + 0.5X_{t-1} + \varepsilon_t \text{ où } \varepsilon_t \sim \mathcal{N}(0, 1),$$

dont les dernières observations ont été 11.391, 12.748, 10.653, 9.285 et 10.738. La prévision pour la date $T + 1$ est alors

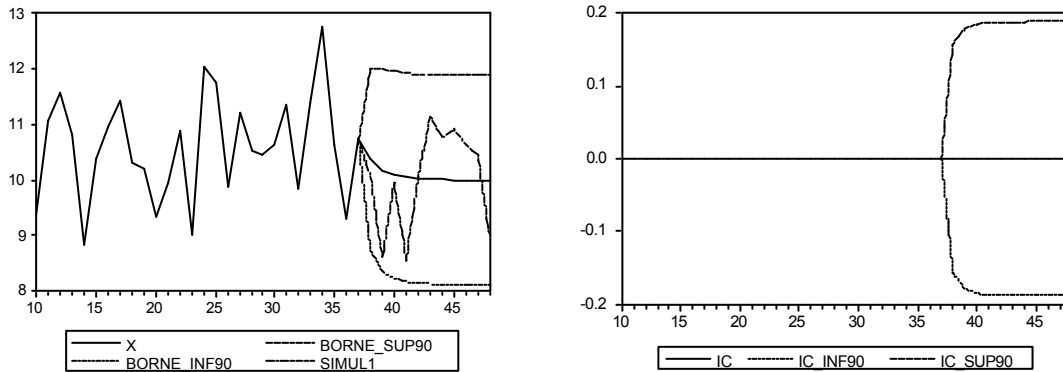
$${}_T X_{T+1}^* = \mathbb{E}(X_{T+1} | X_T, X_{T-1}, \dots, X_1) = \phi_1 X_T + \mu = 5 + 0.5 * 10.738 = 10.369$$

$${}_T X_{T+2}^* = \begin{cases} \phi_1 \cdot {}_T X_{T+1}^* + \mu \\ \mu (1 + \phi_1) + \phi_1 \cdot X_T \end{cases} = 10.184.$$

D'où finalement les prévisions suivantes, avec la variance, l'intervalle de confiance, et la taille relative de l'intervalle de confiance

horizon	0	1	2	3	4	5	6	7
${}_T X_{T+h}^*$	10.738	10.369	10.184	10.092	10.046	10.023	10.012	10.006
\widehat{V}		1.000	1.250	1.313	1.328	1.332	1.333	1.333
$B_{\text{inf}}^{90\%}$		8.739	8.362	8.225	8.168	8.142	8.130	8.124
$B_{\text{sup}}^{90\%}$		11.999	12.007	11.960	11.925	11.904	11.893	11.888
$IC^{90\%}$		$\pm 15.7\%$	$\pm 17.9\%$	$\pm 18.5\%$	$\pm 18.7\%$	$\pm 18.8\%$	$\pm 18.8\%$	$\pm 18.8\%$

Graphiquement, on obtient à gauche les prévisions suivantes (avec la "vraie" valeur de X_t), et à droite la variation relative de l'intervalle de confiance,



8.6.2 Prévision pour un processus $MA(1)$

Considérons le processus stationnaire (X_t) , sous la forme générale

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

La prévision à horizon 1, fait à la date T , s'écrit

$${}_T X_{T+1}^* = \mathbb{E}(X_{T+1} | X_T, X_{T-1}, \dots, X_1) = \mu + \theta_1 \varepsilon_T$$

où ε_T est l'erreur de la dernière observation, à la date T . De façon plus générale, on obtient récursivement la prévision à horizon h ,

$${}_T X_{T+h}^* = \mathbb{E}(X_{T+h} | X_T, X_{T-1}, \dots, X_1) = \mathbb{E}(\mu + \varepsilon_{T+h} + \theta_1 \varepsilon_{T+h-1}) = \mu \tag{40}$$

C'est à dire qu'à partir d'un horizon 2, la meilleure prévision est la moyenne du processus. L'erreur de prévision à horizon h est donnée par

$${}_T \Delta_h = {}_T X_{T+h}^* - X_{T+h} = \varepsilon_{T+h} + \theta_1 \varepsilon_{T+h-1}$$

dont la variance est

$$\widehat{V} = (1 + \theta_1^2) \sigma^2 \text{ où } V(\varepsilon_t) = \sigma^2$$

pour $h \geq 2$. Sinon, pour $h = 1$, la variance est $\widehat{V} = \theta_1^2 \sigma^2$.

Exemple 93 *Considérons le processus suivant,*

$$X_t = 5 + \varepsilon_t - 0.5\varepsilon_{t-1} \text{ où } \varepsilon_t \sim \mathcal{N}(0, 1)$$

dont les dernières observations ont été 4.965, 5.247, 4.686 et 5.654. Pour faire de la prévision, soit on considère la forme AR(∞) du processus, soit on cherche uniquement à exprimer (ε_t) en fonction du passé de (X_t) , ou de $Y_t = X_t - 5$, processus centré

$$\begin{aligned} \varepsilon_t &= Y_t + 0.5\varepsilon_{t-1} = Y_t + 0.5[Y_{t-1} + 0.5\varepsilon_{t-2}] = Y_t + 0.5[Y_{t-1} + 0.5[Y_{t-2} + 0.5\varepsilon_{t-3}]] = \dots \\ &= \sum_{i=0}^{\infty} (0.5)^i Y_{t-i} = \sum_{i=0}^{\infty} (0.5)^i [X_{t-i} - 5] \end{aligned}$$

La prévision pour la date $T + 1$ est alors

$${}_T X_{T+1}^* = \mathbb{E}(X_{T+1} | X_T, X_{T-1}, \dots, X_1) = \mu + \theta_1 \varepsilon_t = 5 - 0.5 \times 0.606 = 3.3049$$

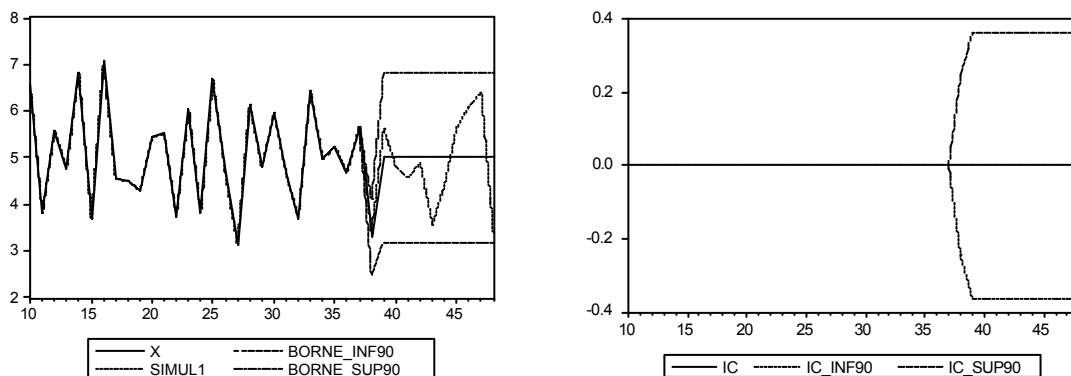
et la prévision à un horizon $h \geq 2$ est alors μ , la constante du modèle,

$${}_T X_{T+2}^* = \mu = 5$$

D'où finalement les prévisions suivantes, avec la variance, l'intervalle de confiance, et la taille relative de l'intervalle de confiance

horizon	0	1	2	3	4	5	6	7
${}_T X_{T+h}^*$	5.654	3.304	5.000	5.000	5.000	5.000	5.000	5.000
\widehat{V}		0.250	1.250	1.250	1.250	1.250	1.250	1.250
$B_{\text{inf}}^{90\%}$		2.489	3.178	3.178	3.178	3.178	3.178	3.178
$B_{\text{sup}}^{90\%}$		4.119	6.822	6.822	6.822	6.822	6.822	6.822
$IC^{90\%}$		$\pm 24.7\%$	$\pm 36.4\%$	$\pm 36.4\%$	$\pm 36.4\%$	$\pm 36.4\%$	$\pm 36.4\%$	$\pm 36.4\%$

Graphiquement, on obtient à gauche les prévisions suivantes (avec la "vraie" valeur de X_t), et à droite la variation relative de l'intervalle de confiance,



8.6.3 Prédiction pour un processus $ARIMA(1, 1, 0)$

Il s'agit ici d'un modèle $AR(1)$ pour la variable intégrée $Y_t = X_t - X_{t-1}$, $Y_t = \phi_1 Y_{t-1} + \mu + \varepsilon_t$. Aussi, la prédiction à horizon $h = 1$ est donnée par

$${}_T X_{T+1}^* = X_T + {}_T Y_{T+1}^*,$$

et de façon plus générale

$${}_T X_{T+h}^* = X_T + {}_T Y_{T+1}^* + {}_T Y_{T+2}^* + \dots + {}_T Y_{T+h}^*.$$

En substituant aux ${}_T Y_{T+i}^*$ ce qui a été trouvé précédemment pour les processus AR , on obtient

$${}_T X_{T+1}^* = (1 + \phi_1) X_T - \phi_1 X_{T-1} + \mu,$$

et, pour un horizon $h = 2$,

$${}_T X_{T+2}^* = (1 + \phi_1 + \phi_1^2) X_T - (\phi_1 + \phi_1^2) X_{T-1} + (\phi_1 + 1) \mu + \mu.$$

L'expression générale de la prédiction à la date h s'obtient récursivement à l'aide de

$$\begin{cases} {}_T Y_{T+h}^* = \phi_1^h Y_T + [\phi_1^{h-1} + \dots + \phi_1 + 1] \mu \\ {}_T X_{T+h}^* = {}_T X_{T+1}^* + \phi_1 \cdot {}_T Y_{T+h-1}^* + \mu. \end{cases}$$

L'erreur faite sur la prédiction à horizon 1 est donnée par

$${}_T \Delta_1 = {}_T X_{T+1}^* - X_{T+1} = {}_T Y_{T+1}^* - Y_{T+1} = \varepsilon_{T+1}, \text{ de variance } \sigma^2.$$

A horizon 2, l'erreur de prédiction est

$${}_T \Delta_2 = {}_T X_{T+2}^* - X_{T+2} = ({}_T Y_{T+1}^* - Y_{T+1}) + ({}_T Y_{T+2}^* - Y_{T+2}) = (1 + \theta_1) \varepsilon_{T+1} + \varepsilon_{T+2},$$

dont la variance est $\hat{V} = [1 + (1 + \phi_1)^2] \sigma^2$. De façon plus générale, l'erreur de prédiction à horizon h est

$$\begin{aligned} {}_T \Delta_h &= ({}_T Y_{T+1}^* - Y_{T+1}) + ({}_T Y_{T+2}^* - Y_{T+2}) + ({}_T Y_{T+3}^* - Y_{T+3}) + \dots + ({}_T Y_{T+h}^* - Y_{T+h}) \\ &= \varepsilon_{T+1} + (\varepsilon_{T+2} + \phi_1 \varepsilon_{T+1}) + \dots + (\varepsilon_{T+h} + \phi_1 \varepsilon_{T+h-1} + \dots + \phi_1^{h-2} \varepsilon_{T+2} + \phi_1^{h-1} \varepsilon_{T+1}) \\ &= \varepsilon_{T+h} + (1 + \phi_1) \varepsilon_{T+h-1} + \dots + (1 + \phi_1 + \dots + \phi_1^{h-1}) \varepsilon_{T+1}, \end{aligned}$$

d'où la variance

$$\hat{V} = \left[\sum_{i=1}^h \left(\sum_{j=0}^{i-1} \phi_1^j \right)^2 \right] \sigma^2.$$

L'erreur de prédiction sur X_{T+h} est alors l'accumulation des erreurs de prédiction de Y_{T+1}, \dots, Y_{T+h} .

Exemple 94 *Considérons le processus (X_t) tel que $X_t - X_{t-1} = Y_t$ où (Y_t) vérifie,*

$$Y_t = 2 + 0.2Y_{t-1} + \varepsilon_t \text{ où } \varepsilon_t \sim \mathcal{N}(0, 1),$$

dont les dernières observations ont été 81.036, 84.074 et 86.586. Le processus (Y_t) sous-jacent peut être obtenu comme différence entre X_t et X_{t-1} . On cherche alors les prévisions de (Y_t) et les prévisions de (X_t) correspondant à la somme des (Y_t) prévus (processus intégré).

(X_t)	70.788	73.606	74.937	78.035	81.036	84.074	86.586
(Y_t)	*	2.818	1.331	3.098	3.001	3.037	2.512

La prédiction pour la date $T + 1$ est alors

$${}_T X_{T+1}^* = X_T + {}_T Y_{T+1}^* \text{ où } {}_T Y_{T+1}^* = \mu + \phi_1 Y_T = 2.502,$$

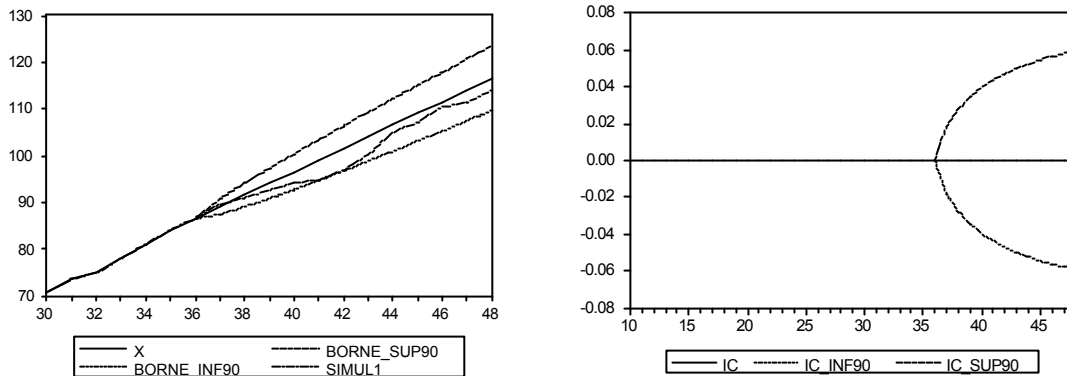
et donc ${}_T X_{T+1}^* = 89.088$. Lla prédiction à un horizon $h \geq 2$ est alors μ

$${}_T X_{T+2}^* = X_T + {}_T Y_{T+1}^* + {}_T Y_{T+2}^* \text{ où } \begin{cases} {}_T Y_{T+1}^* = \mu + \phi_1 Y_T = 2.5024 \\ {}_T Y_{T+2}^* = \mu + \phi_1 \cdot {}_T Y_{T+1}^* = 2.5005 \end{cases}$$

D'où finalement les prévisions suivantes, avec la variance, l'intervalle de confiance, et la taille relative de l'intervalle de confiance

horizon	0	1	2	3	4	5	6	7
${}_T Y_{T+h}^*$	2.512	2.502	2.500	2.500	2.500	2.500	2.500	2.500
${}_T X_{T+h}^*$	86.586	89.088	91.589	94.089	96.589	99.089	101.59	104.09
\widehat{V}		1.000	2.440	3.978	5.535	7.097	8.659	10.22
$B_{\text{inf}}^{90\%}$		87.458	89.043	90.838	92.754	94.747	96.793	98.878
$B_{\text{sup}}^{90\%}$		90.718	94.135	97.340	100.42	103.43	106.39	109.30
$IC^{90\%}$		$\pm 1.8\%$	$\pm 2.8\%$	$\pm 3.5\%$	$\pm 4.0\%$	$\pm 4.4\%$	$\pm 4.7\%$	$\pm 5.0\%$

Graphiquement, on obtient à gauche les prévisions suivantes (avec la "vraie" valeur de X_t), et à droite la variation relative de l'intervalle de confiance,



8.7 Application

8.7.1 Exemple de prévision : cas d'école

Considérons le modèle $ARIMA(1, 1, 1)$ donné par

$$(1 - L)(1 - 0.5L)X_t = (1 - 0.8L)\varepsilon_t,$$

où l'on suppose que ε_t est gaussien, centré, de variance connue $\sigma^2 = 0.5$, avec $X_T = 12$ et ${}_T \widehat{X}_{T+1}^* = EL(X_{T+1} | \underline{X}_T) = 10$, où $\underline{X}_T = \{X_1, \dots, X_T, Z\}$.

La prévision à horizon h faite en T est ${}_T \widehat{X}_{T+h}^* = \mathbb{E}(X_{T+h} | \underline{X}_T)$.

