

COURS DE SERIES TEMPORELLES THEORIE ET APPLICATIONS

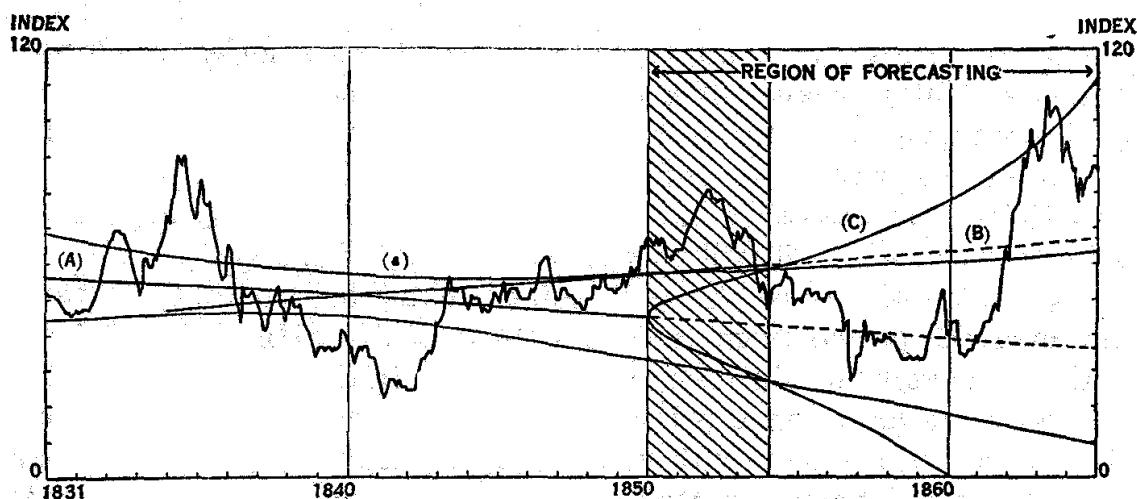


FIGURE 149.—FORECASTING OF RAIL STOCK PRICES.

This chart shows the highly explosive character of trend extrapolations where the basic data are highly variable. The shaded region is the region of forecast corresponding to the standard error of the trend shown in the chart. (a) $y(t) + \sigma_f(t)$; (a') $y(t) - \sigma_f(t)$; (C) $y(t) \pm \sigma(m)$.

VOLUME 2

Modèles linéaires multivariés : VAR et cointégration
 Introduction aux modèles ARCH et GARCH
 Introduction à la notion de mémoire longue
 Exercices corrigés et compléments informatiques

ARTHUR CHARPENTIER
arthur.charpentier@ensae.fr

Contents

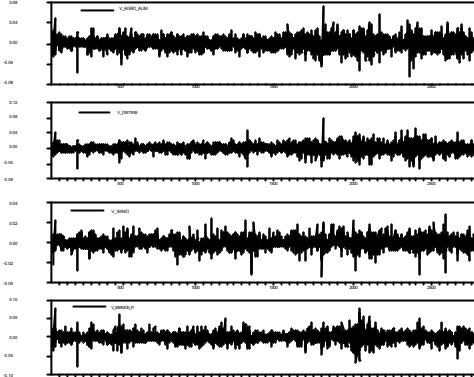
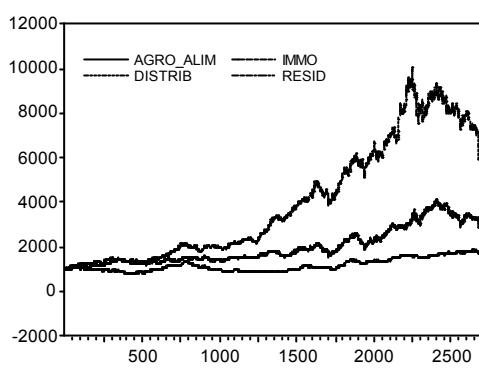
1 Les séries temporelles multivariées	4
1.1 La notion de causalité	5
1.1.1 Dépendance stochastique	5
1.1.2 Causalité au sens de Granger	6
1.2 La notion de cointégration	6
1.2.1 Modèles à correction d'erreur (<i>ECM</i>)	8
1.2.2 Tests de cointégration	9
1.2.3 Généralisation à k variables	9
1.2.4 Différences entre les approches de Engle/Granger et de Johansen	10
1.3 La modélisation <i>VAR</i> (p)	14
1.3.1 La représentation <i>VAR</i> (p)	14
1.3.2 Les modèles <i>ARMAX</i> ou <i>VARMA</i>	16
1.3.3 Estimation des paramètres d'un modèle <i>VAR</i>	17
1.3.4 Autocovariances et autocorrélations	18
1.3.5 Application sur un exemple	18
1.3.6 Prévision à l'aide des modèles <i>VAR</i>	23
1.3.7 Tests de bruit blanc des erreurs	24
1.3.8 Test de causalité (<i>Granger</i> (1969)) sur des modèles <i>VAR</i>	24
1.3.9 Les modèles <i>VAR</i> (p) : analyse des chocs	25
1.4 Application des modèles <i>VAR</i>	28
1.4.1 Application : investissement, revenu et consommation	28
1.4.2 Application des modèles <i>VAR</i> : rendements d'une action et d'un indice	32
1.5 Régression linéaire dans un cadre dynamique	33
1.5.1 Les retards échelonnés	33
1.5.2 Le modèle d'ajustement partiel	34
1.5.3 Les modèles autorégressifs à retards échelonnés (<i>ADL</i> (p, q))	35
1.6 Compléments : modèles multivariés sous <i>SAS</i>	36
1.6.1 Estimation d'un modèle <i>VAR</i>	36
1.6.2 Estimation d'un modèle à correction d'erreur	38
1.7 Conseils bibliographiques	41
2 Les modèles <i>ARCH</i> - Autorégressifs Conditionnellement Hétérosclélastiques	42
2.1 Notions de stationnarité et notions de linéarité	43
2.2 Présentation des modèles <i>ARCH</i>	43
2.2.1 Processus <i>ARCH</i> (1)	44
2.2.2 Processus <i>ARCH</i> (p)	45
2.2.3 Processus <i>GARCH</i> (p, q)	45
2.3 Modèles avec erreurs <i>ARCH</i>	46
2.3.1 Erreurs <i>ARCH</i> (1)	46
2.3.2 Erreurs <i>ARCH</i> (p)	48
2.3.3 Remarque : test de racine unité en présence d'erreurs <i>ARCH</i>	48
2.4 Estimation et prévision	49
2.4.1 Estimation des paramètres d'un modèles <i>ARCH</i>	49
2.4.2 La procédure de Diebold (1987) : test d'autocorrélation en présence d'effet <i>ARCH</i>	49
2.4.3 Prévision et intervalle de confiance	50
2.5 Modèles <i>ARCH</i> et finance	52
2.5.1 Liens entre temps continu et temps discret	53
2.5.2 Modèles avec variance stochastique/déterministe en temps continu	54
2.5.3 Modèles avec variance stochastique/déterministe en temps discret	54
2.6 Autres types de modèles non-linéaires	55
2.6.1 Les modèles bilinéaires - notés <i>BL</i> (p, q, P, Q)	55
2.6.2 Les modèles autorégressifs exponentiels - <i>EXPAR</i>	55
2.6.3 Les modèles autorégressifs à seuils - <i>TAR</i> , <i>STAR</i> ou <i>SETAR</i>	56
2.6.4 Les généralisations des modèles <i>ARCH</i>	57
2.7 Les tests de linéarité	59
2.7.1 Les tests du multiplicateur de Lagrange	59