(i) estimation de ${}_T \widehat{X}_{T+h}^*$: Cette forme $ARIMA$ s'écrit

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = \varepsilon_t + \theta_1 \varepsilon_{t-1}, \text{ avec une racine unité.}$$

Aussi, pour $h \geq 2$, on peut écrire ${}_T \widehat{X}_{T+h}^* - \phi_1 \cdot {}_T \widehat{X}_{T+h-1}^* - \phi_2 \cdot {}_T \widehat{X}_{T+h-2}^* = 0$. Pour expliciter les solutions, on se ramène au problème suivant : "recherche des suites u_n telle que $u_n = \alpha u_{n-1} + \beta u_{n-2}$ "²⁴. Les racines du polynôme caractéristique étant 1 et 1/2, on peut écrire ${}_T \widehat{X}_{T+k}^* = \lambda \cdot 1^k + \mu \cdot (1/2)^k$. Compte tenu du fait que $X_T = 12$ et ${}_T \widehat{X}_{T+1}^* = 10$ on en déduit $\lambda = 8$ et $\mu = 4$. Aussi

$${}_T \widehat{X}_{T+h}^* = 8 + \frac{4}{2^k} \text{ d'où les premières valeurs } \{12, 10, 9, 8.5, 8.25, 8.125, \dots\}$$

(ii) expression de l'intervalle de confiance : L'intervalle de confiance à 95% de la prévision est de la forme

$$\left[{}_T \widehat{X}_{T+h}^* - 1.96 \sqrt{V({}_T \widehat{X}_{T+h}^*)}; {}_T \widehat{X}_{T+h}^* + 1.96 \sqrt{V({}_T \widehat{X}_{T+h}^*)} \right].$$

²⁴Rappel : Pour une relation récurrente $u_n = \alpha u_{n-1} + \beta u_{n-2}$, la forme générale des solutions est $u_n = \lambda r_1^n + \mu r_2^n$ où r_1 et r_2 sont les racines du polynôme $P(x) = x^2 - \alpha x - \beta$, dans le cas où les racines sont distinctes. Dans le cas où P admet une racine double (r), la forme générale est $u_n = (\lambda + \mu r) r^n$. Dans le cas où la suite est entièrement déterminée par les valeurs initiales u_1 et u_2 , alors λ et μ sont entièrement déterminés par la résolution du système

$$\begin{cases} u_1 = \lambda r_1 + \mu r_2 \\ u_2 = \lambda r_1^2 + \mu r_2^2 \end{cases}$$

Cette variance s'obtient en approximant la forme *ARIMA* par une forme *MA*(∞),

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

On note alors que

$$\begin{cases} X_{T+1} - X_T = \varepsilon_{T+1} + b_1\varepsilon_T + b_2\varepsilon_{T-1} + \dots \\ X_{T+2} - X_{T+1} = \varepsilon_{T+2} + b_1\varepsilon_{T+1} + b_2\varepsilon_T + \dots \\ \dots \\ X_{T+h} - X_{T+h-1} = \varepsilon_{T+h} + b_1\varepsilon_{T+h-1} + b_2\varepsilon_{T+h-2} + \dots \end{cases}$$

d'où, par sommation, l'expression de $X_{T+h} - X_T$ et en considérant la variance (*puisque les ε_t sont identiquement distribués, de variance σ^2 , et surtout indépendant*),

$$V\left({}_T\widehat{X}_{T+h}^*\right) = \sigma^2 \left[1 + (1 + b_1)^2 + (1 + b_1 + b_2)^2 + \dots + (1 + b_1 + \dots + b_h)^2\right].$$

Or l'expression des b_i est donnée par la relation $B(L) = \Theta(L)\Phi^{-1}(L) = (1 - 0.8L)(1 - 0.5L)^{-1}$

$$B(L) = (1 - 0.8L)(1 + 0.5L + 0.5^2L^2 + \dots) = 1 - 0.3L - 0.3\frac{L^2}{2} - 0.3\frac{L^3}{2^2} - \dots$$

et donc $b_0 = 1$ et $b_i = -0.3/2^{i-1}$. Aussi

$$1 + b_1 + \dots + b_j = 1 - 0.3 \left[1 + \frac{1}{2} + \frac{1}{4} + \dots + \frac{1}{2^{j-1}}\right] = 1 - 0.6 \left[1 - \frac{1}{2^j}\right] = 0.4 + \frac{0.6}{2^j},$$

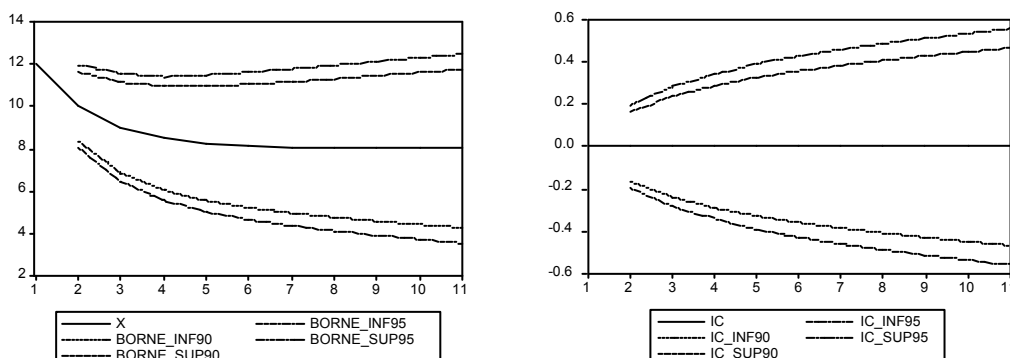
et donc

$$V\left({}_T\widehat{X}_{T+h}^*\right) = \sigma^2 \sum_{j=0}^{h-1} \left[0.4 + \frac{0.6}{2^j}\right]^2.$$

D'où finalement les prévisions et les premiers intervalles de confiance suivant :

h	0	1	2	3	4	5	6	7	8	9	10
$B_{inf}^{95\%}$		8.040	6.444	5.560	5.015	4.639	4.353	4.116	3.906	3.715	3.535
$B_{inf}^{90\%}$		8.360	6.862	6.040	5.543	5.209	4.959	4.755	4.577	4.416	4.265
${}_T\widehat{X}_{T+h}^*$	12.000	10.000	9.000	8.500	8.250	8.125	8.063	8.031	8.016	8.007	8.004
$B_{sup}^{90\%}$		11.640	11.138	10.960	10.957	11.041	11.166	11.307	11.454	11.600	11.743
$B_{sup}^{95\%}$		11.960	11.556	11.440	11.485	11.611	11.772	11.947	12.125	12.301	12.473
$IC_{95\%}$		±19.6%	±28.4%	±34.6%	±39.2%	±42.9%	±46.0%	±48.7%	±51.3%	±53.6%	±55.8%

On notera bien sur cet exemple que les prévisions à l'aide d'un modèle *ARIMA* à moyen terme sont déjà relativement incertaines. Graphiquement, on obtient, à gauche, la prévision suivante (*avec les intervalles de confiance à 90% et 95%*), et à droite, l'évolution relative de l'intervalle de confiance en fonction de l'horizon de prévision



8.7.2 Exemple d'application : cas pratique

Considérons ici la série du nombre de voyageurs SNCF, et la modélisation *ARIMA* que nous avons pu faire,

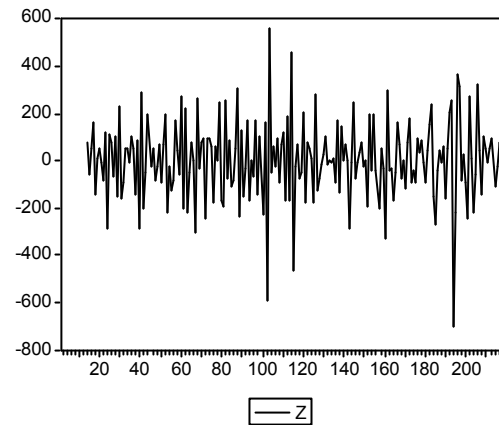
$$(1 - L)(1 - L^{12})X_t = \begin{pmatrix} 1 - 0.8344L \\ (0.0402) \end{pmatrix} \begin{pmatrix} 1 - 0.4926L^{12} \\ (0.0687) \end{pmatrix} \varepsilon_t. \quad (41)$$

La série Z_t peut alors être utilisée pour faire de la prévision, à l'aide de sa modélisation *MA*, sous *SAS*

ARIMA Procedure

Forecasts for variable Z

Obs	Forecast	Std Error	Lower 95%	Upper 95%
205	-56.9839	124.1251	-300.2647	186.2969
206	322.5639	161.6583	5.7195	639.4083
207	43.9931	161.6583	-272.8513	360.8374
208	-146.3447	161.6583	-463.1890	170.4997
209	-100.0972	161.6583	-416.9416	216.7472
210	56.2979	161.6583	-260.5465	373.1422
211	-10.3091	161.6583	-327.1535	306.5353
212	50.8267	161.6583	-266.0176	367.6711
213	95.2525	161.6583	-221.5919	412.0968
214	-112.7559	161.6583	-429.6002	204.0885
215	-23.3419	161.6583	-340.1863	293.5024
216	73.6131	161.6583	-243.2313	390.4574
217	-36.8593	172.8357	-375.6110	301.8924



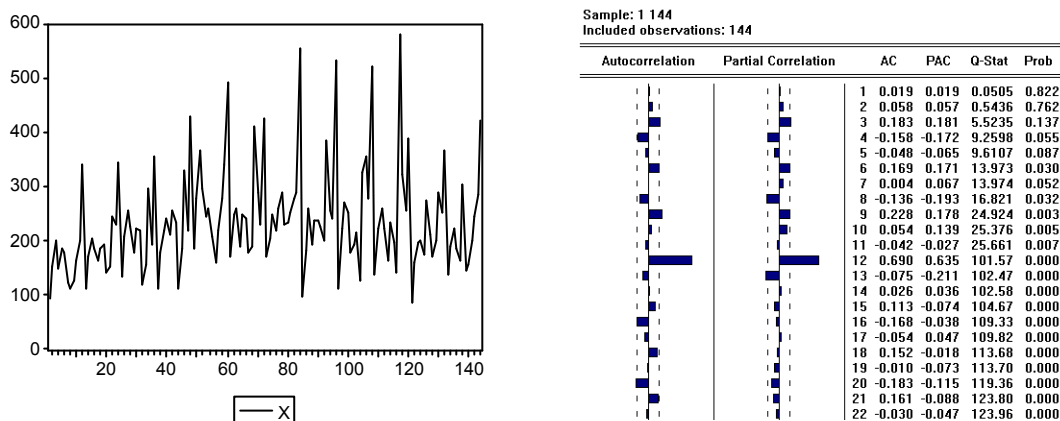
Comme nous le voyons sur cet exemple, la prévision à l'aide d'une modélisation *ARMA* reste relativement floue, même ici à court terme (*un exemple analogue sera repris plus en détails dans la partie suivante*).

9 Applications de la méthode de Box & Jenkins

Nous allons ici nous intéresser à 2 applications de ces modèles sur données réelles. Dans une première partie, nous allons voir comment utiliser cette méthode en assurance-vie, afin de prévoir un nombre de contrats dans un portefeuille. Dans une seconde partie, nous allons voir comment utiliser cette méthode en finance, pour modéliser des taux d'intérêt. Enfin, dans une troisième partie, nous allons nous intéresser à l'application de cette méthode sur des données simulées.

9.1 Application à un portefeuille d'assurance-vie

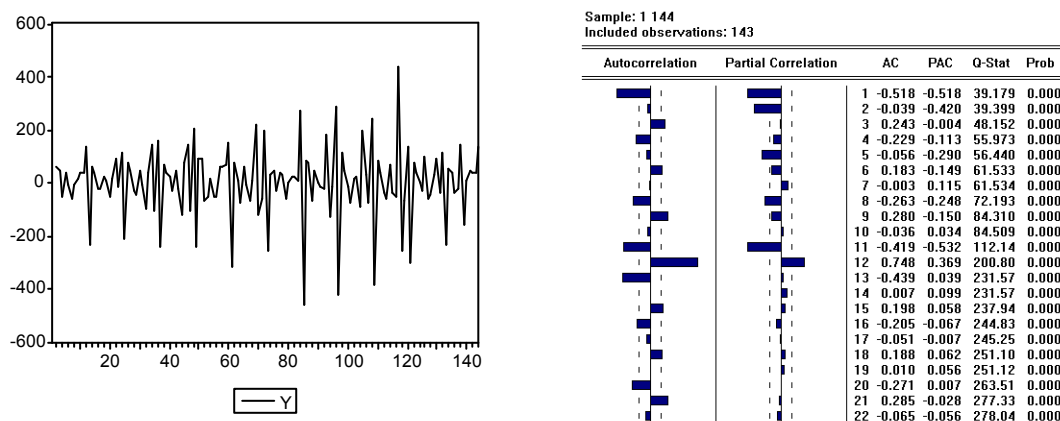
Nous allons considérer ici le nombre de contrats dans un portefeuille, exprimé en nombre de contrats mensuels, à l'aide de données mensuelles, sur une dizaine d'années (*source : Dreesbeke, Fichet, Tassi (1992)*)



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

9.1.1 Modélisation de la série

Afin d'enlever une éventuelle tendance, les auteurs proposent d'étudier $Y_t = (1 - L) X_t = \Delta X_t$



Toutefois, une étude des tests de racine unité sur la série (X_t) aurait permis de confirmer l'hypothèse de stationnarité de la série

Null Hypothesis: X has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic based on SIC, MAXLAG=10)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-8.531272	0.0000
Test critical values:		
1% level	-4.105534	
5% level	-3.480463	
10% level	-3.168039	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(X)
Method: Least Squares
Date: 07/18/03 Time: 07:58
Sample: 80 144
Included observations: 65

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X(-1)	-1.108362	0.129918	-8.531272	0.0000
C	296.8203	43.59923	6.807926	0.0000
@TREND(80)	-0.689283	0.698095	-0.987377	0.3273

R-squared	0.540368	Mean dependent var	2.907692
Adjusted R-squared	0.525541	S.D. dependent var	151.4768
S.E. of regression	104.3387	Akaike info criterion	12.17822
Sum squared resid	674966.5	Schwarz criterion	12.27857
Log likelihood	-392.7920	F-statistic	36.44527
Durbin-Watson stat	1.953637	Prob(F-statistic)	0.000000

Null Hypothesis: X has a unit root
Exogenous: Constant, Linear Trend
Bandwidth: 2 (Newey-West using Bartlett kernel)

	Adj. t-Stat	Prob.*
PhillipsPerron test statistic	-8.537252	0.0000
Test critical values:		
1% level	-4.105534	
5% level	-3.480463	
10% level	-3.168039	

*MacKinnon (1996) one-sided p-values.

Residual variance (no correction) 10384.10
HAC corrected variance (Bartlett kernel) 10040.64

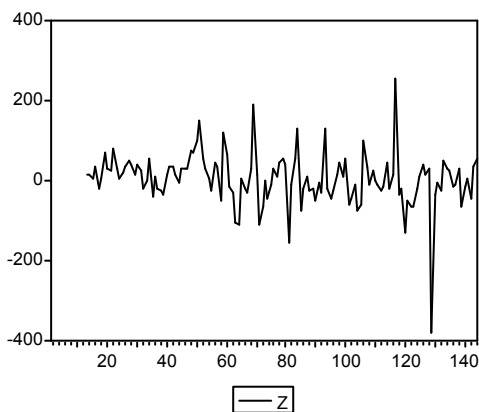
Phillips-Perron Test Equation
Dependent Variable: D(X)
Method: Least Squares

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X(-1)	-1.108362	0.129918	-8.531272	0.0000
C	296.8203	43.59923	6.807926	0.0000
@TREND(80)	-0.689283	0.698095	-0.987377	0.3273

R-squared	0.540368	Mean dependent var	2.907692
Adjusted R-squared	0.525541	S.D. dependent var	151.4768
S.E. of regression	104.3387	Akaike info criterion	12.17822
Sum squared resid	674966.5	Schwarz criterion	12.27857
Log likelihood	-392.7920	F-statistic	36.44527
Durbin-Watson stat	1.953637	Prob(F-statistic)	0.000000

avec le test *ADF* (avec constante et tendance), à gauche, et le test de Philipps et Perron à droite. Ces deux tests rejettent l'hypothèse de présence de racine unité.

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



Date: 12/21/01 Time: 15:51
Sample: 1 144
Included observations: 132

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.168	0.168	3.7953	0.051	
2	-0.032	-0.062	3.9378	0.140	
3	-0.026	-0.010	4.0283	0.258	
4	-0.065	-0.062	4.6037	0.330	
5	-0.111	-0.094	6.3279	0.276	
6	-0.033	-0.004	6.4786	0.372	
7	0.099	0.098	7.8625	0.345	
8	0.122	0.086	9.9991	0.265	
9	0.251	0.228	19.033	0.025	
10	0.211	0.155	25.500	0.004	
11	0.056	0.046	25.961	0.007	
12	-0.413	-0.430	51.152	0.000	
13	-0.119	0.024	53.262	0.000	
14	-0.015	0.005	53.296	0.000	
15	-0.095	-0.074	54.652	0.000	
16	-0.016	-0.058	54.691	0.000	
17	0.103	-0.011	56.331	0.000	
18	0.121	0.011	58.606	0.000	
19	0.100	0.102	60.157	0.000	
20	-0.005	-0.019	60.161	0.000	
21	-0.055	0.159	60.638	0.000	
22	-0.130	0.026	63.351	0.000	
23	-0.039	0.103	63.598	0.000	

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

9.1.2 Estimation de modèles *ARMA*

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

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

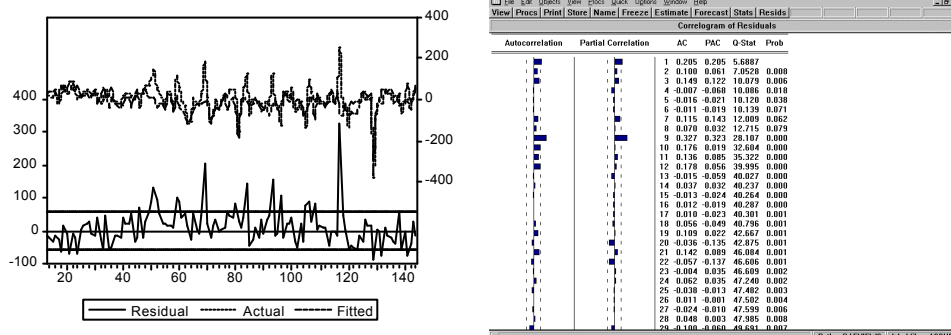
étant donné que les autocorrélations entre 1 et 11 peuvent sembler négligeables. Une méthode basée sur des moindres carrés conditionnels, donne les estimations suivantes (*respectivement avec et sans constante*)

LS // Dependent Variable is Z				LS // Dependent Variables Z					
Sample: 13 144				Sample: 13 144					
Included observations: 132 after adjusting endpoints				Included observations: 132 after adjusting endpoints					
Convergence achieved after 17 iterations				Convergence achieved after 14 iterations					
Variable	Coefficient	Std. Error	T-Statistic	Prob.	Variable	Coefficient	Std. Error	T-Statistic	Prob.
C	2.086042	1.845862	1.130118	0.2605	MA(12)	-0.899009	0.032780	-27.42595	0.0000
MA(12)	-0.881818	0.043054	-20.48182	0.0000					
R-squared	0.258916		Mean dependent var	4.750000	R-squared	0.250655		Mean dependent var	4.750000
Adjusted R-squared	0.253216		S.D. dependent var	65.83328	Adjusted R-squared	0.250655		S.D. dependent var	65.83328
S.E. of regression	56.89094		Akaike info criterion	8.097308	S.E. of regression	56.98839		Akaike info criterion	8.093242
Sum squared resid	420755.3		Schwarz criterion	8.140987	Sum squared resid	425445.6		Schwarz criterion	8.115081
Log likelihood	-719.7222		F-statistic	45.41878	Log likelihood	-720.4538		Durbin-Watson stat	1.502360
Durbin-Watson stat	1.517747		Prob(F-statistic)	0.000000					
Inverted MA Roots	.99	.86 - .49i	.86 + .49i	.49 - .86i	Inverted MA Roots	.99	.86 + .50i	.86 - .50i	.50 + .86i
	.49 + .86i	.00 + .99i	.00 - .99i	-.49 + .86i		.50 - .86i	-.00 - .99i	-.00 + .99i	-.50 - .86i
	-.49 - .86i	-.86 + .49i	-.86 - .49i	-.86 - .49i		-.50 + .86i	-.86 + .50i	-.86 - .50i	-.99

Pour ce premier modèle, nous obtenons l'écriture suivante

$$[1] \quad (1 - L^{12}) X_t = \begin{pmatrix} 1 + 0.89901L^{12} \\ (0.0328) \end{pmatrix} \varepsilon_t$$

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

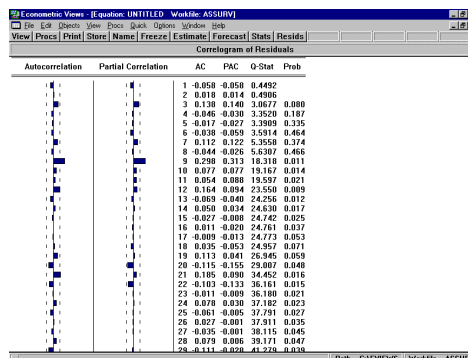
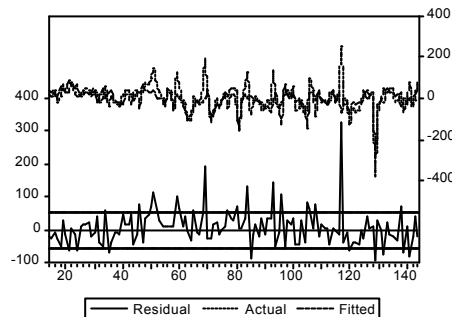


Le modèle présenté est rejeté : les erreurs ne suivent pas un bruit blanc. En particulier, si l'on considère les autocorrélations de (Z_t) la première semble significativement non nulle : ceci pousse à tester un modèle $ARMA(1, 12)$ de la forme $(1 - \phi L) Z_t = (1 - \theta L^{12}) \varepsilon_t$,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

LS // Dependent Variable is Z				LS // Dependent Variable is Z					
Sample: 14 144				Sample: 14 144					
Included observations: 131 after adjusting endpoints				Included observations: 131 after adjusting endpoints					
Convergence achieved after 13 iterations				Convergence achieved after 10 iterations					
Variable	Coefficient	Std. Error	T-Statistic	Prob.	Variable	Coefficient	Std. Error	T-Statistic	Prob.
C	1.959250	2.238375	0.875300	0.3831	AR(1)	0.247243	0.084927	2.911259	0.0042
AR(1)	0.238275	0.085711	2.779982	0.0063	MA(12)	-0.890845	0.034450	-25.85917	0.0000
MA(12)	-0.882966	0.039380	-22.42171	0.0000					
R-squared	0.301621		Mean dependent var	4.656489	R-squared	0.297269		Mean dependent var	4.656489
Adjusted R-squared	0.290709		S.D. dependent var	66.07720	Adjusted R-squared	0.291822		S.D. dependent var	66.07720
S.E. of regression	55.64982		Akaike info criterion	8.060792	S.E. of regression	55.60616		Akaike info criterion	8.051737
Sum squared resid	396403.6		Schwarz criterion	8.126636	Sum squared resid	398873.8		Schwarz criterion	8.095633
Log likelihood	-710.8628		F-statistic	27.64084	Log likelihood	-711.2697		F-statistic	54.56963
Durbin-Watson stat	2.040467		Prob(F-statistic)	0.000000	Durbin-Watson stat	2.049872		Prob(F-statistic)	0.000000
Inverted AR Roots	.24				Inverted AR Roots	.25			
Inverted MA Roots	.99	.86 + .49i	.86 - .49i	.49 - .86i	Inverted MA Roots	.99	.86 + .50i	.86 - .50i	.50 + .86i
	.49 + .86i	.00 + .99i	.00 - .99i	-.49 + .86i		.50 - .86i	-.00 - .99i	-.00 + .99i	-.50 - .86i
	-.49 - .86i	-.86 + .49i	-.86 - .49i	-.86 - .49i		-.50 + .86i	-.86 + .50i	-.86 - .50i	-.99

$$[2] \quad \left(1 - \frac{0.24724L}{(0.084)}\right) (1 - L^{12}) X_t = \begin{pmatrix} 1 + 0.89085L^{12} \\ (0.0345) \end{pmatrix} \varepsilon_t$$

dont les résidus ε_t ont le comportement suivant



Là encore, l'hypothèse de bruit blanc des résidus est rejetée.

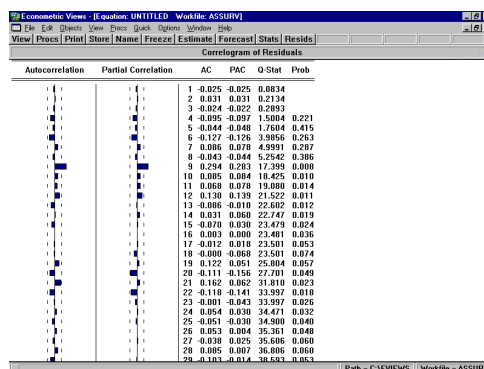
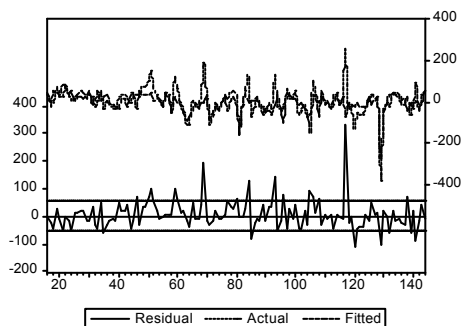
A titre de comparaison (de la même façon que cela est fait dans Droesbeke, Fichet, Tassi (1992)), l'estimation d'un modèle $ARMA(3, 12)$ donne

LS // Dependent Variable is Z				LS // Dependent Variable is Z					
Sample: 16 144				Sample: 16 144					
Included observations: 129 after adjusting endpoints				Included observations: 129 after adjusting endpoints					
Convergence achieved after 8 iterations				Convergence achieved after 8 iterations					
Variable	Coefficient	Std. Error	T-Statistic	Prob.	Variable	Coefficient	Std. Error	T-Statistic	Prob.
AR(1)	0.203724	0.087952	2.316310	0.0222	AR(1)	0.215663	0.085560	2.520598	0.0130
AR(2)	0.054332	0.090117	0.602907	0.5477	AR(3)	0.164294	0.086724	1.894447	0.0605
AR(3)	0.152170	0.089481	1.700579	0.0915	MA(12)	-0.892081	0.034309	-26.00140	0.0000
MA(12)	-0.893591	0.034620	-25.81143	0.0000					
R-squared	0.323645	Mean dependent var	4.558140		R-squared	0.321653	Mean dependent var	4.558140	
Adjusted R-squared	0.307412	S.D. dependent var	66.58241		Adjusted R-squared	0.310865	S.D. dependent var	66.58241	
S.E. of regression	55.41111	Akaike info criterion	8.060077		S.E. of regression	55.27198	Akaike info criterion	8.047513	
Sum squared resid	383798.9	Schwarz criterion	8.148754		Sum squared resid	384929.0	Schwarz criterion	8.114021	
Log likelihood	-698.9180	F-statistic	19.93805		Log likelihood	-699.1077	F-statistic	29.87288	
Durbin-Watson stat	1.988017	Prob(F-statistic)	0.000000		Durbin-Watson stat	2.011660	Prob(F-statistic)	0.000000	
Inverted AR Roots	.65	-.22+ .43i	-.22 - .43i		Inverted AR Roots	.63	-.21+ .47i	-.21 - .47i	
Inverted MA Roots	.99	.86 - .50i	.86 + .50i	.50 - .86i	Inverted MA Roots	.99	.86 + .50i	.86 - .50i	.50 + .86i
		.50+ .86i	.00+ .99i	-.50+ .86i			.00 - .99i	-.00 + .99i	-.50 - .86i
		-.50 - .86i	-.86+ .50i	-.86 - .50i			-.50+ .86i	-.86 - .50i	-.99

Le modèle avec des composantes $AR(1)$, $AR(3)$ et $MA(12)$ est significatif.

$$[3] \quad \left(1 - \frac{0.21566L}{(0.086)} - \frac{0.16429L^3}{(0.087)} \right) (1 - L^{12}) X_t = \left(1 + \frac{0.89208L^{12}}{(0.0343)} \right) \varepsilon_t$$

On peut également noter que les racines des polynômes de la composante AR et de la composante MA sont distinctes deux à deux. Toutefois, l'allure des résidus, représentée ci-dessous, pousse encore une fois à rejeter l'hypothèse de bruit blanc des erreurs,



En revenant au modèle précédent, on peut noter la présence d'une forte autocorrélation à l'ordre 9. On peut alors

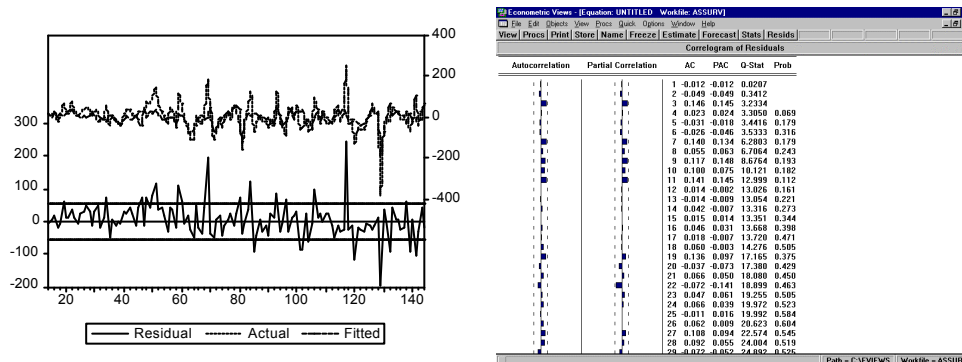
tenter de modéliser la série à l'aide d'un modèle de la forme $(1 - \phi L) Z_t = (1 - \alpha L^9 - \beta L^{12}) \varepsilon_t$.

LS // Dependent Variable is Z					LS // Dependent Variable is Z				
Sample: 14 144					Sample: 14 144				
Included observations: 131 after adjusting endpoints					Included observations: 131 after adjusting endpoints				
Convergence achieved after 13 iterations					Convergence achieved after 13 iterations				
Variable	Coefficient	Std. Error	T-Statistic	Prob.	Variable	Coefficient	Std. Error	T-Statistic	Prob.
C	3.716450	3.689678	1.012745	0.3131	AR(1)	0.214099	0.087981	2.433366	0.0163
AR(1)	0.208215	0.088354	2.356593	0.0200	MA(12)	-0.674685	0.060045	-11.23631	0.0000
MA(9)	0.226890	0.058097	3.906922	0.0002	MA(9)	0.234509	0.057176	4.101524	0.0001
MA(12)	-0.684472	0.060959	-11.22845	0.0000					
R-squared	0.338724		Mean dependent var	4.656489	R-squared	0.333420		Mean dependent var	4.656489
Adjusted R-squared	0.323103		S.D. dependent var	66.07720	Adjusted R-squared	0.323004		S.D. dependent var	66.07720
S.E. of regression	54.36419		Akaike info criterion	8.021470	S.E. of regression	54.36816		Akaike info criterion	8.014192
Sum squared resid	375344.1		Schwartz criterion	8.109262	Sum squared resid	378354.7		Schwartz criterion	8.080036
Log likelihood	-707.2872		F-statistic	21.68428	Log likelihood	-707.8105		F-statistic	32.01242
Durbin-Watson stat	1.995098		Prob(F-statistic)	0.000000	Durbin-Watson stat	1.997797		Prob(F-statistic)	0.000000
Inverted AR Roots	.21				Inverted AR Roots	.21			
Inverted MA Roots	.94	.85	-.46i	.85+.46i	.50+.86i	.94	.85+.46i	.85-.46i	.50-.86i
	.50-.86i	-.02+.97i	-.02-.97i	-.47+.82i		.50+.86i	-.03+.97i	-.03-.97i	-.47-.81i
	-.47-.82i	-.83-.51i	-.83+.51i	-.99		-.47+.81i	-.83+.51i	-.83-.51i	-.99

d'où le modèle

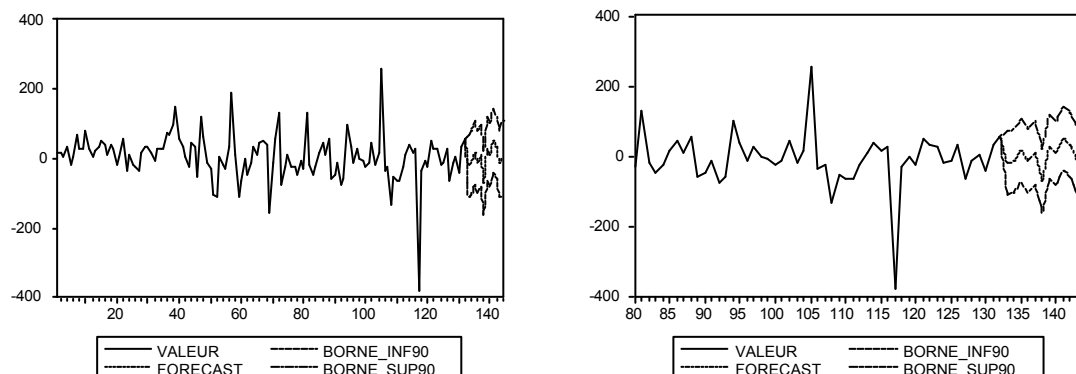
$$[4] \quad \left(1 - \frac{0.21409L}{(0.088)}\right) (1 - L^{12}) X_t = \left(1 - \frac{0.23450L^9}{(0.060)} + \frac{0.67498L^{12}}{(0.060)}\right) \varepsilon_t$$

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



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

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



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

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

le polynôme $MA B(L)$ s'écrit alors

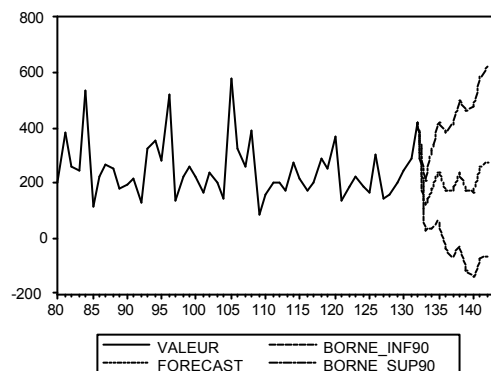
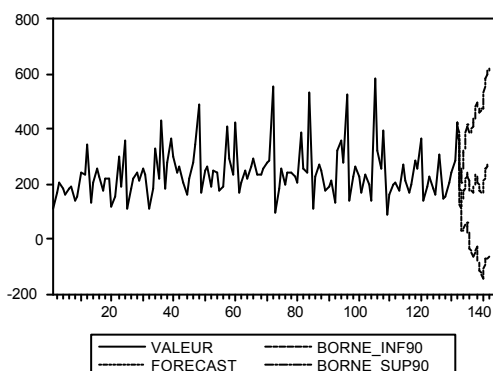
$$B(L) = \sum_{i=0}^{\infty} b_i L^i \text{ où } b_i = \begin{cases} 1 & \text{pour } i = 0 \\ 0.2141^i & \text{pour } 1 \leq i \leq 8 \\ 0.2141^i - 0.2141^{i-9} \times 0.2345 & \text{pour } 9 \leq i \leq 11 \\ 0.2141^i - 0.2141^{i-9} \times 0.2345 + 0.2141^{i-12} \times 0.6745 & \text{pour } 11 \leq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

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

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

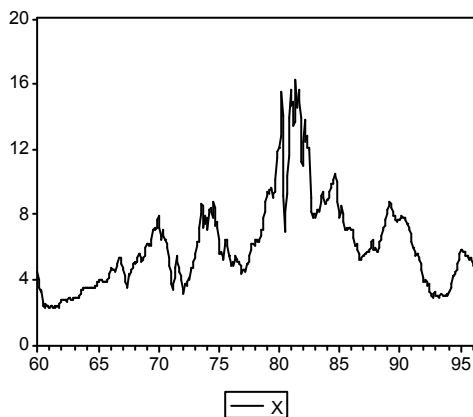
soit ici, puisque σ^2 est estimé par $s^2 = 54.368^2$,

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



9.2 Application de la série des taux d'intérêt à 3 mois

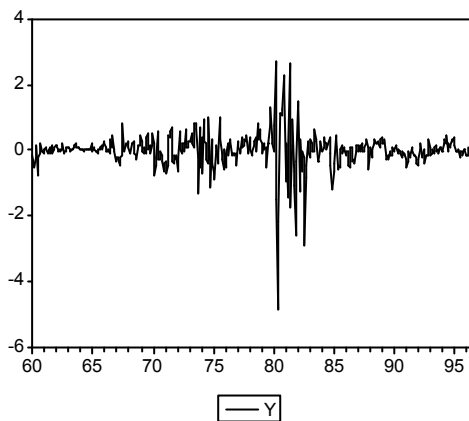
Nous allons considérer ici les taux à 3 mois du trésor américain (*source :Pindyck et Rubinfeld (1998)*), données mensuelles, de Janvier 1960 à Mars 1996



Sample: 1960:01 1996:03
Included observations: 435

	Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.980	0.980	420.61	0.000		
2	0.949	-0.283	816.02	0.000		
3	0.922	0.181	1190.5	0.000		
4	0.900	-0.000	1547.6	0.000		
5	0.879	0.022	1889.1	0.000		
6	0.856	-0.087	2213.4	0.000		
7	0.839	0.244	2526.1	0.000		
8	0.830	0.038	2832.7	0.000		
9	0.817	-0.151	3130.5	0.000		
10	0.796	-0.108	3414.1	0.000		
11	0.773	0.041	3681.7	0.000		
12	0.749	-0.083	3934.0	0.000		
13	0.730	0.106	4174.0	0.000		
14	0.710	-0.069	4401.5	0.000		
15	0.683	-0.153	4612.5	0.000		
16	0.660	0.117	4810.1	0.000		
17	0.636	-0.172	4994.2	0.000		
18	0.609	-0.069	5163.4	0.000		
19	0.578	-0.047	5315.9	0.000		
20	0.547	0.119	5453.2	0.000		
21	0.526	0.059	5580.4	0.000		
22	0.511	0.054	5700.7	0.000		

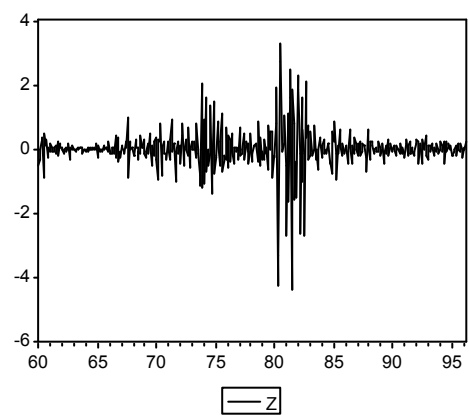
L'autocorrélogramme de la série bute des taux (X_t) permet de se rendre compte rapidement que la série n'est pas stationnaire. La série différenciée $Y_t = X_t - X_{t-1}$ a l'allure suivante



Sample: 1960:01 1996:03
Included observations: 435

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.272	0.272	32.519	0.000	
2	-0.116	-0.206	38.446	0.000	
3	-0.102	-0.009	43.057	0.000	
4	-0.042	-0.033	43.828	0.000	
5	0.055	0.064	45.174	0.000	
6	-0.180	-0.260	59.606	0.000	
7	-0.192	-0.048	75.978	0.000	
8	0.102	0.147	80.634	0.000	
9	0.207	0.092	99.762	0.000	
10	0.067	-0.045	101.78	0.000	
11	-0.007	0.068	101.81	0.000	
12	-0.107	-0.122	106.92	0.000	
13	0.032	0.062	107.39	0.000	
14	0.169	0.149	120.34	0.000	
15	-0.103	-0.141	125.14	0.000	
16	0.008	0.170	125.16	0.000	
17	0.097	0.044	129.45	0.000	
18	0.098	0.022	133.79	0.000	
19	-0.021	-0.135	134.00	0.000	
20	-0.240	-0.080	160.35	0.000	
21	-0.162	-0.072	172.43	0.000	
22	-0.004	-0.044	172.44	0.000	

La série ainsi formée semble stationnaire. A titre comparatif, la série Z_t obtenue en différenciant 2 fois donne des résultats ne semblant pas significativement différents



Sample: 1960:01 1996:03
Included observations: 435

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	-0.232	-0.232	23.541	0.000	
2	-0.276	-0.349	57.041	0.000	
3	-0.033	-0.240	57.511	0.000	
4	-0.026	-0.270	57.804	0.000	
5	0.230	0.058	81.202	0.000	
6	-0.155	-0.179	91.811	0.000	
7	-0.210	-0.309	111.44	0.000	
8	0.129	-0.203	118.90	0.000	
9	0.168	-0.048	131.56	0.000	
10	-0.045	-0.153	132.48	0.000	
11	0.018	0.042	132.62	0.000	
12	-0.164	-0.142	144.70	0.000	
13	0.001	-0.205	144.70	0.000	
14	0.281	0.088	180.42	0.000	
15	-0.263	-0.217	211.71	0.000	
16	0.015	-0.079	211.81	0.000	
17	0.061	-0.052	213.49	0.000	
18	0.082	0.101	216.58	0.000	
19	0.068	0.032	218.89	0.000	
20	-0.203	0.020	237.53	0.000	
21	-0.055	-0.014	238.94	0.000	
22	0.110	-0.034	244.55	0.000	

Aussi, différencier 1 fois suffit pour obtenir un modèle stationnaire.