2.7.2	Le test du <i>CUSUM</i>	60
2.7.3	Le test <i>BDS</i>	61
2.7.4	Le test <i>RESET</i> (<i>Regression error specification test</i>)	62
2.7.5	La procédure <i>AUTOREG</i> sous <i>SAS/ETS</i>	62
2.8	Application sur des données réelles	65
2.8.1	Modélisation <i>GARCH</i> du <i>CAC40</i>	65
2.8.2	Modélisation des rendements du cours <i>Intel</i> : modèle <i>ARCH</i>	67
2.8.3	Modélisation des rendements de l'indice <i>S&P</i> : modèle <i>GARCH</i>	68
2.8.4	Modélisation des rendements de l'action <i>IBM</i> (avec dividendes) : modèle <i>TAR</i>	69
2.8.5	Modélisation du taux de chômage aux Etats Unis : modèles <i>TAR</i> et <i>ARCH</i>	69
2.9	Conseils bibliographiques	72
2.10	Compléments sur les séries temporelles en finance	73
2.10.1	Introduction historique	73
2.10.2	Complément sur les données disponibles	73
3	Introduction à la notion de "mémoire longue"	75
3.1	Processus self-similaires	76
3.1.1	Accroissements stationnaires de processus self-similaires	76
3.2	Processus <i>FARIMA - ARIMA</i> Fractionnaires	76
3.2.1	Processus fractionnaire sans composante <i>AR</i> et <i>MA</i> : processus <i>FARIMA</i> (0, d , 0)	77
3.2.2	Autocorrélations des processus <i>ARFIMA</i>	79
3.3	Estimation du paramètre d	79
3.3.1	Méthode du log-autocorrélogramme	79
3.3.2	Méthode du log-périodogramme	80
3.3.3	Méthode de Whittle	81
3.4	Exemples d'applications des processus à mémoire longue	81
3.4.1	Applications en finance : rendements d'indices boursier	81
3.4.2	Applications en finance : taux de change	84
3.4.3	Applications en hydrologie	84
3.4.4	Applications en économie	85
3.5	Conseils bibliographiques	85
4	Compléments : exercices	86
4.1	Exercices avec correction	86
4.2	Examen de 2001/2002	91
4.3	Examen de 2002/2003	95
4.4	Examen de 2003/2004	100
5	Compléments : simulation de séries temporelles	113
5.1	Simulation de processus en temps discret	113
5.1.1	Simulation d'un bruit blanc gaussien	113
5.1.2	Simulation d'un processus <i>ARMA</i>	114
5.1.3	Simulation d'un processus <i>ARCH</i>	115
5.2	Introduction à la simulation de processus en temps continu	115
6	Compléments : Logiciels de séries temporelles	116
6.1	Introduction à EViews	116
6.1.1	Les données sous EViews	117
6.1.2	Les graphiques sous Eviews	117
6.2	La régression linéaire sous Eviews	119
6.2.1	Estimation des paramètres (1) – (2)	120
6.2.2	Statistique de Student <i>F</i> (3) – (4)	121
6.2.3	Coefficient R^2 (5) – (6)	123
6.2.4	Somme des carrés (7) – (8)	124
6.2.5	Log-vraisemblance du modèle (9)	124
6.2.6	Test du Durbin-Watson (10)	125
6.2.7	'Dependent var' (11) – (12)	126
6.2.8	Critère d'Aikaike (13) et de Schwarz (14)	126

6.2.9 Statistique de Fisher F (15) – (16)	126
6.2.10 Statistique sur les résidus	127
6.2.11 Copier/coller des sorties dans un rapport	128
6.2.12 La régression linéaire sur d'autres logiciels	130
6.3 Lecture de l'autocorrélogramme sous Eviews	132
6.4 Estimation d'un modèle $ARMA$ sous Eviews	133
6.4.1 Racines des polynômes AR et MA (17)	133
6.4.2 Tests sur les erreurs ε_t	134
6.5 Les sorties de la procédure $ARIMA$ sous SAS	136
6.6 Utilisation de SAS avec des séries chronologiques	138
6.7 Méthode de lissage exponentiel : Holt-Winters	139
6.8 Modèles $ARIMA$ sous SAS	139
6.8.1 Syntaxe de la procédure $ARIMA$ sous SAS	139
6.8.2 La lecture des sorties SAS	140

1 Les séries temporelles multivariées

Les graphiques ci-dessous donnent l'évolution des indices sectoriels du *CAC*, pour les secteurs de l'agro-alimentaire, de la distribution, des services financiers, et de l'immobilier. Un portefeuille "diversifié" pourrait correspondre à un portefeuille comportant des titres de chacun de ces secteurs. Pour prévoir les rendements futurs du portefeuille, il convient de modéliser conjointement l'évolution de ces différents indices. Comme le confirment les graphiques ci-dessous, l'évolution de ces indices ne se fait pas indépendamment les uns des autres.



Le graphique de gauche correspond à l'évolution conjointe de ces indices, alors que les graphiques de droites permettent de comparer les rendements de ces indices.

Nous allons introduire dans cette partie un certain nombre de concepts afin de bien comprendre la dynamique de la dépendance entre différentes séries (*notions de dépendance, de causalité, de cointégration...*). Nous verrons également une classe de modèles, généralisant les modèles $AR(p)$ univariés dans un cadre multivarié : les modèles $VAR(p)$.