9.2.1 Modélisation de la série

Compte tenu de l'allure des autocorrélogrammes de Y_t , nous pouvons penser modéliser la série X_t par un processus $ARMA(p,q)$. La méthode du coin, décrite auparavant, donne le tableau suivant

$i \setminus j$	1	2	3	4	5	6
1	0.272	0.189	-0.007	0.024	0.041	0.148
2	-0.116	0.041	-0.006	0.001	-0.003	0.040
3	-0.102	0.006	0.003	-0.001	0.001	0.011
4	-0.042	0.007	-0.002	0.002	0.003	0.003
5	0.055	-0.004	-0.005	0.002	-0.001	0.001
6	-0.180	0.043	-0.012	0.003	0.001	0.000

9.2.2 Estimation des paramètres d'une modélisation $ARIMA(1,1,1)$

L'estimation donne les résultats suivants (*la constante était clairement non significative*),

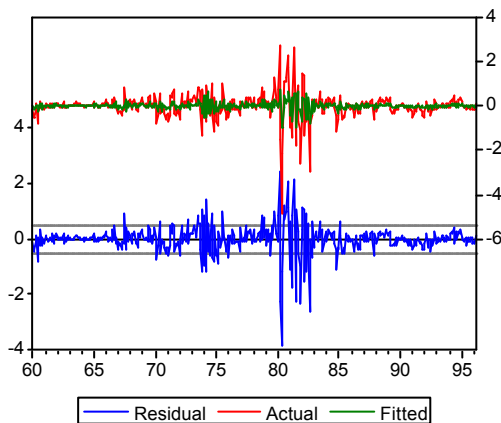
$$\left(1 + \frac{0.324L}{(0.105)}\right) Y_t = \left(1 + \frac{0.662L}{(0.083)}\right) \varepsilon_t$$

LS // Dependent Variable is Y
 Sample: 1960:01 1996:03
 Included observations: 435
 Convergence achieved after 4 iterations

Variable	Coefficient	Std. Error	T-Statistic	Prob.
AR(1)	-0.324261	0.104881	-3.091706	0.0021
MA(1)	0.662163	0.082710	8.005835	0.0000
R-squared	0.112854	Mean dependent var	0.000892	
Adjusted R-squared	0.110805	S.D. dependent var	0.546661	
S.E. of regression	0.515486	Akaike info criterion	-1.320703	
Sum squared resid	115.0593	Schwartz criterion	-1.301966	
Log likelihood	-327.9853	F-statistic	55.08189	
Durbin-Watson stat	1.989933	Prob(F-statistic)	0.000000	
Inverted AR Roots	-.32			
Inverted MA Roots	-.66			

Sample: 1960:01 1996:03
 Included observations: 435
 Q-statistic probabilities adjusted for 2 ARMA term[s]

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.005	0.005	0.0105		
2	-0.010	-0.010	0.0568		
3	-0.108	-0.108	5.1537	0.023	
4	-0.036	-0.036	5.7249	0.057	
5	0.081	0.080	8.6262	0.035	
6	-0.159	-0.175	19.831	0.001	
7	-0.151	-0.162	29.925	0.000	
8	0.082	0.105	32.915	0.000	
9	0.173	0.154	46.336	0.000	
10	0.024	-0.040	46.600	0.000	
11	0.011	0.041	46.650	0.000	
12	-0.082	-0.036	49.684	0.000	
13	-0.019	-0.082	49.846	0.000	
14	0.218	0.230	71.378	0.000	
15	-0.176	-0.123	85.373	0.000	
16	0.068	0.004	87.469	0.000	
17	0.040	0.102	88.202	0.000	
18	0.086	0.034	91.570	0.000	
19	-0.001	-0.093	91.571	0.000	
20	-0.202	-0.110	110.23	0.000	
21	-0.113	-0.069	116.06	0.000	



Conditional Least Squares Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag
MU	0.0009002	0.03114	0.03	0
MA1,1	-0.65661	0.08545	-7.68	1
AR1,1	-0.31902	0.10726	-2.97	1

Constant Estimate = 0.00118737
 Variance Estimate = 0.26700634
 Std Error Estimate = 0.51672656
 AIC = 661.538656*
 SBC = 673.75779*
 Number of Residuals = 434
 * Does not include log determinant.

Autocorrelation Check of Residuals

To	Chi	DF	Prob	Autocorrelations							
Lag	Square			6	12	18	24	30	36	42	48
6	19.81	4	0.001	0.005	-0.012	-0.106	-0.037	0.081	-0.159		
12	49.71	10	0.000	-0.151	0.082	0.173	0.024	0.011	-0.083		
18	91.46	16	0.000	-0.019	0.218	-0.176	0.068	0.041	0.085		
24	116.80	22	0.000	-0.000	-0.203	-0.112	0.007	-0.003	-0.041		
30	125.85	28	0.000	-0.009	0.043	0.104	0.043	-0.069	-0.012		
36	130.82	34	0.000	-0.065	0.007	-0.066	-0.022	-0.033	-0.019		
42	131.09	40	0.000	0.001	0.016	-0.012	0.010	0.008	0.003		
48	137.03	46	0.000	0.012	0.002	-0.020	0.023	-0.104	0.017		

Si les estimations semblent significative (*compte tenu du niveau de la T-stat*), le résidu ne semble pas être un bruit blanc. L'étape suivante est donc d'augmenter le nombre de paramètres.

9.2.3 Estimation des paramètres d'une modélisation ARIMA (2, 1, 2)

L'estimation donne les résultats suivants (*la constante étant là aussi non significative*),

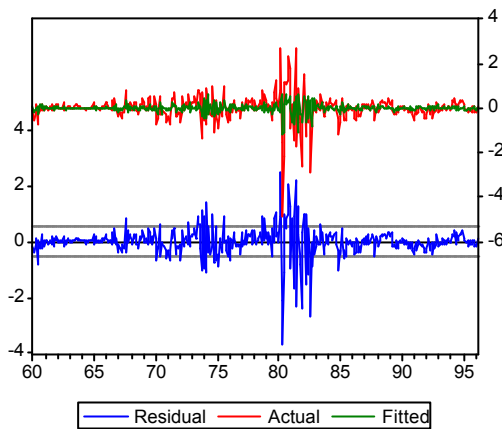
$$\left(\begin{matrix} 1 - 0.564L - 0.125L^2 \\ (0.195) \quad (0.141) \end{matrix} \right) Y_t = \left(\begin{matrix} 1 - 0.238L - 0.461L^2 \\ (0.186) \quad (0.156) \end{matrix} \right) \varepsilon_t$$

LS // Dependent Variable is Y
 Sample: 1960:01 1996:03
 Included observations: 435
 Convergence achieved after 11 iterations

Variable	Coefficient	Std. Error	T-Statistic	Prob.
AR(1)	0.563991	0.195947	2.878290	0.0042
AR(2)	0.125419	0.140515	0.892566	0.3726
MA(1)	-0.238277	0.185762	-1.282700	0.2003
MA(2)	-0.461184	0.155933	-2.957584	0.0033
R-squared	0.119815	Mean dependent var	0.000892	
Adjusted R-squared	0.113689	S.D. dependent var	0.546661	
S.E. of regression	0.514649	Akaike info criterion	-1.319386	
Sum squared resid	114.1564	Schwartz criterion	-1.281911	
Log likelihood	-326.2718	F-statistic	19.55661	
Durbin-Watson stat	1.986863	Prob(F-statistic)	0.000000	
Inverted AR Roots	.73	-.17		
Inverted MA Roots	.81	-.57		

Sample: 1960:01 1996:03
 Included observations: 435
 Q-statistic probabilities adjusted for 4 ARMA term[s]

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.006	0.006	0.0176		
2	0.027	0.027	0.3468		
3	-0.043	-0.044	1.1762		
4	-0.030	-0.030	1.5759		
5	0.120	0.123	7.9408	0.005	
6	-0.141	-0.146	16.716	0.000	
7	-0.123	-0.133	23.392	0.000	
8	0.089	0.119	26.891	0.000	
9	0.174	0.186	40.344	0.000	
10	0.018	-0.038	40.496	0.000	
11	0.035	0.054	41.037	0.000	
12	-0.075	-0.038	43.540	0.000	
13	-0.014	-0.078	43.630	0.000	
14	0.225	0.228	66.395	0.000	
15	-0.178	-0.121	80.770	0.000	
16	0.073	0.069	83.218	0.000	
17	0.034	0.077	83.746	0.000	
18	0.067	0.025	95.768	0.000	
19	0.004	-0.117	95.777	0.000	
20	-0.202	-0.120	104.46	0.000	
21	-0.107	-0.075	109.68	0.000	



Conditional Least Squares Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag
μ	0.0018425	0.02470	0.07	0
MA1,1	0.23910	0.20740	1.15	1
MA1,2	0.39060	0.17076	2.29	2
AR1,1	0.56668	0.21504	2.64	1
AR1,2	0.06125	0.14731	0.42	2

Constant Estimate = 0.00068555
 Variance Estimate = 0.26832254
 Std Error Estimate = 0.51606448
 AIC = 662.407167*
 SBC = 682.77239*
 Number of Residuals = 434
 * Does not include log determinant.

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	17.05	2	0.000	0.002	-0.022	-0.029	-0.038	0.123	-0.144
12	43.83	8	0.000	-0.128	0.089	-0.168	0.012	0.039	-0.079
18	86.23	14	0.000	-0.013	0.226	-0.181	0.072	0.035	0.059
24	111.58	20	0.000	0.008	-0.205	-0.105	0.000	-0.016	-0.047
30	118.83	26	0.000	-0.012	0.028	0.093	0.031	-0.070	-0.013
36	124.99	32	0.000	-0.070	-0.001	-0.071	-0.031	-0.035	-0.030
42	125.20	38	0.000	-0.004	0.008	-0.016	0.005	0.006	-0.006
48	131.58	44	0.000	0.013	-0.009	-0.023	0.019	-0.108	0.016

A titre d'information, le modèle avec constante s'écrit

$$\left(\begin{matrix} 1 - 0.565L - 0.129L^2 \\ (0.194) \quad (0.140) \end{matrix} \right) Y_t = 0.004 + \left(\begin{matrix} 1 - 0.239L - 0.465L^2 \\ (0.183) \quad (0.154) \end{matrix} \right) \varepsilon_t$$

Encore une fois, l'hypothèse de bruit blanc des résidus est rejetée, de part la présence d'autocorrélations significativement non nulles.

9.2.4 Estimation des paramètres d'une modélisation ARIMA (4, 1, 4)

L'estimation donne les résultats suivants (*la constante étant là aussi non significative*),

$$\left(\begin{matrix} 1 - 0.656L + 0.563L^2 - 0.386L^3 + 0.456L^4 \\ (0.102) \quad (0.118) \quad (0.118) \quad (0.098) \end{matrix} \right) Y_t = \left(\begin{matrix} 1 - 0.341L + 0.254L^2 - 0.179L^3 - 0.732L^4 \\ (0.079) \quad (0.084) \quad (0.083) \quad (0.077) \end{matrix} \right) \varepsilon_t$$

LS // Dependent Variable is Y
 Sample: 1960:01 1996:03
 Included observations: 435
 Convergence achieved after 21 iterations

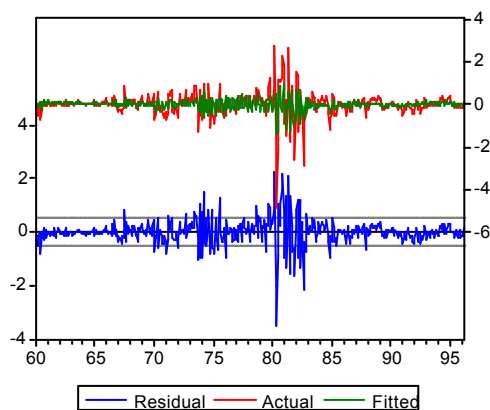
Variable	Coefficient	Std. Error	T-Statistic	Prob.
AR(1)	0.656518	0.101944	6.440006	0.0000
AR(2)	-0.562635	0.118552	-4.745877	0.0000
AR(3)	0.385954	0.118100	3.268017	0.0012
AR(4)	0.456434	0.098600	4.629142	0.0000
MA(1)	-0.341849	0.079426	-4.304007	0.0000
MA(2)	0.254008	0.084136	3.018997	0.0027
MA(3)	-0.179371	0.083193	-2.156092	0.0316
MA(4)	-0.732405	0.076658	-9.554205	0.0000

R-squared 0.186601 Mean dependent var 0.000892
 Adjusted R-squared 0.173266 S.D. dependent var 0.546661
 S.E. of regression 0.497051 Akaike info criterion -1.379905
 Sum squared resid 105.4946 Schwartz criterion -1.304956
 Log likelihood -309.1089 F-statistic 13.99393
 Durbin-Watson stat 1.966731 Prob(F-statistic) 0.000000

Inverted AR Roots .98 .08+.97i .08-.97i -.49
 Inverted MA Roots 1.00 .04-.99i .04+.99i -.75

Sample: 1960:01 1996:03
 Included observations: 435
 Q-statistic probabilities adjusted for 8 ARMA term[s]

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.016	0.016	0.1100		
2	0.027	0.027	0.4404		
3	-0.036	-0.037	1.0075		
4	-0.029	-0.029	1.3817		
5	-0.010	-0.007	1.4220		
6	-0.144	-0.144	10.627		
7	-0.068	-0.067	12.686		
8	0.129	0.140	20.082		
9	0.083	0.076	23.157	0.009	
10	0.004	-0.021	23.163	0.000	
11	0.101	0.105	27.746	0.000	
12	-0.019	-0.030	27.915	0.000	
13	-0.081	-0.113	30.842	0.000	
14	0.166	0.229	43.271	0.000	
15	-0.112	-0.076	48.899	0.000	
16	0.142	0.120	58.074	0.000	
17	0.014	0.049	58.167	0.000	
18	0.023	0.006	58.402	0.000	
19	0.014	-0.047	58.493	0.000	
20	-0.123	-0.085	65.388	0.000	
21	-0.121	-0.101	72.143	0.000	



Conditional Least Squares Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag
μ	0.0015377	0.02530	0.06	0
MA1,1	-0.93247	0.20932	-4.45	1
MA1,2	-0.51607	0.34370	-1.50	2
MA1,3	0.33273	0.33343	1.00	3
MA1,4	0.32631	0.16789	1.94	4
AR1,1	-0.56650	0.21310	-2.66	1
AR1,2	-0.37411	0.28332	-1.32	2
AR1,3	0.29264	0.25505	1.15	3
AR1,4	-0.05048	0.12451	-0.41	4

Constant Estimate = 0.00261178
 Variance Estimate = 0.24960404
 Std Error Estimate = 0.49960369
 AIC = 638.20434*
 SBC = 674.86174*
 Number of Residuals = 434

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	0.00	0	0.000	0.001	-0.006	-0.001	0.049	0.004	-0.100
12	25.70	4	0.000	-0.085	0.003	0.189	0.037	-0.029	-0.016
18	53.83	10	0.000	-0.013	0.184	-0.123	0.045	0.007	0.105
24	75.58	16	0.000	-0.038	-0.192	-0.082	-0.020	-0.017	-0.041
30	82.92	22	0.000	-0.017	0.032	0.087	0.036	-0.074	-0.005
36	88.91	28	0.000	-0.060	-0.013	-0.075	-0.019	-0.050	-0.021

9.2.5 Estimation des paramètres d'une modélisation ARIMA (8, 1, 2)

L'estimation donne les résultats suivants,

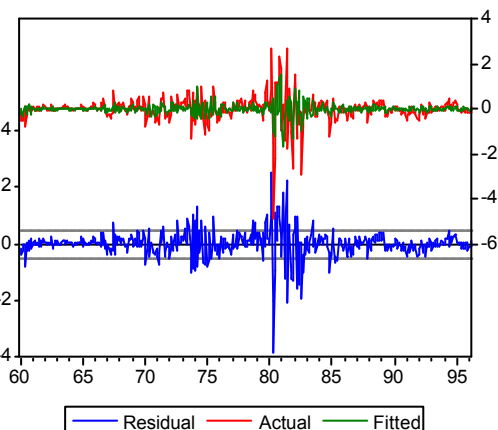
LS // Dependent Variable is Y
 Sample: 1960:01 1996:03
 Included observations: 435
 Convergence achieved after 6 iterations

Variable	Coefficient	Std. Error	T-Statistic	Prob.
AR(1)	0.832927	0.234792	3.547511	0.0004
AR(2)	-0.713119	0.268505	-2.655889	0.0082
AR(3)	0.227474	0.126585	1.769058	0.0776
AR(4)	-0.182781	0.087076	-2.099087	0.0364
AR(5)	0.196717	0.076248	2.579957	0.0102
AR(6)	-0.326733	0.089358	-3.656453	0.0003
AR(7)	0.085217	0.115162	0.739972	0.4597
AR(8)	0.063868	0.087090	0.733359	0.4637
MA(1)	-0.506487	0.232037	-2.182785	0.0296
MA(2)	0.375754	0.223857	1.678544	0.0940

R-squared	0.205297	Mean dependent var	0.000892
Adjusted R-squared	0.188468	S.D. dependent var	0.546661
S.E. of regression	0.492460	Akaike info criterion	-1.393963
Sum squared resid	103.0698	Schwartz criterion	-1.300277
Log likelihood	-304.0513	F-statistic	12.19898
Durbin-Watson stat	1.997998	Prob(F-statistic)	0.000000

Sample: 1960:01 1996:03
 Included observations: 435
 Q-statistic probabilities adjusted for 10 ARMA term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.001	0.001	0.0003		
2	0.003	0.003	0.0031		
3	-0.004	-0.004	0.0097		
4	0.004	0.004	0.0154		
5	0.013	0.013	0.0871		
6	-0.029	-0.029	0.4656		
7	0.034	0.034	0.9821		
8	-0.010	-0.010	1.0283		
9	-0.013	-0.013	1.0995		
10	-0.013	-0.013	1.1777		
11	0.131	0.132	8.9133	0.003	
12	-0.019	-0.022	9.0716	0.011	
13	-0.077	-0.077	11.766	0.008	
14	0.191	0.197	28.202	0.000	
15	-0.128	-0.141	35.618	0.000	
16	0.096	0.103	39.782	0.000	
17	0.035	0.048	40.339	0.000	
18	0.025	0.004	40.620	0.000	
19	-0.060	-0.067	42.271	0.000	
20	-0.155	-0.139	53.293	0.000	
21	-0.077	-0.094	55.990	0.000	



Conditional Least Squares Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag
MU	0.0013416	0.02589	0.05	0
MA1,1	0.48361	0.24886	1.94	1
MA1,2	-0.30774	0.24844	-1.24	2
AR1,1	0.80989	0.24937	3.25	1
AR1,2	-0.63535	0.29763	-2.13	2
AR1,3	0.19796	0.13780	1.44	3
AR1,4	-0.16464	0.08924	-1.84	4
AR1,5	0.19163	0.07493	2.56	5
AR1,6	-0.30877	0.09036	-3.42	6
AR1,7	0.06239	0.11887	0.52	7
AR1,8	0.09180	0.09070	1.01	8

Constant Estimate = 0.00101303
 Variance Estimate = 0.24414457
 Std Error Estimate = 0.49410988
 AIC = 630.559134*
 SBC = 675.362624*
 Number of Residuals = 434

Autocorrelation Check of Residuals

Lag	To	Chi Square	DF	Prob	Autocorrelations
12	8.79	2	0.012	0.038	-0.018 -0.014 -0.013 -0.015 -0.022
24	56.98	14	0.000	-0.060 -0.157 -0.076 -0.017 -0.015 -0.042	
36	67.16	26	0.000	-0.069 -0.013 -0.059 -0.001 -0.059 -0.061	

9.2.6 Estimation des paramètres d'une modélisation ARIMA (8, 1, 4)

L'estimation donne les résultats suivants,

LS // Dependent Variable is Y
 Sample: 1960:01 1996:03
 Included observations: 435
 Convergence achieved after 10 iterations

Variable	Coefficient	Std. Error	T-Statistic	Prob.
AR(1)	0.453658	0.145520	3.117492	0.0019
AR(2)	-0.766228	0.135070	-5.697621	0.0000
AR(3)	0.313465	0.110000	2.849606	0.0046
AR(4)	-0.792703	0.113391	-6.990908	0.0000
AR(5)	0.334490	0.076348	4.381140	0.0000
AR(6)	-0.366035	0.073980	-4.947751	0.0000
AR(7)	-0.017904	0.076594	-0.233757	0.8153
AR(8)	0.035879	0.074208	0.483495	0.6290
MA(1)	-0.111388	0.137724	-0.808774	0.4191
MA(2)	0.562973	0.130136	4.326037	0.0000
MA(3)	-0.136900	0.092114	-1.486189	0.1380
MA(4)	0.630228	0.092707	6.798074	0.0000

R-squared	0.227772	Mean dependent var	0.000892
Adjusted R-squared	0.207691	S.D. dependent var	0.546661
S.E. of regression	0.486593	Akaike info criterion	-1.413457
Sum squared resid	100.1548	Schwartz criterion	-1.301034
Log likelihood	-297.8114	F-statistic	11.34236
Durbin-Watson stat	2.002816	Prob(F-statistic)	0.000000

Sample: 1960:01 1996:03
 Included observations: 435
 Q-statistic probabilities adjusted for 12 ARMA term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	-0.002	-0.002	0.0010		
2	0.003	0.003	0.0051		
3	0.011	0.011	0.0565		
4	-0.020	-0.020	0.2255		
5	0.018	0.018	0.3643		
6	-0.032	-0.032	0.8199		
7	0.058	0.059	2.3369		
8	-0.078	-0.079	5.0151		
9	0.062	0.065	6.7438		
10	-0.009	-0.013	6.7804		
11	0.014	0.020	6.8637		
12	0.032	0.024	7.3314		
13	-0.005	0.005	7.3411	0.007	
14	0.189	0.181	23.515	0.000	
15	-0.109	-0.105	28.929	0.000	
16	0.049	0.045	30.011	0.000	
17	-0.011	-0.011	30.069	0.000	
18	0.057	0.068	31.564	0.000	
19	-0.054	-0.073	32.896	0.000	
20	-0.161	-0.149	44.833	0.000	
21	-0.038	-0.067	45.496	0.000	

9.2.7 Choix du modèle

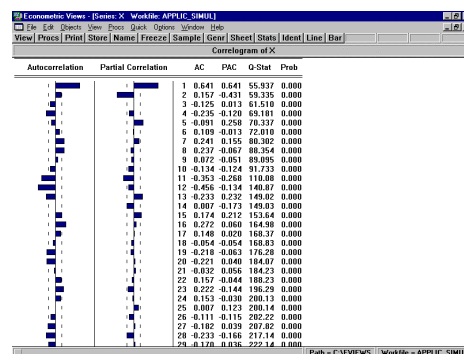
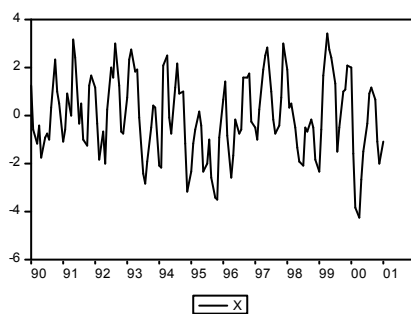
Les différents critères sont

	σ	R^2	\overline{R}^2	F-stat	AIC
ARIMA(1, 1, 1)	0.515	0.113	0.111	55.082	-1.3207
ARIMA(2, 1, 2)	0.515	0.120	0.114	19.556	-1.3193
ARIMA(4, 1, 4)	0.497	0.186	0.173	13.994	-1.3049
ARIMA(8, 1, 2)	0.492	0.205	0.188	12.199	-1.3939
ARIMA(8, 1, 4)	0.487	0.228	0.208	11.342	-1.4134

Aucun modèle ne semble vraiment bien modéliser la série. En fait, aucun modèle de type ARMA ne pourra prendre en compte le pic de volatilité au début des années 80. En fait, nous le verrons par la suite, les modèles ARCH permettent de prendre en compte ce genre de comportement.

9.3 Application à des données simulées

Considérons désormais la courbe suivante, représentant des données mensuelles, sur la période Janvier 1990-Janvier 2001.



Au vu de l'allure de la série (à gauche) et de son corrélogramme (à droite), on peut affirmer que le processus X_t est stationnaire. Nous allons donc essayer de modéliser X_t à l'aide d'un processus ARMA(p, q). Au regard de l'autocorrélogramme partiel, une modélisation à l'aide d'un processus AR(2) semblerait possible. Les sorties ci-dessous correspondent à l'estimation des paramètres d'un processus AR(2), respectivement avec et sans constante. La constante ne semblant pas significative, on peut retenir l'idée d'un modèle AR(2) sans constante.

Variable	Coefficient	Std. Error	T-Statistic	Prob.
C	-0.836662	0.198298	-0.184885	0.8536
AR[1]	0.929812	0.079039	11.76269	0.0000
AR[2]	-0.439617	0.079250	-5.547230	0.0000

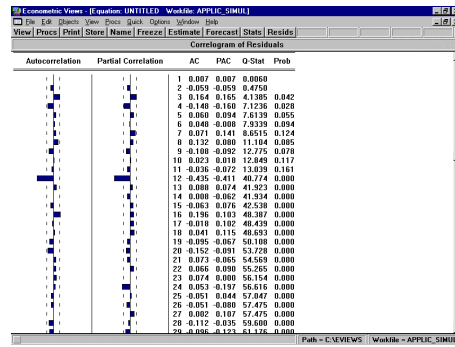
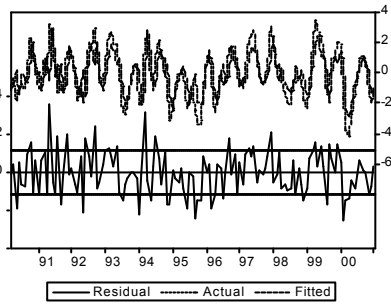
R-squared: 0.531651 Mean dependent var: -0.054480
 Adjusted R-squared: 0.524333 S.D. dependent var: 1.677402
 S.E. of regression: 1.159382 Akaike info criterion: 0.314091
 Sum squared resid: 171.3120 Schwarz criterion: 0.379935
 Log likelihood: -293.4538 F-statistic: 72.65010
 Durbin-Watson stat: 1.985067 Prob(F-statistic): 0.000000
 Inverted AR Roots: .46+ .47i .46 - .47i

Variable	Coefficient	Std. Error	T-Statistic	Prob.
AR[1]	0.938291	0.078714	11.81742	0.0000
AR[2]	-0.439636	0.078951	-5.559197	0.0000

R-squared: 0.531526 Mean dependent var: -0.054480
 Adjusted R-squared: 0.527894 S.D. dependent var: 1.677402
 S.E. of regression: 1.152543 Akaike info criterion: 0.290890
 Sum squared resid: 171.3077 Schwarz criterion: 0.345386
 Log likelihood: -293.4714 F-statistic: 146.3610
 Durbin-Watson stat: 1.985313 Prob(F-statistic): 0.000000
 Inverted AR Roots: .47+ .47i .47 - .47i

On peut faire plusieurs remarques sur cette estimation :

- les paramètres sont significatifs,
- le modèle est validé par le test de Fisher (*statistique F valant 146*)
- le Durbin-Watson est proche de 2 : on rejette l'hypothèse d'autocorrélation à l'ordre 1 des résidus. Au regard de l'autocorrélogramme des résidus, présenté ci-dessous, on peut en revanche noter que les résidus sont autocorrélés à l'ordre 12.

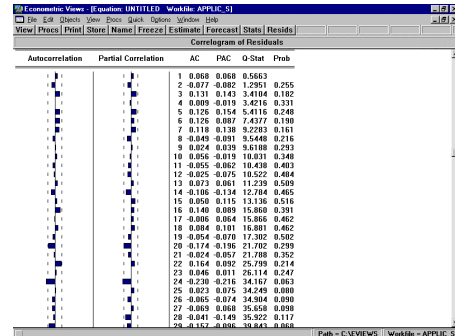


Sur cette série des résidus $\hat{\varepsilon}_t$, il est possible de tester un modèle $AR(12)$,

EViews software output for an AR(12) model:

Variable	Coefficient	Std. Error	T-Statistic	Prob.
AR(12)	-0.473034	0.081873	-5.777684	0.0000

Additional statistics: R-squared: 0.229389, Adjusted R-squared: 0.228389, S.E. of regression: 1.026810, Sum squared resid: 124.8485, Log likelihood: -171.7083.

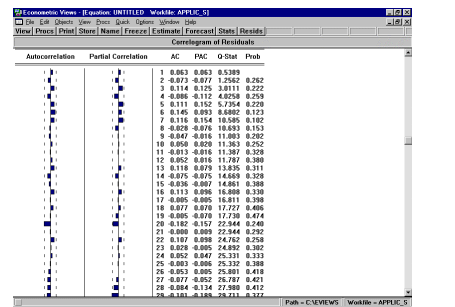


dont l'autocorrélogramme des résidus suit un bruit blanc : l'hypothèse de modèle $AR(12)$ pour $\hat{\varepsilon}_t$. Il est aussi possible de tester un modèle $MA(12)$,

EViews software output for an MA(12) model:

Variable	Coefficient	Std. Error	T-Statistic	Prob.
MA(12)	-0.785858	0.061825	-11.41795	0.0000

Additional statistics: R-squared: 0.298689, Adjusted R-squared: 0.298089, S.E. of regression: 0.981386, Sum squared resid: 120.1566, Log likelihood: -169.2217.



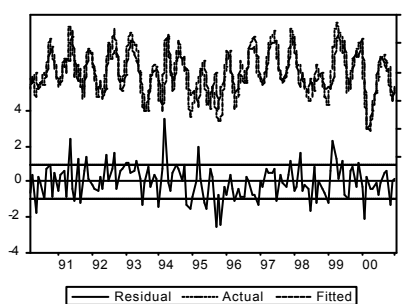
Cette modélisation $MA(12)$ est alors validée. Nous avons donc vu qu'en écrivant $\Phi(L)X_t = \varepsilon_t$ le processus ε_t pouvait se mettre sous la forme $\varepsilon_t = \Theta(L)\eta_t$, et donc $\Phi(L)X_t = \Theta(L)\eta_t$.

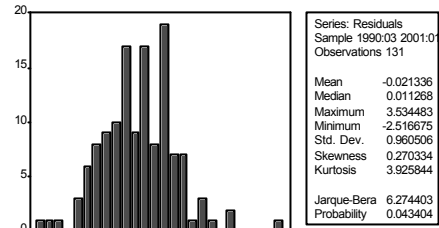
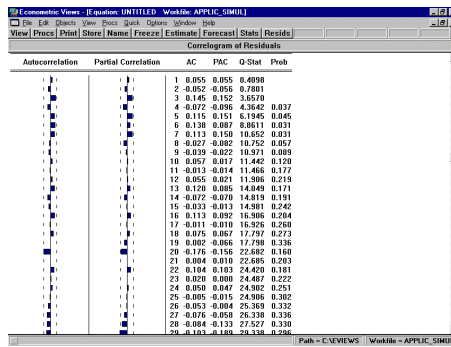
On peut alors tester un modèle $ARMA(2, 1)$, avec 2 retards sur la composante AR et un terme d'ordre 12 (*uniquement*) pour la composante MA . L'estimation est donnée ci-dessous, en haut à gauche et les autres graphiques donnent des informations sur les résidus,

EViews software output for an ARMA(2, 12) model:

Variable	Coefficient	Std. Error	T-Statistic	Prob.
AR(1)	0.941131	0.077793	12.09880	0.0000
AR(2)	-0.463480	0.077840	-5.93918	0.0000
MA(12)	-0.711461	0.061851	-11.53826	0.0000

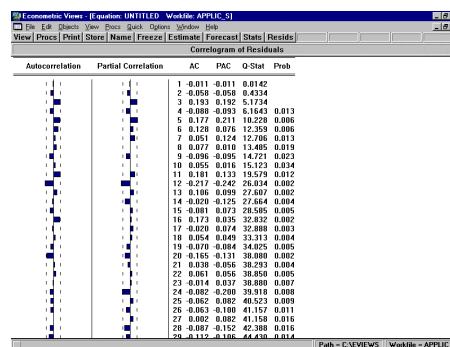
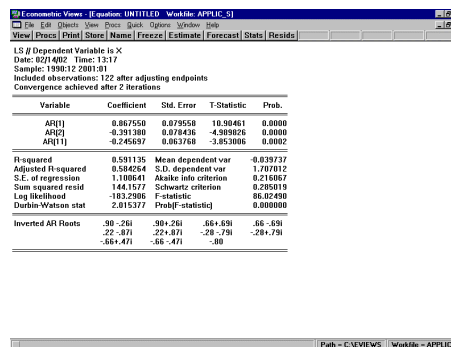
Additional statistics: R-squared: 0.671349, Adjusted R-squared: 0.658873, S.E. of regression: 0.908222, Sum squared resid: 115.0941, Log likelihood: -180.1330, Durbin-Watson stat: 1.895324.





Le modèle semble être validé. On peut d'ailleurs noter que les racines des polynômes AR et MA sont distinctes deux à deux, et à l'extérieur du disque unité (*EViews donne l'inverse des racines, qui doivent être à l'intérieur du disque unité*).