Exemple 1 Le graphique ci-dessous¹ représente l'évolution du spread des obligations d'Etat pour l'Argentine, le Brésil et le Mexique, entre 1997 et 2000,

¹source : Bazdresch, S & Werner, A.M. *Contagion of International financial crises : the case of Mexico* (2001).

1.1 La notion de causalité

1.1.1 Dépendance stochastique

Dépendance à partir les lois conditionnelles : information de Kullback On peut définir l'information de Kullback par

$$K(y|x) = \mathbb{E} \left(\log \left(\frac{f(y|x)}{f(y)} \right) |x \right).$$

D'après la propriété de Jensen,

$$K(y|x) = \mathbb{E} \left(\log \left(\frac{f(y|x)}{f(y)} \right) |x \right) = -\mathbb{E} \left(\log \left(\frac{f(y)}{f(y|x)} \right) |x \right) \geq -\log \mathbb{E} \left(\left(\frac{f(y)}{f(y|x)} \right) |x \right) = 0,$$

et on peut noter que $K(y|x) = 0$ si et seulement si $f(y|X=x) = f(y)$ presque sûrement.

Remarque 1 Dans le cas gaussien, $\text{cov}(X, Y) = 0$ si et seulement si X et Y sont indépendants. Supposons que

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{bmatrix} \right).$$

Alors les lois marginales et conditionnelles vérifient $Y \sim \mathcal{N}(0, \Sigma_Y)$, et

$$Y|X \sim \mathcal{N}(\mathbb{E}(Y|X), V(Y|X)) \text{ où } \begin{cases} \mathbb{E}(Y|X) = \Sigma_{YX}\Sigma_X^{-1}X \\ V(Y|X) = \Sigma_Y - \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}. \end{cases}$$

Aussi la loi du rapport figurant dans l'information de Kullback est donné par

$$\frac{f(y|x)}{f(y)} = \frac{\sqrt{\det(V(Y))}}{\sqrt{\det(V(Y|X))}} \exp \left(-\frac{1}{2} [Y - \mathbb{E}(Y|X)]' V(Y|X) [Y - \mathbb{E}(Y|X)] \right),$$

d'où finalement (après calculs)

$$\log \left(\frac{f(y)}{f(y|x)} \right) = \frac{1}{2} \log \left[\det \left(V(Y) . V(Y|X)^{-1} \right) \right].$$

La notion de causalité La loi du processus est la loi du processus du couple (X_t, Y_t) . Cette dernière s'écrit, à la date t , conditionnellement au passé (*noté \underline{x}_{t-1} et \underline{y}_{t-1}*), $\ell(x_t, y_t | \underline{x}_{t-1}, \underline{y}_{t-1})$. Dans le cas où les processus (X_t) et (Y_t) sont indépendants, alors

$$\ell(x_t, y_t | \underline{x}_{t-1}, \underline{y}_{t-1}) = \ell(x_t | \underline{x}_{t-1}) \times \ell(y_t | \underline{y}_{t-1}).$$

On peut montrer que

$$\frac{\ell(x_t, y_t | \underline{x}_{t-1}, \underline{y}_{t-1})}{\ell(x_t | \underline{x}_{t-1}) \ell(y_t | \underline{y}_{t-1})} = \frac{\ell(x_t | \underline{x}_{t-1}, \underline{y}_{t-1})}{\ell(x_t | \underline{x}_{t-1})} \times \frac{\ell(y_t | \underline{x}_{t-1}, \underline{y}_{t-1})}{\ell(y_t | \underline{y}_{t-1})} \times \frac{\ell(x_t, y_t | \underline{x}_{t-1}, \underline{y}_{t-1})}{\ell(x_t | \underline{x}_{t-1}, \underline{y}_{t-1}) \ell(y_t | \underline{x}_{t-1}, \underline{y}_{t-1})}.$$

En passant au log et en prenant l'espérance (*et en multipliant par 2*) alors

$$\underbrace{2\mathbb{E} \left(\log \frac{\ell(x_t, y_t | \underline{x}_{t-1}, \underline{y}_{t-1})}{\ell(x_t | \underline{x}_{t-1}) \ell(y_t | \underline{y}_{t-1})} \right)}_{\text{Mesure de dépendance } C_{X,Y}} = \underbrace{2\mathbb{E} \left(\log \frac{\ell(x_t | \underline{x}_{t-1}, \underline{y}_{t-1})}{\ell(x_t | \underline{x}_{t-1})} \right)}_{\text{Causalité unidirectionnelle } C_{Y \rightarrow X}} + \underbrace{2\mathbb{E} \left(\log \frac{\ell(y_t | \underline{x}_{t-1}, \underline{y}_{t-1})}{\ell(y_t | \underline{y}_{t-1})} \right)}_{\text{Causalité unidirectionnelle } C_{X \rightarrow Y}} + \underbrace{2\mathbb{E} \left(\log \frac{\ell(x_t, y_t | \underline{x}_{t-1}, \underline{y}_{t-1})}{\ell(x_t | \underline{x}_{t-1}, \underline{y}_{t-1}) \ell(y_t | \underline{x}_{t-1}, \underline{y}_{t-1})} \right)}_{\text{Causalité instantanée } C_{X \leftrightarrow Y}}.$$

1.1.2 Causalité au sens de Granger

Soient (X_t) et (Y_t) deux séries temporelles, et notons le passé de (X_t) et (Y_t) , $\underline{X}_t = \{X_t, X_{t-1}, \dots\}$ et $\underline{Y}_t = \{Y_t, Y_{t-1}, \dots\}$. Granger a introduit en 1969 différentes notions de causalité :

Définition 1 (i) Y cause X à la date t si et seulement si

$$\mathbb{E}(X_t | \underline{X}_{t-1}, \underline{Y}_{t-1}) \neq \mathbb{E}(X_t | \underline{X}_{t-1}).$$

(ii) Y cause X instantanément à la date t si et seulement si

$$\mathbb{E}(X_t | \underline{X}_{t-1}, Y_t) \neq \mathbb{E}(X_t | \underline{X}_{t-1}, Y_{t-1}).$$

Il y a équivalence entre

- (1) X ne cause pas Y instantanément à la date t
- (2) Y ne cause pas X instantanément à la date t

Exemple 2 Soient (X_t) et (Y_t) définis par

$$\begin{cases} X_t = \varepsilon_t + \alpha Z_t + \beta Z_{t-1} \\ Y_t = Z_t, \end{cases}$$

où (ε_t) et (Z_t) sont des bruits blancs indépendants.

- (i) $\mathbb{E}(X_t | \underline{X}_{t-1}, \underline{Y}_{t-1}) = \beta Z_{t-1} = \beta Y_{t-1}$: Y cause X à la date t si et seulement si $\beta \neq 0$,
- (ii) $\mathbb{E}(X_t | \underline{X}_{t-1}, Y_t) = \alpha Z_t + \beta Z_{t-1} = \alpha Y_t + \beta Y_{t-1}$: il y a causalité instantanée de Y vers X si et seulement si $\alpha \neq 0$.

1.2 La notion de cointégration

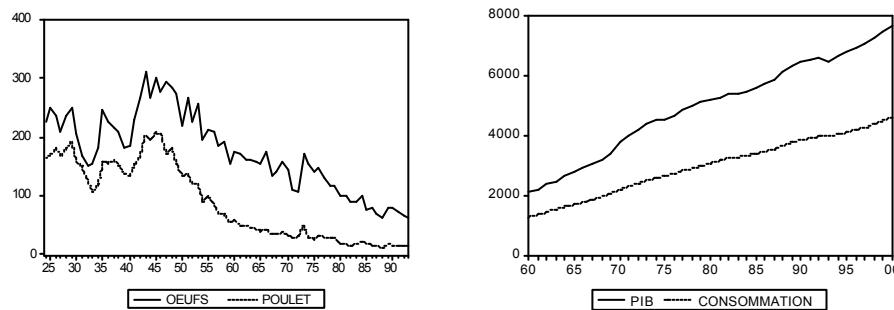
L'analyse de la cointégration permet d'identifier la relation entre plusieurs variables. Cette notion a été introduite dès 1974 par Engle et Newbold, sous le nom de "spurious regressions", ou régressions fallacieuses, puis formalisée par Engle et Granger en 1987, et enfin par Johansen en 1991 et 1995.

Une série est intégrée d'ordre d s'il convient de la différencier d fois avant de la stationnariser.

Définition 2 La série (X_t) sera dite **intégrée d'ordre d** ($d \geq 1$) si $\Delta^{d-1} X_t$ n'est pas stationnaire et $\Delta^d X_t$ est stationnaire. Une série stationnaire sera dite intégrée d'ordre 0.

Remarque 2 Soient (X_t) une série stationnaire et (Y_t) intégrée d'ordre 1, alors $(X_t + Y_t)$ est intégrée d'ordre 1. Toutefois, si (X_t) et (Y_t) sont intégrées d'ordre d , alors $(X_t + Y_t)$ peut être soit intégrée d'ordre d , soit stationnaire (dans le cas où les tendances s'annulent).

Le graphique ci-dessous à gauche présente deux séries non-cointégrées, alors que le graphique de droite correspond à des séries cointégrées



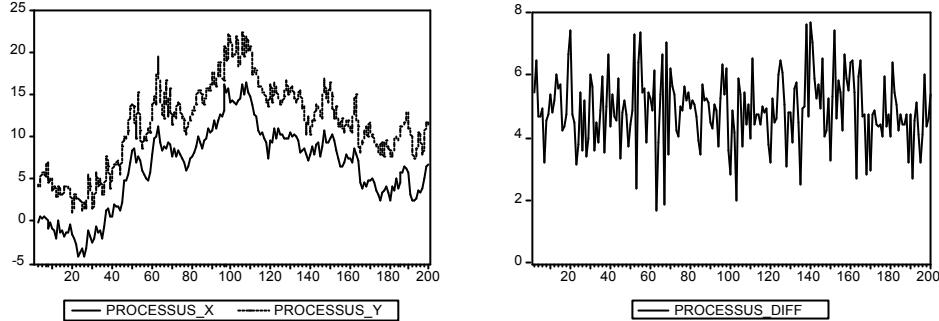
avec à droite, les *PIB* et la consommation agrégée en France, et à gauche, le prix de la douzaine d'oeufs et d'un poulet, aux Etats Unis. Comme nous le verrons plus tard, même si les courbes de gauche semblent avoir une tendance commune, elles ne sont pas pour autant coïntégrées.

Définition 3 Deux séries (X_t) et (Y_t) sont **cointégrées** si

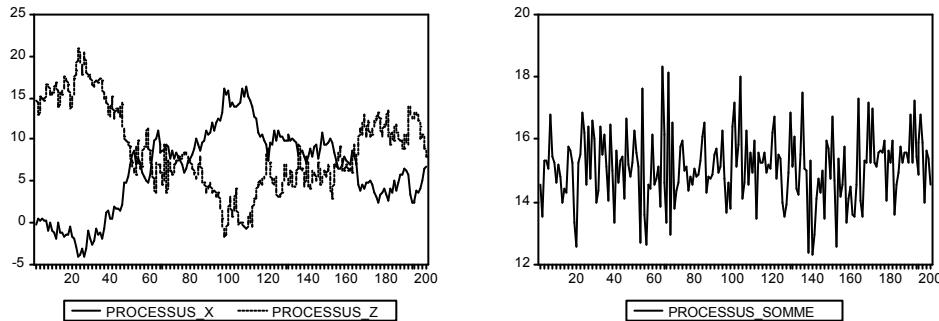
- (i) (X_t) et (Y_t) sont intégrées d'ordre d
- (ii) il existe une combinaison linéaire de ces séries qui soit intégrée d'ordre strictement inférieur à d , noté $d - b$

Dans le cas de l'intégration, on notera $X_t \sim I(d)$, et pour la cointégration $X_t, Y_t \sim CI(d, b)$. Le vecteur (α, β) tel que $\alpha X_t + \beta Y_t \sim I(d - b)$ sera appelé vecteur de cointégration.