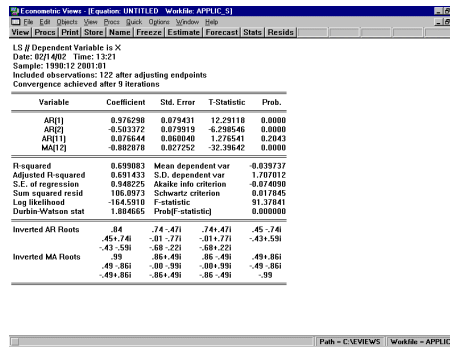
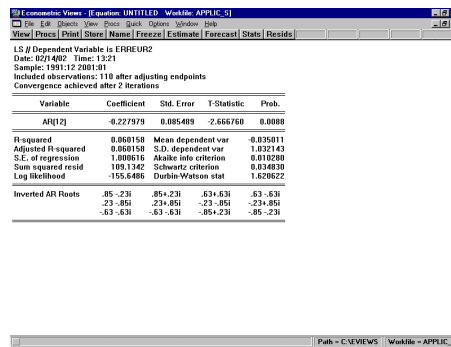
L'autocorrélogramme aurait aussi pu suggérer un modèle avec un retard d'ordre 11 sur la série initiale,



Comme on le voit sur la sortie (*ci-dessus à gauche*) la composante AR d'ordre 11 est significativement non nulle. Toutefois, comme auparavant, les résidus ne suivent pas un bruit blanc, l'autocorrélation partielle d'ordre 12 étant significativement non nulle, on peut alors tester un modèle AR(12) sur cette série. Le modèle pouvant être retenu, on peut alors modéliser la série X_t par un équation de la forme

$$X_t = \alpha X_{t-1} + \beta X_{t-2} + \gamma X_{t-11} + \varepsilon_t - \theta \varepsilon_{t-12}$$

Toutefois, la sortie ci-dessous à gauche montre que la composante autocorrégressive d'ordre 11 n'est pas significative. On en revient alors à considérer le modèle précédent, $X_t = \alpha X_{t-1} + \beta X_{t-2} + \varepsilon_t - \theta \varepsilon_{t-12}$



Les sorties ci-dessous correspondent à l'estimation du même modèle sous SAS,

```

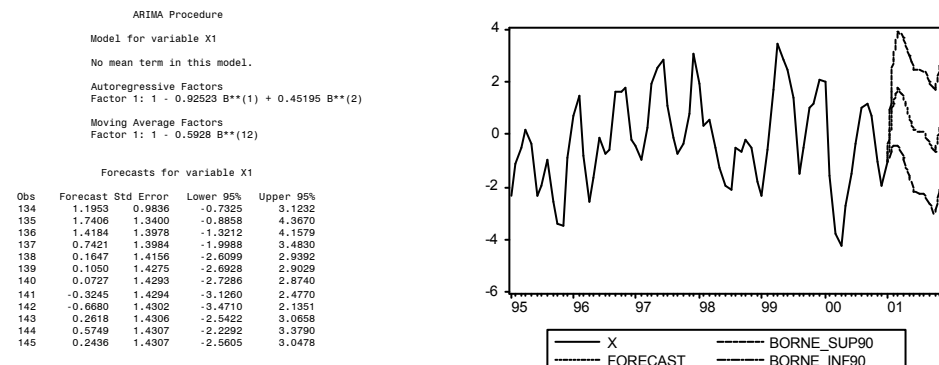
ARIMA Procedure
Lag Covariance Correlation
0 2.765782 1.00000
1 1.779821 0.64127
2 0.435508 -0.15744
3 -0.347034 -0.12547
4 -0.649323 -0.23477
5 -0.251108 -0.09079
6 0.300811 0.10876
7 0.667198 0.24123
8 0.654841 0.23677
9 0.197798 0.07152
10 -0.371829 -0.13444
11 -0.076420 -0.35304
12 -1.259860 -0.45552
13 -0.645783 -0.23349
14 0.018853 0.00682
15 0.481153 0.17397
16 0.751913 0.27186
17 0.409392 0.14802
18 -0.149954 -0.05422
19 -0.601785 -0.21758
20 -0.612205 -0.22135
21 -0.088030 -0.03183
22 0.434812 0.15721
23 0.614522 0.22219
24 0.422268 0.15268
Std
0 0.086711
0.117059
0.118640
0.119634
0.123049
0.123552
0.124270
0.127742
0.131000
0.131293
0.132324
0.139226
0.150014
0.152722
0.152724
0.154207
0.157769
0.158910
0.158949
0.161173
0.163443
0.163489
0.164622
0.166862
Inverse Autocorrelations
Lag Correlation
1 -0.77526
2 0.52645
3 -0.40775
4 0.36394
5 -0.31446
6 0.24728
7 -0.25439
8 0.23272
9 -0.19539
10 0.19393
11 -0.25485
12 0.33300
13 -0.23752
14 0.10711
15 -0.01011
16 -0.02764
17 0.02954
18 -0.05536
19 0.03948
20 -0.01601
21 0.03232
22 -0.04912
23 0.02409
24 0.00858
Autocorrelation Check for White Noise
To Ch1 Autocorrelations
Lag Square DF Prob
6 72.01 6 0.000 0.641 0.157 -0.125 -0.235 -0.091 0.109
12 140.87 12 0.000 0.241 0.237 0.072 -0.134 -0.353 -0.456
18 168.83 18 0.000 -0.233 0.007 0.174 0.272 0.148 -0.054
24 200.13 24 0.000 -0.218 -0.221 -0.032 0.157 0.222 0.153
ARIMA Procedure
Conditional Least Squares Estimation
Iteration SSE MA1,1 AR1,1 AR1,2 Lambda R Crit
0 133.59976 0.4555 0.9177 -0.4310 0.00001 1
1 130.16169 0.6174 0.9204 -0.4411 1E-6 0.162007
2 130.10177 0.5998 0.9180 -0.4457 1E-7 0.023948
3 130.10044 0.6024 0.9184 -0.4452 1E-8 0.003393
4 130.10042 0.6020 0.9183 -0.4452 1E-9 0.000525
Maximum Likelihood Estimation
Iter Loglike MA1,1 AR1,1 AR1,2 Lambda R Crit
0 -188.11308 0.6020 0.9183 -0.4452 0.00001 1
1 -188.10091 0.5914 0.9253 -0.4525 1E-6 0.01455
2 -188.10067 0.5931 0.9252 -0.4519 1E-7 0.002016
3 -188.10066 0.5928 0.9252 -0.4520 1E-8 0.000329
ARIMA Estimation Optimization Summary
Estimation Method: Maximum Likelihood
Parameters Estimated: 3
Termination Criteria: Maximum Relative Change in Estimates
Iteration Stopping Value: 0.001
Criteria Value: 0.00048293
Alternate Criteria: Relative Change in Objective Function
Alternate Criteria Value: 4.91058E-6
Maximum Absolute Value of Gradient: 0.04331511
R-Square (Relative Change in Regression SSE) from Last Iteration
Step: 0.00032913
Objective Function: Log Gaussian Likelihood
Objective Function Value: -188.10066
Marquardt's Lambda Coefficient: 1E-8
Numerical Derivative Perturbation Delta: 0.001
Iterations: 3
Maximum Likelihood Estimation
Parameter Estimate Approx. Std Error T Ratio Lag
MA1,1 0.59280 0.08028 7.38 12
AR1,1 0.92523 0.07630 12.13 1
AR1,2 -0.45195 0.07619 -5.93 2
Variance Estimate = 0.96749844
Std Error Estimate = 0.98361499
AIC = 382.20133
SBC = 390.872377
Number of Residuals= 133
    
```

Là aussi, le modèle est validé (avec des valeurs légèrement différentes des estimateurs, venant des méthodes de calculs différentes sous SAS et EViews). Le modèle retenu est alors le suivant

$$X_t = 0.92523X_{t-1} - 0.45195X_{t-2} + \varepsilon_t - 0.5928\varepsilon_{t-12} \text{ où } \varepsilon_t \sim BB(0, 0.984)$$

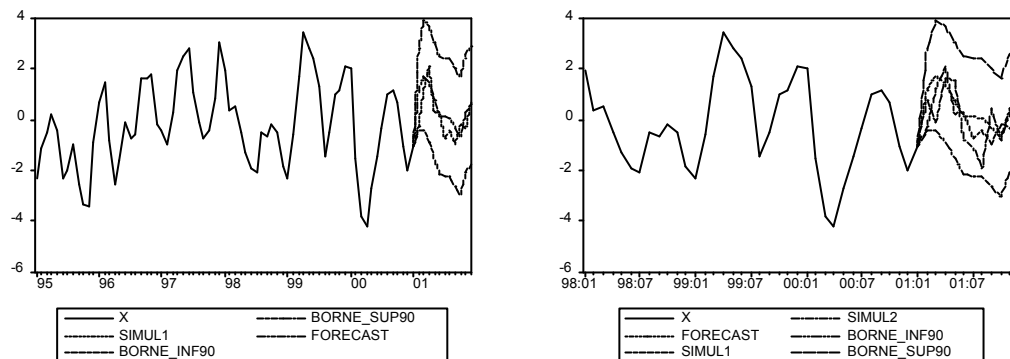
(0.0763) (0.0762) (0.0803)

SAS donne également des prévisions (avec intervalle de confiance à 90%) de (X_t) . Ces prévisions ont été représentées sur le graphique ci-dessous à droite,



Les graphiques ci-dessous montrent le processus (X_t) , sa prévision (avec intervalle de confiance), et 1 puis 2 simulations

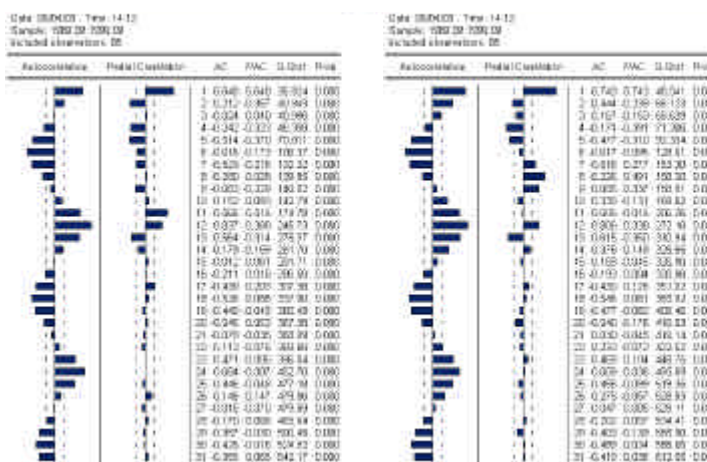
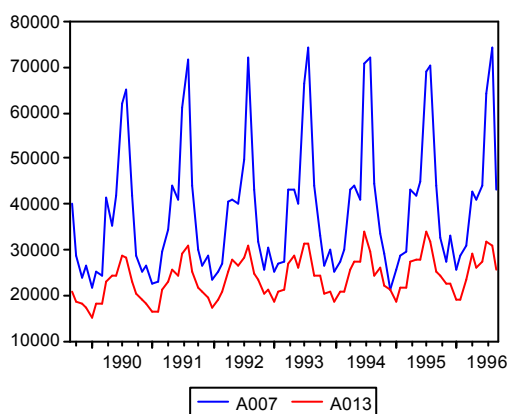
du processus prolongé jusqu'à fin 2001,



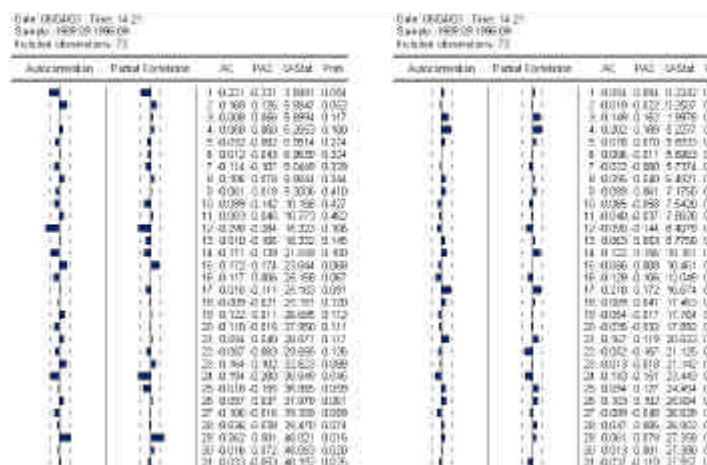
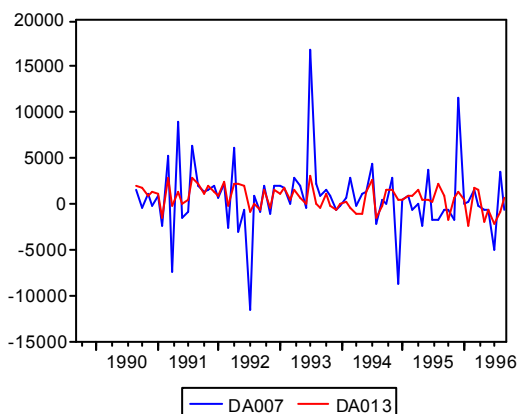
Comme on peut le noter, la prévision à l'aide de processus *ARMA* donne de relativement bons résultats.

9.4 Modélisation du trafic autoroutier

Les deux séries suivantes correspondent à du trafic automobile, sur deux autoroutes françaises, l'A7 et l'A13, en données mensuelles,



Les deux séries sont certes, très corrélées entre elles (*non développeront ce point dans la partie suivante, sur la modélisation des séries multivariées*), mais elles sont surtout cycliques, comme le montrent les autocorrélogrammes. Nous allons donc travailler ici sur les séries $Y_t = (1 - L^{12}) X_t = X_t - X_{t-12}$,



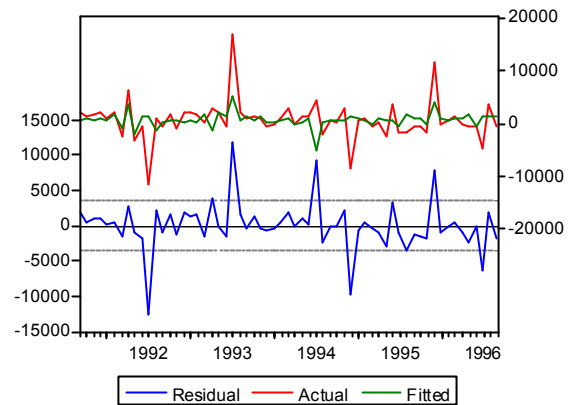
9.4.1 Modélisation du trafic sur l'autoroute A7

- Modélisation à l'aide d'un processus AR(12)

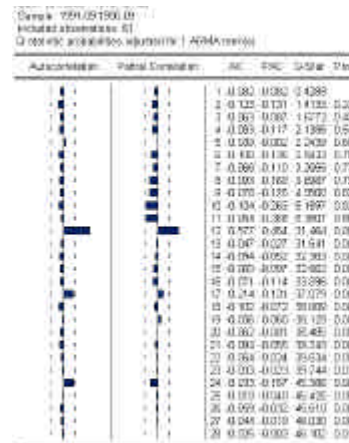
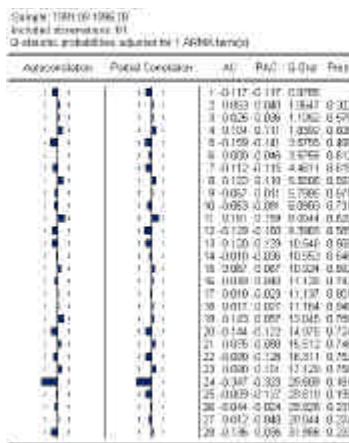
Tentons tout d'abord une modélisation ne faisant intervenir que le retard à l'ordre 12, on obtient alors

Dependent Variable: DA007
 Method: Least Squares
 Date: 07/18/03 Time: 07:44
 Sample(adjusted): 1991:09 1996:09
 Included observations: 61 after adjusting endpoints
 Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	664.4503	331.7515	2.002856	0.0498
AR(12)	-0.349683	0.119648	-2.922612	0.0049
R-squared	0.126465	Mean dependent var	645.1803	
Adjusted R-squared	0.111659	S.D. dependent var	3709.991	
S.E. of regression	3496.734	Akaike info criterion	19.18928	
Sum squared resid	7.21E+08	Schwarz criterion	19.25849	
Log likelihood	-583.2731	F-statistic	8.541663	
Durbin-Watson stat	2.226113	Prob(F-statistic)	0.004914	
Inverted AR Roots	.88+.24i	.88-.24i	.65-.65i	.65-.65i
	.24+.88i	.24-.88i	-.24-.88i	-.24+.88i
	-.65+.65i	-.65-.65i	-.88-.24i	-.88+.24i



La constante et la composante $AR(12)$ sont ici toutes deux significatives. De plus, l'autocorrélogramme des résidus (présenté ci-dessous à gauche) ainsi que le test de Box Pierce conduisent à accepter l'hypothèse de bruit blanc (au sens L^2) des résidus. Néanmoins, si l'on regarde la série des résidus, présentée ci-dessus à droite, on peut noter que certains pics périodiques apparaissent, mais le signe de ces pics ne permet pas de mettre en avant une autocorrélation éventuelle de ces résidus. Mais si l'on regarde l'autocorrélogramme du carré des résidus, on peut noter que l'hypothèse de bruit blanc au sens fort est à rejeter : les carrés des résidus sont clairement autocorrélés. Nous sommes ici en présence d'un effet $GARCH$ sur les résidus, notion que nous étudierons plus en détails par la suite.



9.4.2 Modélisation du trafic sur l'autoroute A13

- Modélisation par un bruit blanc

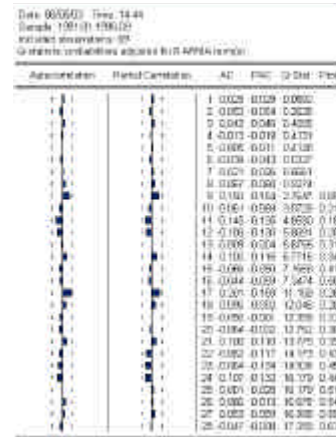
Compte tenu de l'autocorrélogramme de la série 'désaisonnalisée' (i.e. $Y_t = (1 - L^{12}) X_t = X_t - X_{t-12}$), l'hypothèse de bruit blanc (au sens L^2) est vérifiée, autrement dit, la série (Y_t) est purement aléatoire.

- Modélisation à l'aide d'un processus ARMA(4, 4)

Afin d'améliorer la prévision, il peut être intéressant de tester un modèle $ARMA(4, 4)$ puisque ce sont les dernières valeurs qui pourraient sembler significatives sur les autocorrélogrammes.