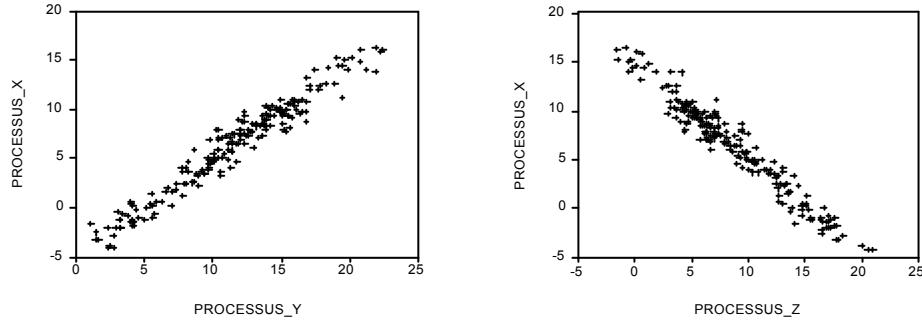
Exemple 3 Le cas le plus simple est le suivant : deux séries (X_t) et (Y_t) intégrées d'ordre 1 seront cointégrées s'il existe une combinaison linéaire $\alpha X_t + \beta Y_t$ telle que le processus (Z_t) défini par $Z_t = \alpha X_t + \beta Y_t$ soit stationnaire. Aussi, si X_t est intégré d'ordre 1 (par exemple une marche aléatoire), et si (ε_t) un bruit blanc (X_t) et $(X_t + \varepsilon_t)$ sont cointégrés. Ce cas est représenté ci-dessous, avec à gauche (X_t) et (Y_t) où $Y_t = X_t + \varepsilon_t + \kappa$, et à droite la différence $(\kappa + \varepsilon_t)$.



Les graphiques ci-dessous représentent les séries (X_t) et (Z_t) où $Z_t = -X_t + \varepsilon_t + \kappa$, et à droite la somme $(\kappa - \varepsilon_t)$.

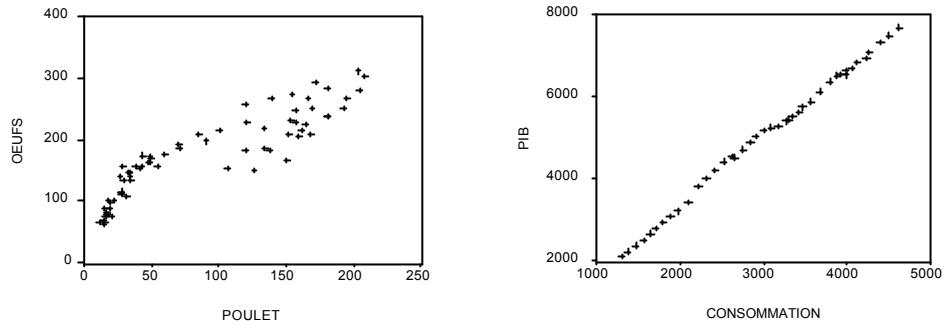


Les deux combinaisons linéaires étant stationnaires, les séries sont cointégrées. La notion de cointégration peut également se visualiser en représentant le nuage de points (scatterplot) des observations (X_t, Y_t) - à gauche - et (X_t, Z_t) - à droite



Remarque 3 Le dernier point de l'exemple précédent permet d'avoir d'avoir une intuition visuelle sur la présence ou absence de relation de cointégration (linéaire) entre les variables présentées initialement, avec à gauche les observations

des prix du poulet et des oeufs, et à droite, le PIB et la consommation



1.2.1 Modèles à correction d'erreur (ECM)

Les modèles dits à correction d'erreur ont été introduits au début des années 80, par Hendry en particulier. Ces modèles dynamiques permettent d'intégrer les évolutions à long terme et à courte terme des variables.

Considérons deux variables (X_t) et (Y_t) cointégrées d'ordre 1 ($X_t, Y_t \sim CI(1, 1)$) et soit $[\beta, -1]$ le vecteur de cointégration. L'idée des modèles à correction d'erreur est de considérer des relations de la forme

$$\Delta Y_t = \lambda \Delta X_t + \mu [Y_{t-1} - \beta X_{t-1}] + \varepsilon_t, \quad (1)$$

ce qui revient à décomposer un processus stationnaire (ΔY_t) en une somme de deux processus stationnaires (ΔX_t et $Y_{t-1} - \beta X_{t-1}$). De façon plus générale que (1), ces modèles s'écrivent

$$\Delta Y_t = \mu + \sum_{i=1}^p \alpha_i \Delta Y_{t-i} + \sum_{j=0}^q b_i \Delta X_{t-i} + c [Y_{t-1} - \beta X_{t-1}] + \eta_t, \quad (2)$$

où les variables interviennent soit à travers leurs différences premières (*supposées stationnaires*), soit à travers un terme d'écart à la cible à long terme, à la période précédente (*qui doit être stationnaire si la théorie économique sous-jacente est pertinente*).

Propriété 1 Théorème de représentation de Granger - Toutes les séries cointégrées peuvent être représentées par un modèle à correction d'erreur

Preuve. Engle,R.E. & Granger,C.W.J. (1987). 'Cointegration and error-correction : representation, estimation and testing'. *Econometrica*. 55 ■

Remarque 4 Le théorème de représentation de Granger présente l'intérêt de faire la synthèse entre les deux approches:

- l'approche ECM issue de l'idée de concilier des préoccupations de théorie économique avec une écriture rigoureuse des équations économétriques

- l'approche VAR (que nous développeront dans la partie suivante) issue d'une approche purement statistique, de type "boîte noire".

Exemple 4 Considérons les variables économiques (C_t) et (R_t), correspondant respectivement à la consommation et au revenu. D'un point de vue théorique, il est souvent admis qu'il existe une relation à long terme de la forme $C_t = \kappa R_t$, ce qui donne, sous forme logarithmique, $c_t = k + r_t$. D'un point de vue économique, l'expression "long terme" ne correspond pas uniquement à une valeur limite, mais plus généralement, à un comportement asymptotique : on suppose que l'on a une chemin de croissance équilibré à taux constant, i.e. $c_t - c_{t-1} = cste$. On suppose qu'il existe une relation dynamique liant consommation et revenu de la forme suivante,

$$A(L)c_t = \mu + B(L) + \varepsilon_t \quad (3)$$

où A et B sont ici supposés de degré p et q respectivement, et où (ε_t) est un bruit blanc. Or, d'après la formule de Taylor, il est possible d'écrire

$$A(L) = A(1) + (1-L) A^*(L) \text{ où } A^*(L) = \sum_{k=1}^p \frac{(-1)^k}{k!} A^{(k)}(1) (1-L)^{k-1},$$

et de même $B(L) = B(1) + (1 - L)B^*(L)$. Par substitution dans (3), on peut alors écrire

$$A(1)c_t = \mu - A^*(L)\Delta c_t + B(1)r_t + B^*(L)\Delta r_t + \varepsilon_t.$$

On dit que cette expression combine les termes en niveau et les termes en différence première. D'après la relation de long terme ($\Delta c_t = \mu_C$, $\Delta r_t = \mu_R$ et $c_t = k + r_t$) on en déduit

$$A(1)c_t = \mu - A^*(L)\mu_C + B(1)r_t + B^*(L)\mu_R, \quad (4)$$

en supposant le terme d'erreur nul à long terme. (4) est dite **relation de long terme** associée à (3). Il faut toutefois rajouter l'équation suivante, correspondant à l'hypothèse d'élasticité unitaire de la consommation au revenu, $A(1) = B(1)$. Dans le cas par exemple, où $p = q = 1$, alors $A(L) = 1 + a_1L$ et $B(L) = b_0 + b_1L$, alors (3) s'écrit

$$c_t + a_1c_{t-1} = \mu + b_0r_t + b_1r_{t-1} + \varepsilon_t,$$

d'où finallement

$$\Delta c_t = \mu - (1 + a_1)c_{t-1} + b_0\Delta r_t + (b_0 + b_1)r_{t-1} + \varepsilon_t.$$

Et en utilisant l'égalité $A(1) = B(1) = 1$, cette équation s'écrit

$$\Delta c_t = \mu - (1 + a_1)[c_{t-1} - r_{t-1}] + b_0\Delta r_t + \varepsilon_t.$$

qui correspond à l'écriture proposée dans l'équation (2). Sous la forme

$$\Delta c_t = [\mu + (1 + a_1)k] - (1 + a_1)[c_{t-1} - r_{t-1} - k] + b_0\Delta r_t + \varepsilon_t.$$

il s'agit de la forme générale d'un modèle à correction d'erreur, décrivant l'ajustement instantané de la consommation Δc_t aux variations de revenu et à l'écart à la cible à long terme, $(c_{t-1} - r_{t-1} - k)$.

De façon pratique, supposons que l'on dispose de 2 (ou plus) séries (X_t) et (Y_t) intégrées à l'ordre 1. Notons $Z_t = (X_t, Y_t)'$. D'après le théorème de Wold, on peut représenter (Z_t) sous la forme $(1 - L)Z_t = \Theta(L)\varepsilon_t$ où (ε_t) est un bruit blanc (en dimension 2), $\Theta(0) = \mathbb{I}$ et aucune ligne de $\Theta(1)$ n'est nulle (*sinon une des composantes est stationnaire*).

(i) si les composantes (X_t) et (Y_t) ne sont pas cointégrées, alors le vecteur (Z_t) peut s'écrire sous forme *VAR*, ou *VARMA*, en différences premières, $\Phi(L)\Delta Z_t = \varepsilon_t$ ou $\Phi(L)\Delta Z_t = (L)\varepsilon_t$.

(ii) si les composantes (X_t) et (Y_t) sont cointégrées, alors le vecteur (Z_t) peut s'écrire sous forme *VARMA*, $A(L)\Delta Z_t = B(L)\varepsilon_t$, avec une représentation à correction d'erreur $A^*(L)\Delta Z_t = -\gamma Z_{t-1} + B(L)\varepsilon_t$. La relation *VARMA* est dite relation de court terme, et la seconde relation de long terme, dans laquelle apparaît une force de rappel vers la cible de long terme.

1.2.2 Tests de cointégration

Afin de tester la cointégration entre deux variables, on peut utiliser l'algorithme mis en place par Engle et Granger.

(1) tester l'ordre d'intégration des variables : une des conditions nécessaire pour qu'il y ait cointégration étant que les deux séries doivent être intégrées de même ordre. Ces tests reposent sur l'utilisation des tests de Dikey & Fuller. On cherche alors d tel que $X_t \sim I(d)$ et $Y_t \sim I(d)$.

(2) estimation de la relation de long-terme.

Compte tenu du théorème de représentation de Granger, il est également possible de tester la cointégration en estimant le modèle à correction d'erreur associé. Cette estimation peut se faire en deux étapes, en estimant par *mco* la relation de long-terme

$$Y_t = \hat{\alpha} + \hat{\beta}X_t + \hat{\varepsilon}_t,$$

puis en estimant, toujours par *mco* la relation de court-terme

$$\Delta Y_t = \lambda \Delta X_t + \mu \hat{\varepsilon}_{t-1} + \eta_t.$$

Le coefficient μ doit alors être significativement négatif : dans le cas contraire, on rejette l'hypothèse d'une modélisation de la forme *ECM*.

1.2.3 Généralisation à k variables

La notion de cointégration peut se généraliser de 2 à k variables.

1.2.4 Différences entre les approches de Engle/Granger et de Johansen

Comme nous venons de le voir la méthode de Engle et Granger (1987) s'articule en deux étapes

(1) régression statistique entre variables intégrées (*après avoir vérifié que les variables sont intégrées de même ordre*)

(2) test de vérification de la stationnarité des résidus (*test de Dickey & Fuller*)

La méthode Johansen consiste à tester les restrictions imposées par la cointégration sur le modèle *VAR* non restreint. On considère un modèle *VAR*(p) de la forme

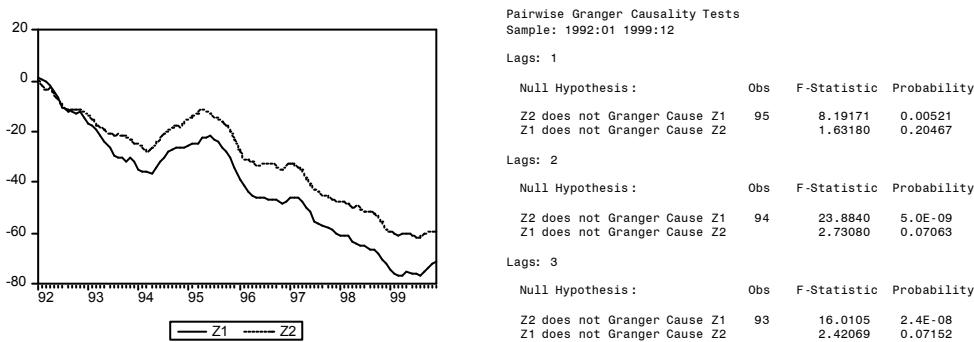
$$Y_t = A_1 Y_{t-1} + \dots + A_p Y_{t-p} + \varepsilon_t \text{ ou } \Delta Y_t = \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \Gamma_2 \Delta Y_{t-2} + \dots + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t, \quad (5)$$

où $\Pi = (A_1 + \dots + A_p) - I$ et $\Gamma_i = -(A_{i+1} + \dots + A_p)$. Le théorème de représentation de Granger affirme alors que le coefficient de la matrice Π a un rang réduit $r < k$, nombre de variables cointégrées, et donc, il existe des matrices U et V de rang r , telles que $\Pi = UV'$ et telles que $V'Y_t$ soit stationnaire.

r est alors le nombre de relations de cointégration (*rang de cointégration*), chaque colonne de V est le vecteur de cointégration, et les éléments de U sont des paramètres d'ajustement.