Dependent Variable: DA013
 Method: Least Squares
 Date: 06/05/03 Time: 14:43
 Sample(adjusted): 1991:01 1996:09
 Included observations: 69 after adjusting endpoints
 Convergence achieved after 22 iterations
 Backcast: 1990:09 1990:12

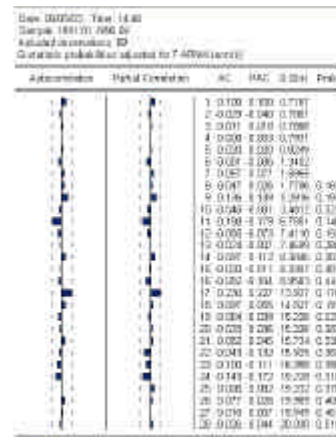
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	542.3256	255.2653	2.124557	0.0378
AR(1)	0.231052	0.120712	1.914081	0.0604
AR(2)	0.354618	0.112041	3.165075	0.0024
AR(3)	0.260947	0.116229	2.245119	0.0285
AR(4)	-0.378479	0.121658	-3.110988	0.0029
MA(1)	-0.222646	0.020711	-10.75022	0.0000
MA(2)	-0.480386	0.023240	-20.67077	0.0000
MA(3)	-0.207426	0.022297	-9.303057	0.0000
MA(4)	0.934617	0.026384	35.40974	0.0000
R-squared	0.340761	Mean dependent var	613.8261	
Adjusted R-squared	0.252862	S.D. dependent var	1264.917	
S.E. of regression	1093.358	Akaike info criterion	16.95300	
Sum squared resid	71725932	Schwarz criterion	17.24441	
Log likelihood	-575.8786	F-statistic	3.876750	
Durbin-Watson stat	1.933998	Prob(F-statistic)	0.000974	
Inverted AR Roots	.69+ .31i	.69- .31i	-.58+ .58i	-.58- .58i
Inverted MA Roots	.83+ .52i	.83- .52i	-.72+ .66i	-.72- .66i



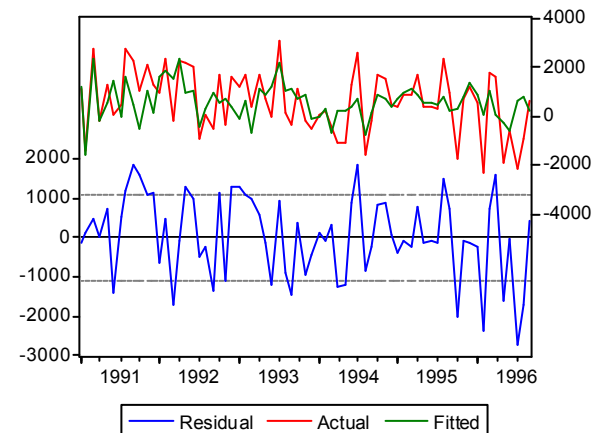
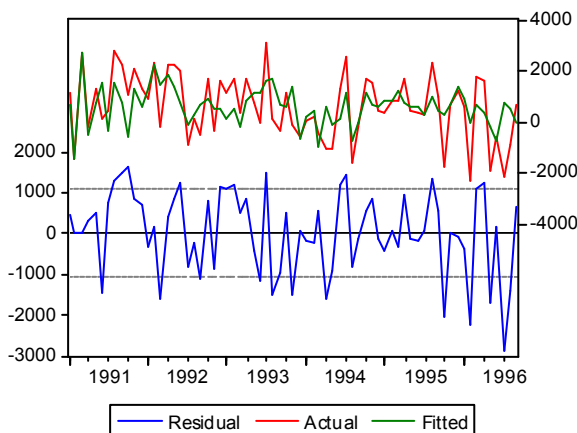
On peut noter que tous les retards semblent significatifs, et le bruit associé vérifie l'hypothèse de bruit blanc : un modèle $ARMA(4, 4)$ convient. Toutefois, en toute rigueur, le premier retard de sa composante d'autocorrélation ($AR(1)$) ne passe pas le seuil de 5% : il est possible d'essayer de l'enlever de la régression. On obtient alors

Dependent Variable: DA013
 Method: Least Squares
 Date: 06/05/03 Time: 14:46
 Sample(adjusted): 1991:01 1996:09
 Included observations: 69 after adjusting endpoints
 Convergence achieved after 16 iterations
 Backcast: 1990:09 1990:12

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	520.8973	242.9550	2.144007	0.0360
AR(2)	0.539645	0.108087	4.992716	0.0000
AR(3)	0.270148	0.101843	2.652590	0.0102
AR(4)	-0.404489	0.108105	-3.741632	0.0004
MA(1)	-0.120464	0.013922	-8.652958	0.0000
MA(2)	-0.660822	0.015333	-43.09873	0.0000
MA(3)	-0.103716	0.013349	-7.769263	0.0000
MA(4)	0.954459	0.016257	58.71185	0.0000
R-squared	0.308872	Mean dependent var	613.8261	
Adjusted R-squared	0.229562	S.D. dependent var	1264.917	
S.E. of regression	1110.276	Akaike info criterion	16.97126	
Sum squared resid	75195450	Schwarz criterion	17.23028	
Log likelihood	-577.5083	F-statistic	3.894504	
Durbin-Watson stat	1.797891	Prob(F-statistic)	0.001440	
Inverted AR Roots	.68- .30i	.68+ .30i	-.68+ .54i	-.68- .54i
Inverted MA Roots	.84+ .53i	.84- .53i	-.78+ .60i	-.78- .60i



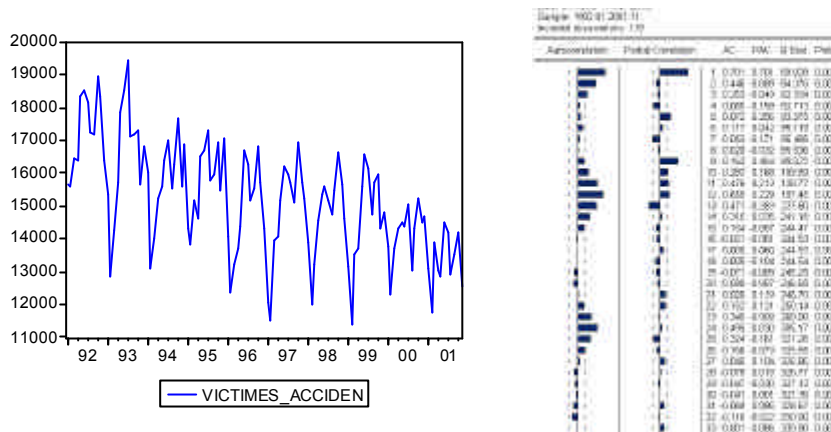
où, cette fois-ci, tous les coefficients sont significatifs au seuil de 5%. Et là aussi, l'hypothèse de bruit blanc pour les résidus est validée. Les résidus des deux régressions sont présentés ci-dessous, de gauche à droite,



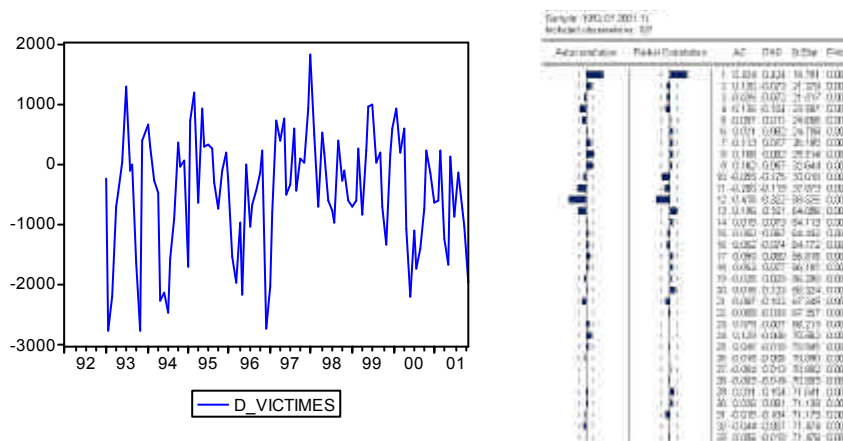
[A INSERER]

9.5 Modélisation du nombre de victimes sur les routes

Considérons la série suivante, représentant le nombre de victimes sur les routes, en France, en données mensuelles,



avec la série brute (X_t) à gauche, et son autocorrélogramme à droite. Cette série présentant clairement un comportement saisonnier, nous allons étudier $Y_t = (1 - L^{12}) X_t = X_t - X_{t-12}$,



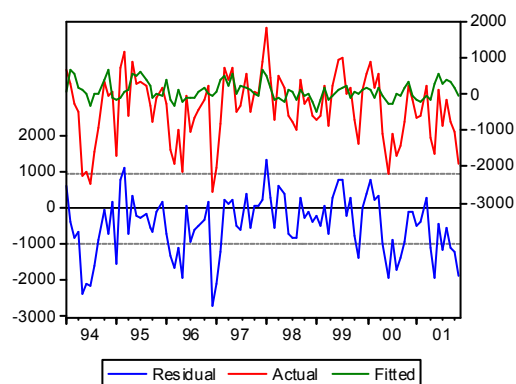
En utilisant l'autocorrélogramme de cette série, nous allons modéliser (Y_t) à l'aide d'un processus $ARMA(p, q)$.

- **Modélisation à l'aide d'un processus AR(12)**

Compte tenu de la forme de l'autocorrélogramme, nous pouvons, par le principe de parcimonie, tenter une modélisation relativement simple, à l'aide d'un processus $AR(12)$. Une modélisation à l'aide d'un polynôme $1 - \phi L^{12}$ donne les résultats suivants,

Dependent Variable: D_VICTIMES
 Method: Least Squares
 Date: 07/18/03 Time: 06:59
 Sample(adjusted): 1994:01 2001:11
 Included observations: 95 after adjusting endpoints
 Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(12)	-0.252692	0.092678	-2.726572	0.0076
R-squared	-0.092661	Mean dependent var	-387.5368	
Adjusted R-squared	-0.092661	S.D. dependent var	920.6434	
S.E. of regression	962.3526	Akaike info criterion	16.58711	
Sum squared resid	87055511	Schwarz criterion	16.61399	
Log likelihood	-786.8877	Durbin-Watson stat	0.808405	
Inverted AR Roots	.86 +.23i	.86 -.23i	.63 -.63i	.63+.63i
	.23 -.86i	.23+.86i	-.23+.86i	-.23 -.86i
	-.63 -.63i	-.63 -.63i	-.86 -.23i	-.86+.23i



Le paramètre ϕ est effectivement significatif, et le résidu semble correspondre à un bruit blanc (hypothèse validée par le test de Box Pierce), toutefois, la qualité de l'ajustement ne peut manifestement pas convenir. L'estimation d'un

modèle AR(12) complet (avec les 12 composantes), en enlevant ensuite, une à une, les composantes non-significatives (une des étapes étant proposée à droite)

Dependent Variable: D_VICTIMES
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-403.1510	88.13528	-4.574229	0.0000
AR(1)	0.373193	0.100646	3.707971	0.0004
AR(2)	-0.028210	0.105584	-0.267185	0.7900
AR(3)	0.041577	0.104470	0.397978	0.6917
AR(4)	0.003752	0.103359	0.036302	0.9711
AR(5)	-0.008166	0.103532	-0.078870	0.9373
AR(6)	0.012129	0.103316	0.117397	0.9068
AR(7)	0.062211	0.104500	0.595317	0.5533
AR(8)	-0.052965	0.104848	-0.505165	0.6148
AR(9)	0.133121	0.105012	1.267679	0.2085
AR(10)	-0.075842	0.105460	-0.719149	0.4741
AR(11)	-0.011995	0.104314	-0.114987	0.9087
AR(12)	-0.357608	0.095975	-3.726044	0.0004

R-squared 0.378505 Mean dependent var -387.5368
Adjusted R-squared 0.287555 S.D. dependent var 920.6434
S.E. of regression 777.0825 Akaike info criterion 16.27550
Sum squared resid 49516291 Schwarz criterion 16.62497
Log likelihood -760.0861 F-statistic 4.161668
Durbin-Watson stat 1.851256 Prob(F-statistic) 0.000044

Inverted AR Roots
.93 +.22i .93 -.22i .69+.66i .69-.66i
.25 -.89i .25+.89i -.23 -.87i -.23+.87i
-.58 -.64i -.58+.64i -.88+.27i -.88-.27i

Dependent Variable: D_VICTIMES
Method: Least Squares
Date: 07/18/03 Time: 07:25
Sample(adjusted): 1994:01 2001:11
Included observations: 95 after adjusting endpoints
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-404.4320	87.86986	-4.613125	0.0000
AR(1)	0.362837	0.090355	4.015707	0.0001
AR(3)	0.031491	0.085447	0.368541	0.7134
AR(7)	0.049055	0.083384	0.588299	0.5578
AR(9)	0.112071	0.093093	1.203859	0.2319
AR(10)	-0.083440	0.092096	-0.906003	0.3674
AR(12)	-0.355504	0.083454	-4.259889	0.0001

R-squared 0.375645 Mean dependent var -387.5368
Adjusted R-squared 0.333075 S.D. dependent var 920.6434
S.E. of regression 751.8479 Akaike info criterion 16.15377
Sum squared resid 49744218 Schwarz criterion 16.34195
Log likelihood -760.3042 F-statistic 8.824230
Durbin-Watson stat 1.826335 Prob(F-statistic) 0.000000

Inverted AR Roots
.93+.22i .93-.22i .69-.66i .69+.66i
.24+.88i .24-.88i -.21+.86i -.21-.86i
-.59+.65i -.59-.65i -.88+.27i -.88-.27i

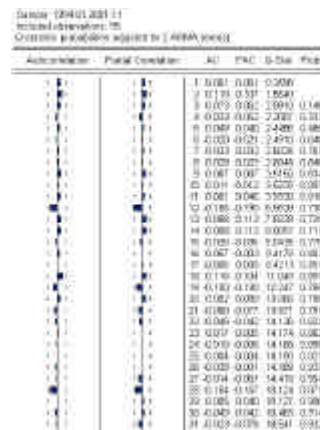
Le modèle final, où toutes les composantes sont significative comporte toujours le retard d'ordre 12, ainsi que le retard d'ordre 1, ainsi que la constante, que nous avons omise auparavant,

Dependent Variable: D_VICTIMES
Method: Least Squares
Date: 07/18/03 Time: 07:26
Sample(adjusted): 1994:01 2001:11
Included observations: 95 after adjusting endpoints
Convergence achieved after 3 iterations

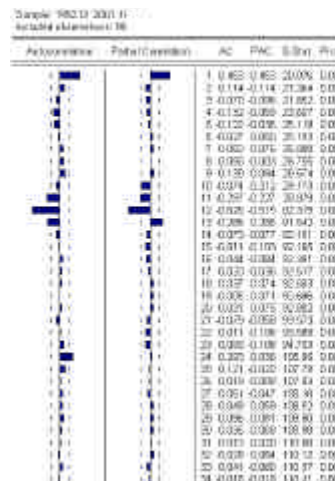
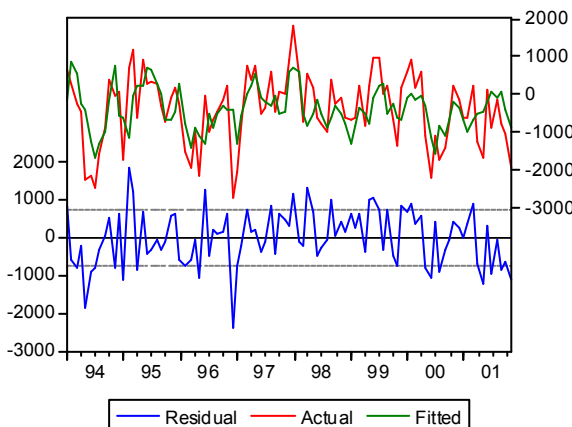
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-398.4806	76.31964	-5.221207	0.0000
AR(1)	0.367802	0.086803	4.237205	0.0001
AR(12)	-0.370033	0.079309	-4.665704	0.0000

R-squared 0.359326 Mean dependent var -387.5368
Adjusted R-squared 0.345398 S.D. dependent var 920.6434
S.E. of regression 744.8694 Akaike info criterion 16.09536
Sum squared resid 51044394 Schwarz criterion 16.17601
Log likelihood -761.5298 F-statistic 25.79936
Durbin-Watson stat 1.842573 Prob(F-statistic) 0.000000

Inverted AR Roots
.93 -.24i .93+.24i .69-.65i .69+.65i
.27+.88i .27-.88i -.21+.88i -.21-.88i
-.62+.65i -.62-.65i -.86-.24i -.86+.24i



Comme le montre le graphique de droite, l'hypothèse de bruit blanc des résidus est validée par le test de Box Pierce. De plus, l'ajustement est nettement meilleur que dans le cas précédent, comme le montre le graphique ci-dessous à gauche.



Remarque 60 Au vu de l'autocorrélogramme de (Y_t) on pourrait croire reconnaître un processus SARIMA d'ordre 12 (cd section sur les modèles SARIMA). Mais si l'on différencie (Y_t) en notant $Z_t = (1 - L^{12})Y_t$, on obtient l'autocorrélogramme présenté ci-dessus à droite : la différenciation n'apporte rien.

- Modélisation à l'aide d'un processus ARMA (12,12)

Compte tenu de la présence d'une autocorrélation partielle et d'une autocorrélation significatives à l'ordre 12, il pourrait être possible, également, de tester un modèle $ARMA(12, 12)$. On inclus pour cela les 24 retards, et la constante, comme cela est fait ci-dessous, à gauche,

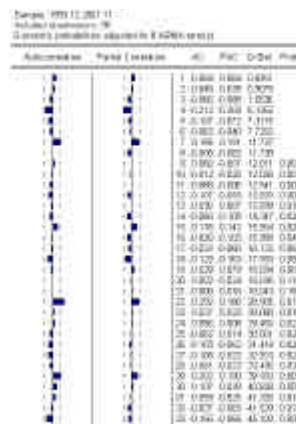
Dependent Variable: D_VICTIMES
Backcast: 1993:01 1993:12

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-305.0056	99.72524	-3.058459	0.0030
AR(1)	0.420150	0.099681	4.214941	0.0001
AR(3)	-0.031261	0.109530	-0.285409	0.7761
AR(4)	0.275419	0.134000	2.053772	0.0431
AR(5)	-0.113776	0.118891	-0.956980	0.3415
AR(7)	0.309327	0.097592	3.169605	0.0022
AR(8)	-0.036742	0.101708	-0.361249	0.7189
AR(11)	-0.214746	0.115685	-1.853090	0.0676
AR(12)	-0.134458	0.109928	-1.223142	0.2249
MA(3)	0.058559	0.094135	0.622072	0.5357
MA(4)	-0.368560	0.115378	-3.194377	0.0020
MA(5)	-0.063851	0.065686	-0.972065	0.3340
MA(7)	-0.428437	0.090813	-4.717804	0.0000
MA(8)	-0.225824	0.097233	-2.322506	0.0228
MA(11)	0.398335	0.111243	3.580766	0.0006
MA(12)	-0.307450	0.109088	-2.818742	0.0061
R-squared	0.511111	Mean dependent var	-387.5368	
Adjusted R-squared	0.418285	S.D. dependent var	920.6434	
S.E. of regression	702.1773	Akaike info criterion	16.05896	
Sum squared resid	3661191	Schwarz criterion	16.52979	
Log likelihood	-748.6864	F-statistic	5.506068	
Durbin-Watson stat	1.989013	Prob(F-statistic)	0.000000	
Inverted AR Roots	.93 + .17i	.93 - .17i	.65 - .69i	.65 + .69i
			.16 - .84i	.21 - .88i
			.65 - .39i	.67 - .17i
			.67 - .17i	.67 + .17i
Inverted MA Roots	.99	.62 - .28i	.62 + .28i	.57 + .76i
	.57 - .76i	.01 - .99i	.01 - .99i	.36 - .82i
	.36 + .82i	.85 + .48i	.85 - .48i	.99

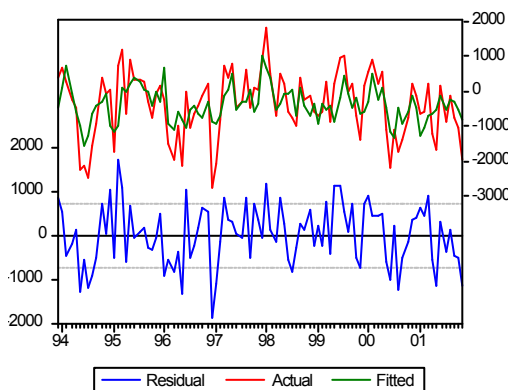
Après élimination, une à une, des variables non significatives, on aboutit au modèle suivant,