La méthode de Johansen consiste à estimer la matrice Π et de voir si on peut rejeter des restrictions impliquées par le rang réduit de Π : si on a k variables endogènes, toutes avec une racine unitaire (*non stationnaire, mais stationnaire après différenciation - une fois*), on peut avoir $k-1$ relations de cointégration linéaires indépendantes. S'il n'y a aucune relation de cointégration, des analyses standards de séries chronologiques telles que les *VAR* peuvent être appliquées aux différences premières des données.

Exemple 5 Sur les données détaillées dans l'exemple (1.3.5), intégrées², le test de causalité de Granger proposé par EViews donne les résultats suivants, à droite



avec à gauche les courbes des processus Z_1 et Z_2 . Trois cas sont présentés ici : lorsque le modèle contient 1, 2 et 3 retards. L'hypothèse nulle est ici qu'il n'y a pas causalité, à savoir que Z_2 ne cause pas Z_1 dans la première équation, et que X ne cause pas Y dans la seconde. Ici, on rejette l'hypothèse nulle selon laquelle Z_2 ne cause pas Z_1 car la puissance associée aux tests est très faible (dans les trois cas), et au seuil 5, on accepte l'hypothèse que Z_1 cause Z_2 au sens de Granger. Au contraire, on rejette l'hypothèse de causalité de Z_2 vers Z_1 .

²Ce test (test de Johansen) est un test de coïntégration : il n'est donc valable que sur des variables non-stationnaires.

Le test de Johansen est présenté ci-dessous

Included observations: 91 Test assumption: No deterministic trend in the data Series: Z1 Z2 Lags interval: 1 to 4						Included observations: 91 Test assumption: No deterministic trend in the data Series: Z1 Z2 Lags interval: 1 to 4					
Eigenvalue	Likelihood Ratio	5 Percent Critical Value	1 Percent Critical Value	Hypothesized No. of CE(s)	Eigenvalue	Likelihood Ratio	5 Percent Critical Value	1 Percent Critical Value	Hypothesized No. of CE(s)		
0.005347	3.968995	12.53	16.31	None	0.104344	10.82946	15.41	20.04	None		
0.007598	0.694084	3.84	6.51	At most 1	0.006586	0.601353	3.76	6.65	At most 1		
(*)* denotes rejection of the hypothesis at 5% significance level L.R. rejects any cointegration at 5% significance level											
Unnormalized Cointegrating Coefficients:											
Z1	Z2				Z1	Z2					
0.019804	-0.023022				0.018941	-0.016819					
-0.028400	0.040442				-0.037981	0.049285					
Normalized Cointegrating Coefficients: 1 Cointegrating Equation(s)											
Z1	Z2	C			Z1	Z2	C				
1.000000	-1.162462	(0.17327)			1.000000	-0.887939	13.86919				
Log likelihood	-390.7528				Log likelihood	-385.9191					
Included observations: 93 Test assumption: No deterministic trend in the data Series: Z1 Z2 Lags interval: 1 to 2											
Eigenvalue	Likelihood Ratio	5 Percent Critical Value	1 Percent Critical Value	Hypothesized No. of CE(s)	Rank	No Intercept	Intercept	Linear	Quadratic	Linear	Quadratic
0.032396	3.741320	12.53	16.31	None	0	-409.0860	-409.0860	-407.4978	-407.4978	-402.6242	-402.6242
0.007270	0.678590	3.84	6.51	At most 1	1	-407.9648	-402.8793	-402.0757	-402.0757	-397.9195	-397.9195
(*)* denotes rejection of the hypothesis at 5% significance level L.R. rejects any cointegration at 5% significance level											
Unnormalized Cointegrating Coefficients:											
Z1	Z2				0	8.797548	8.797548	8.763383	8.763383	8.658885	8.658885
0.023160	-0.027864				1	8.773436	8.664070	8.646789	8.646789	8.557409	8.557409
-0.022981	0.032209				2	8.772081	8.639970	8.639970	8.639970	8.544793	8.544793
Normalized Cointegrating Coefficients: 1 Cointegrating Equation(s)											
Z1	Z2	C			0	9.187449	9.187449	9.250770	9.250770	9.243436	9.243436
1.000000	-1.201111	(0.111159)			1	9.358288	9.297695	9.329116	9.329116	9.376997	9.376997
Log likelihood	-437.2565				2	9.551584	9.517247	9.517247	9.517247	9.519545	9.519545
Log Likelihood by Model and Rank											
Akaike Information Criteria by Model and Rank											
Schwarz's Criteria by Model and Rank											
Normalized Cointegrating Coefficients: 1 Cointegrating Equation(s)											
Z1	Z2	C			0	9.187449	9.187449	9.250770	9.250770	9.243436	9.243436
1.000000	-1.201111	(0.111159)			1	9.358288	9.297695	9.329116	9.329116	9.376997	9.376997
Log likelihood	-437.2565				2	9.551584	9.517247	9.517247	9.517247	9.519545	9.519545
L.R. Test: Rank = 0											

EViews propose plusieurs options pour les tests de cointégration³. La sortie en bas à droite est la sortie récapitulative. Pour les autres, les valeurs propres sont présentées dans la première colonne. Le LR - Likelihood Radio - est défini par

$$Q(r) = -T \sum_{i=r+1}^k \log(1 - \lambda_i), \text{ pour } r = 0, 1, \dots, k-1,$$

où λ_i est la i ème valeur propre (lorsqu'elles sont rangées par ordre croissant). Cette statistique est aussi appelée "statistique de la trace". La première ligne de la cinquième colonne (*Hypothesized No. of CE(s)*) teste l'hypothèse de non-cointégration ($r=0$: relation de cointégration). La deuxième ligne teste l'hypothèse d'une relation de coïntégration ($r=1$), la deuxième ligne teste l'hypothèse de deux relations de coïntégration ($r=2$)...etc. Toutes ces hypothèses sont testées contre l'hypothèse alternative que la matrice est de plein rang (c'est à dire que toutes les séries sont stationnaires). Dans l'exemple ci-dessus, la statistique de la trace ne rejette aucune des hypothèses au seuil de 5% : il n'y a donc aucune relation de cointégration car on n'a pas rejeté 0 (*None*) relation de cointégration pour au moins une.