Dependent Variable: D_VICTIMES
Method: Least Squares
Date: 07/18/03 Time: 19:56
Sample(adjusted): 1993:12 2001:11
Included observations: 96 after adjusting endpoints
Convergence achieved after 10 iterations
Backcast: 1992:12 1993:11

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-361.4454	70.05417	-5.159513	0.0000
AR(1)	0.323792	0.083297	3.887199	0.0002
AR(4)	0.221787	0.123963	1.790575	0.0768
AR(7)	0.360634	0.095223	3.787262	0.0003
AR(11)	-0.223918	0.118132	-1.895485	0.0613
MA(4)	-0.243286	0.126263	-1.926817	0.0573
MA(7)	-0.367712	0.080979	-4.540834	0.0000
MA(11)	0.290547	0.127322	2.281979	0.0249
MA(12)	-0.464710	0.081240	-5.720233	0.0000
R-squared	0.432562	Mean dependent var	-379.4479	
Adjusted R-squared	0.380384	S.D. dependent var	919.2082	
S.E. of regression	723.5614	Akaike info criterion	16.09531	
Sum squared resid	45548073	Schwarz criterion	16.33572	
Log likelihood	-763.5748	F-statistic	8.229081	
Durbin-Watson stat	1.627925	Prob(F-statistic)	0.000000	
Inverted AR Roots	.88 - .10i	.88 + .10i	.61 - .70i	.61 + .70i
	.08 - .84i	.08 + .84i	-.24 - .83i	-.24 + .83i
	-.77 + .43i	-.77 - .43i		-.80
Inverted MA Roots	.97	.74 - .39i	.74 + .39i	.52 - .79i
	.52 + .79i	-.00 - .96i	-.00 + .96i	-.39 + .86i
	-.39 - .86i	.87 - .48i	.87 + .48i	-.96



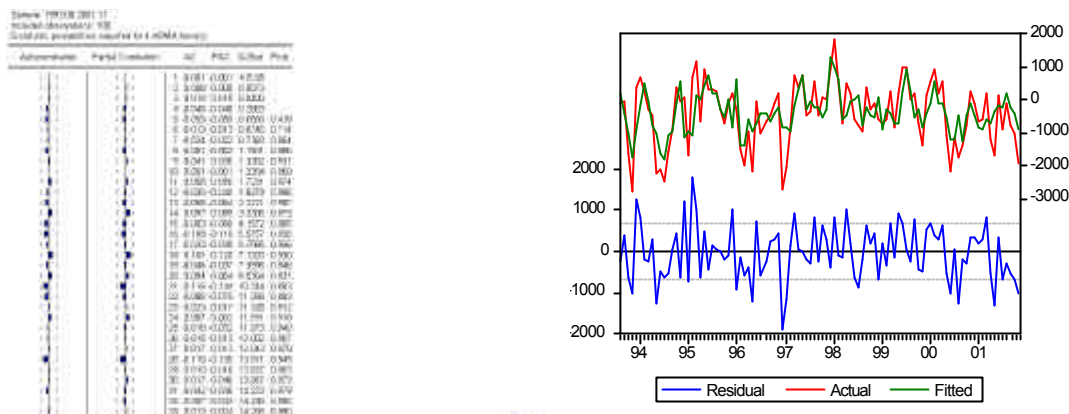
Néanmoins, on peut noter que les résidus associés ne suivent pas un bruit blanc, comme le montre l'autocorrélogramme ci-dessus à droite. Les graphiques associés sont présentés ci-dessous à gauche. En éliminant les variables non significatives dans le modèle précédent (p -valeur supérieure à 5%), on obtient la sortie ci-dessous à droite,



Dependent Variable: D_VICTIMES
Method: Least Squares
Date: 07/18/03 Time: 19:57
Sample(adjusted): 1993:08 2001:11
Included observations: 100 after adjusting endpoints
Convergence achieved after 15 iterations
Backcast: 1992:08 1993:07

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-285.5781	58.35439	-4.893858	0.0000
AR(1)	0.429447	0.077826	5.518073	0.0000
AR(7)	0.212892	0.092498	2.301586	0.0235
MA(7)	-0.240872	0.075975	-3.170413	0.0020
MA(12)	-0.656765	0.075974	-8.644549	0.0000
R-squared	0.499844	Mean dependent var	-408.9800	
Adjusted R-squared	0.478785	S.D. dependent var	940.4888	
S.E. of regression	678.9883	Akaike info criterion	15.92779	
Sum squared resid	43797389	Schwarz criterion	16.05805	
Log likelihood	-791.3896	F-statistic	23.73516	
Durbin-Watson stat	1.969953	Prob(F-statistic)	0.000000	
Inverted AR Roots	.88	.57 - .61i	.57 + .61i	-.12 - .77i
		-.12 + .77i	-.67 + .34i	-.67 - .34i
Inverted MA Roots	.99	.81 + .48i	.81 - .48i	.51 - .84i
	.51 + .84i	-.02 + .97i	-.02 - .97i	-.46 + .84i
	-.46 - .84i	-.86 - .48i	-.86 + .48i	-.94

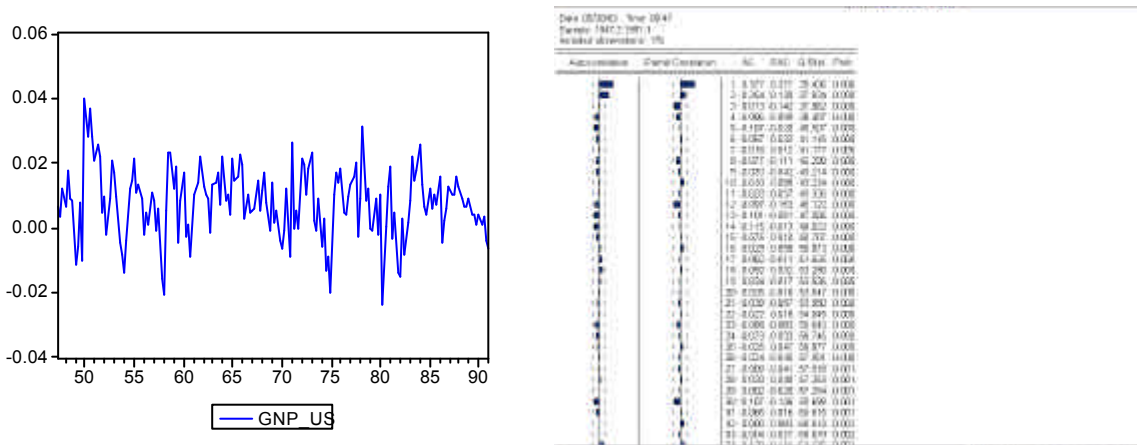
c'est à dire que (Y_t) est modélisable par un processus $ARMA(7, 12)$. En effet, l'hypothèse de bruit blanc des résidus est validée,



[A CONTINUER]

9.6 Modélisation du taux de croissance du PIB américain

La série suivante correspond au taux de croissance trimestriel du PIB américain entre 1948 et 1991,



Compte tenu de la forme des autocorrélations, il est possible de tester un modèle $AR(3)$, i.e.

Dependent Variable: GNP_US
 Method: Least Squares
 Date: 05/28/03 Time: 09:43
 Sample(adjusted): 1948:1 1991:1
 Included observations: 173 after adjusting endpoints
 Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.007682	0.001229	6.251827	0.0000
AR(1)	0.350924	0.076268	4.601216	0.0000
AR(2)	0.180937	0.079722	2.269604	0.0245
AR(3)	-0.144305	0.076504	-1.886233	0.0610
R-squared	0.176968	Mean dependent var		0.007748
Adjusted R-squared	0.162358	S.D. dependent var		0.010811
S.E. of regression	0.009894	Akaike info criterion		-6.370866
Sum squared resid	0.016545	Schwarz criterion		-6.297957
Log likelihood	555.0799	F-statistic		12.11276
Durbin-Watson stat	2.019838	Prob(F-statistic)		0.000000
Inverted AR Roots	.44+.29i	.44-.29i	-.52	

Date: 05/28/03 Time: 09:43
 Sample: 1948:1 1991:1
 Included observations: 173

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.3509	0.3509	0.3509	0.0000	0.0000
2	0.1809	0.1809	0.0000	0.0000	0.0000
3	-0.1443	-0.1443	0.0000	0.0000	0.0000
4	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.0000	0.0000	0.0000	0.0000	0.0000
9	0.0000	0.0000	0.0000	0.0000	0.0000
10	0.0000	0.0000	0.0000	0.0000	0.0000
11	0.0000	0.0000	0.0000	0.0000	0.0000
12	0.0000	0.0000	0.0000	0.0000	0.0000
13	0.0000	0.0000	0.0000	0.0000	0.0000
14	0.0000	0.0000	0.0000	0.0000	0.0000
15	0.0000	0.0000	0.0000	0.0000	0.0000
16	0.0000	0.0000	0.0000	0.0000	0.0000
17	0.0000	0.0000	0.0000	0.0000	0.0000
18	0.0000	0.0000	0.0000	0.0000	0.0000
19	0.0000	0.0000	0.0000	0.0000	0.0000
20	0.0000	0.0000	0.0000	0.0000	0.0000
21	0.0000	0.0000	0.0000	0.0000	0.0000
22	0.0000	0.0000	0.0000	0.0000	0.0000
23	0.0000	0.0000	0.0000	0.0000	0.0000
24	0.0000	0.0000	0.0000	0.0000	0.0000
25	0.0000	0.0000	0.0000	0.0000	0.0000
26	0.0000	0.0000	0.0000	0.0000	0.0000
27	0.0000	0.0000	0.0000	0.0000	0.0000
28	0.0000	0.0000	0.0000	0.0000	0.0000
29	0.0000	0.0000	0.0000	0.0000	0.0000
30	0.0000	0.0000	0.0000	0.0000	0.0000
31	0.0000	0.0000	0.0000	0.0000	0.0000
32	0.0000	0.0000	0.0000	0.0000	0.0000
33	0.0000	0.0000	0.0000	0.0000	0.0000
34	0.0000	0.0000	0.0000	0.0000	0.0000
35	0.0000	0.0000	0.0000	0.0000	0.0000
36	0.0000	0.0000	0.0000	0.0000	0.0000
37	0.0000	0.0000	0.0000	0.0000	0.0000
38	0.0000	0.0000	0.0000	0.0000	0.0000
39	0.0000	0.0000	0.0000	0.0000	0.0000
40	0.0000	0.0000	0.0000	0.0000	0.0000
41	0.0000	0.0000	0.0000	0.0000	0.0000
42	0.0000	0.0000	0.0000	0.0000	0.0000
43	0.0000	0.0000	0.0000	0.0000	0.0000
44	0.0000	0.0000	0.0000	0.0000	0.0000
45	0.0000	0.0000	0.0000	0.0000	0.0000
46	0.0000	0.0000	0.0000	0.0000	0.0000
47	0.0000	0.0000	0.0000	0.0000	0.0000
48	0.0000	0.0000	0.0000	0.0000	0.0000
49	0.0000	0.0000	0.0000	0.0000	0.0000
50	0.0000	0.0000	0.0000	0.0000	0.0000
51	0.0000	0.0000	0.0000	0.0000	0.0000
52	0.0000	0.0000	0.0000	0.0000	0.0000
53	0.0000	0.0000	0.0000	0.0000	0.0000
54	0.0000	0.0000	0.0000	0.0000	0.0000
55	0.0000	0.0000	0.0000	0.0000	0.0000
56	0.0000	0.0000	0.0000	0.0000	0.0000
57	0.0000	0.0000	0.0000	0.0000	0.0000
58	0.0000	0.0000	0.0000	0.0000	0.0000
59	0.0000	0.0000	0.0000	0.0000	0.0000
60	0.0000	0.0000	0.0000	0.0000	0.0000
61	0.0000	0.0000	0.0000	0.0000	0.0000
62	0.0000	0.0000	0.0000	0.0000	0.0000
63	0.0000	0.0000	0.0000	0.0000	0.0000
64	0.0000	0.0000	0.0000	0.0000	0.0000
65	0.0000	0.0000	0.0000	0.0000	0.0000
66	0.0000	0.0000	0.0000	0.0000	0.0000
67	0.0000	0.0000	0.0000	0.0000	0.0000
68	0.0000	0.0000	0.0000	0.0000	0.0000
69	0.0000	0.0000	0.0000	0.0000	0.0000
70	0.0000	0.0000	0.0000	0.0000	0.0000
71	0.0000	0.0000	0.0000	0.0000	0.0000
72	0.0000	0.0000	0.0000	0.0000	0.0000
73	0.0000	0.0000	0.0000	0.0000	0.0000
74	0.0000	0.0000	0.0000	0.0000	0.0000
75	0.0000	0.0000	0.0000	0.0000	0.0000
76	0.0000	0.0000	0.0000	0.0000	0.0000
77	0.0000	0.0000	0.0000	0.0000	0.0000
78	0.0000	0.0000	0.0000	0.0000	0.0000
79	0.0000	0.0000	0.0000	0.0000	0.0000
80	0.0000	0.0000	0.0000	0.0000	0.0000
81	0.0000	0.0000	0.0000	0.0000	0.0000
82	0.0000	0.0000	0.0000	0.0000	0.0000
83	0.0000	0.0000	0.0000	0.0000	0.0000
84	0.0000	0.0000	0.0000	0.0000	0.0000
85	0.0000	0.0000	0.0000	0.0000	0.0000
86	0.0000	0.0000	0.0000	0.0000	0.0000
87	0.0000	0.0000	0.0000	0.0000	0.0000
88	0.0000	0.0000	0.0000	0.0000	0.0000
89	0.0000	0.0000	0.0000	0.0000	0.0000
90	0.0000	0.0000	0.0000	0.0000	0.0000
91	0.0000	0.0000	0.0000	0.0000	0.0000
92	0.0000	0.0000	0.0000	0.0000	0.0000
93	0.0000	0.0000	0.0000	0.0000	0.0000
94	0.0000	0.0000	0.0000	0.0000	0.0000
95	0.0000	0.0000	0.0000	0.0000	0.0000
96	0.0000	0.0000	0.0000	0.0000	0.0000
97	0.0000	0.0000	0.0000	0.0000	0.0000
98	0.0000	0.0000	0.0000	0.0000	0.0000
99	0.0000	0.0000	0.0000	0.0000	0.0000
100	0.0000	0.0000	0.0000	0.0000	0.0000
101	0.0000	0.0000	0.0000	0.0000	0.0000
102	0.0000	0.0000	0.0000	0.0000	0.0000
103	0.0000	0.0000	0.0000	0.0000	0.0000
104	0.0000	0.0000	0.0000	0.0000	0.0000
105	0.0000	0.0000	0.0000	0.0000	0.0000
106	0.0000	0.0000	0.0000	0.0000	0.0000
107	0.0000	0.0000	0.0000	0.0000	0.0000
108	0.0000	0.0000	0.0000	0.0000	0.0000
109	0.0000	0.0000	0.0000	0.0000	0.0000
110	0.0000	0.0000	0.0000	0.0000	0.0000
111	0.0000	0.0000	0.0000	0.0000	0.0000
112	0.0000	0.0000	0.0000	0.0000	0.0000
113	0.0000	0.0000	0.0000	0.0000	0.0000
114	0.0000	0.0000	0.0000	0.0000	0.0000
115	0.0000	0.0000	0.0000	0.0000	0.0000
116	0.0000	0.0000	0.0000	0.0000	0.0000
117	0.0000	0.0000	0.0000	0.0000	0.0000
118	0.0000	0.0000	0.0000	0.0000	0.0000
119	0.0000	0.0000	0.0000	0.0000	0.0000
120	0.0000	0.0000	0.0000	0.0000	0.0000
121	0.0000	0.0000	0.0000	0.0000	0.0000
122	0.0000	0.0000	0.0000	0.0000	0.0000
123	0.0000	0.0000	0.0000	0.0000	0.0000
124	0.0000	0.0000	0.0000	0.0000	0.0000
125	0.0000	0.0000	0.0000	0.0000	0.0000
126	0.0000	0.0000	0.0000	0.0000	0.0000
127	0.0000	0.0000	0.0000	0.0000	0.0000
128	0.0000	0.0000	0.0000	0.0000	0.0000
129	0.0000	0.0000	0.0000	0.0000	0.0000
130	0.0000	0.0000	0.0000	0.0000	0.0000
131	0.0000	0.0000	0.0000	0.0000	0.0000
132	0.0000	0.0000	0.0000	0.0000	0.0000
133	0.0000	0.0000	0.0000	0.0000	0.0000
134	0.0000	0.0000	0.0000	0.0000	0.0000
135	0.0000	0.0000	0.0000	0.0000	0.0000
136	0.0000	0.0000	0.0000	0.0000	0.0000
137	0.0000	0.0000	0.0000	0.0000	0.0000
138	0.0000	0.0000	0.0000	0.0000	0.0000
139	0.0000	0.0000	0.0000	0.0000	0.0000
140	0.0000	0.0000	0.0000	0.0000	0.0000
141	0.0000	0.0000	0.0000	0.0000	0.0000
142	0.0000	0.0000	0.0000	0.0000	0.0000
143	0.0000	0.0000	0.0000	0.0000	0.0000
144	0.0000	0.0000	0.0000	0.0000	0.0000
145	0.0000	0.0000	0.0000	0.0000	0.0000
146	0.0000	0.0000	0.0000	0.0000	0.0000
147	0.0000	0.0000	0.0000	0.0000	0.0000
148	0.0000	0.0000	0.0000	0.0000	0.0000
149	0.0000	0.0000	0.0000	0.0000	0.0000
150	0.0000	0.0000	0.0000	0.0000	0.0000
151	0.0000	0.0000	0.0000	0.0000	0.0000
152	0.0000	0.0000	0.0000	0.0000	0.0000
153	0.0000	0.0000	0.0000	0.0000	0.0000
154	0.0000	0.0000	0.0000	0.0000	0.0000
155	0.0000	0.0000	0.0000	0.0000	0.0000
156	0.0000	0.0000	0.0000	0.0000	0.0000
157	0.0000	0.0000	0.0000	0.0000	0.0000
158	0.0000	0.0000	0.0000	0.0000	0.0000
159	0.0000	0.0000	0.0000	0.0000	0.0000
160	0.0000	0.0000	0.0000	0.0000	

En reprenant ce qui a été fait dans la partie (???) sur les modèles $AR(2)$, on peut noter que le polynôme autorégressif s'écrit

$$(1 - 0.35L - 0.18L^2 + 0.14L^3) = (1 + 0.52L)(1 - 0.87L + 0.27L^2)$$

où le second terme a des racines complexes conjuguées. On peut alors noter que la longueur moyenne du cycle stochastique est alors de 10.83 trimestres, c'est à dire entre 2 ans et demi et 3 ans.