Le vecteur de cointégration n'est pas identifié, à moins d'imposer une normalisation arbitraire : EViews adopte une normalisation de sorte que les r premières séries dans le vecteur Y_t soient normalisés à une matrice identité. Dans le cas présenté en haut à droite, par exemple, la relation de cointégration normalisée qui fait l'hypothèse (fictive) d'une relation de cointégration est donnée par

Normalized Cointegrating Coefficients: 1 Cointegrating Equation(s)					
Z1	Z2	C			
1.000000	-0.887939	13.86919			
(0.25230)					

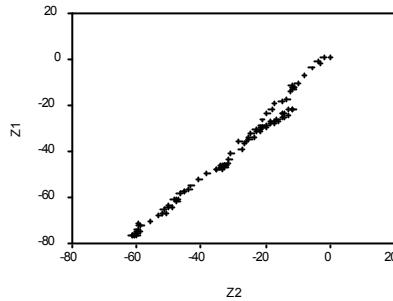
Log likelihood -385.9191

ce qui donne la relation de cointégration de la forme $Z_1 - 0.8873.Z_2 + 13.869$. Enfin, le graphique ci-dessous montre le scatterplot de (Z_1, Z_2) , qui montre (graphiquement) qu'il n'y a pas de relation de cointégration a priori, compte tenu

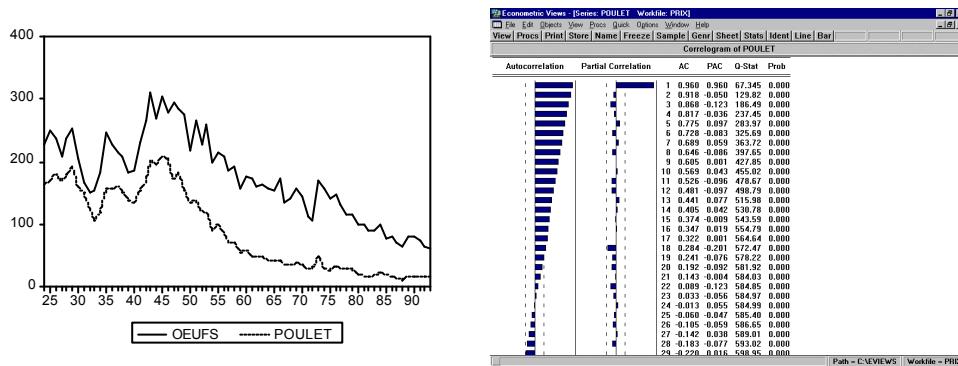
³Les 5 cas étudiés sont les suivants :

- (i) pas de trend, pas de constante dans la relation de coïntégration $\Pi Y_{t-1} = \alpha \beta' Y_{t-1}$
- (ii) pas de trend dans la relation de coïntégration $\Pi Y_{t-1} = \alpha (\beta' Y_{t-1} + \mu)$
- (iii) trend linéaire dans la relation de coïntégration $\Pi Y_{t-1} = \alpha (\beta' Y_{t-1} + \mu) + \alpha^\perp \gamma$
- (iv) trend linéaire dans la relation de coïntégration $\Pi Y_{t-1} = \alpha (\beta' Y_{t-1} + \mu + \lambda t) + \alpha^\perp \gamma$
- (v) trend quadratique dans la relation de coïntégration $\Pi Y_{t-1} = \alpha (\beta' Y_{t-1} + \mu + \lambda t) + \alpha^\perp (\gamma + \delta t)$

de cette rupture pour les valeurs positives, avec deux tendances distinctes.



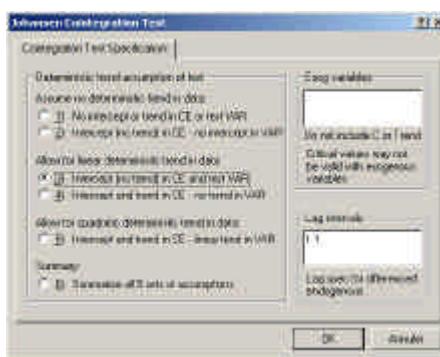
Pour reprendre l'exemple introductif, les prix de la douzaine d'oeufs et d'un poulet, bien que non-stationnaires, ne sont pas pour autant coïntégrées,



comme cela peut se voir sur les sorties des tests de Johansen

```
Included observations: 67
Test assumption: Linear deterministic trend in the data
Series: OEUFFS POULET
Lags interval: 1 to 2
Eigenvalue      Likelihood Ratio      5 Percent Critical Value      1 Percent Critical Value      Hypothesized No. of CE(s)
0.072879      9.161157      15.41      20.04      None
0.059236      4.091221      3.76      6.65      At most 1 *
(*++) denotes rejection of the hypothesis at 5%(1) significance level
L.R. rejects any cointegration at 5% significance level
Unnormalized Cointegrating Coefficients:
OEUFFS          POULET
-0.004465      0.003299
-0.001669      0.003287
Normalized Cointegrating Coefficients: 1 Cointegrating Equation(s)
OEUFFS          POULET          C
1.000000      -0.738944      -109.7839
(0.20045)
Log likelihood -634.8795
```

La lecture des sorties des tests de cointégration sous Eviews 4.1 se fait de la façon suivante,



On considère un modèle *VAR* pour (Y_t) , à savoir $Y_t = A_1 Y_{t-1} + \dots + A_p Y_{t-p} + \varepsilon_t$, où (Y_t) est un vecteur de processus $I(1)$. On réécrit alors ce modèle sous la forme décrite ci-dessus, à savoir

$$\Delta Y_t = \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \Gamma_2 \Delta Y_{t-2} + \dots + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t,$$

où $\Pi = (A_1 + \dots + A_p) - I$ et $\Gamma_i = -(A_{i+1} + \dots + A_p)$. D'après le théorème de représentation de Granger, il existe α et β matrices $k \times r$, où $r < k$, telles que $\Pi = \alpha\beta'$ et telles que $\beta'Y_t$ soit stationnaire. r est alors le nombre de relations de cointégration, et les colonnes de β correspondent aux vecteurs de cointégration. De la même façon que Johansen (1995) 5 cas sont distingués, suivant la tendance retenue,

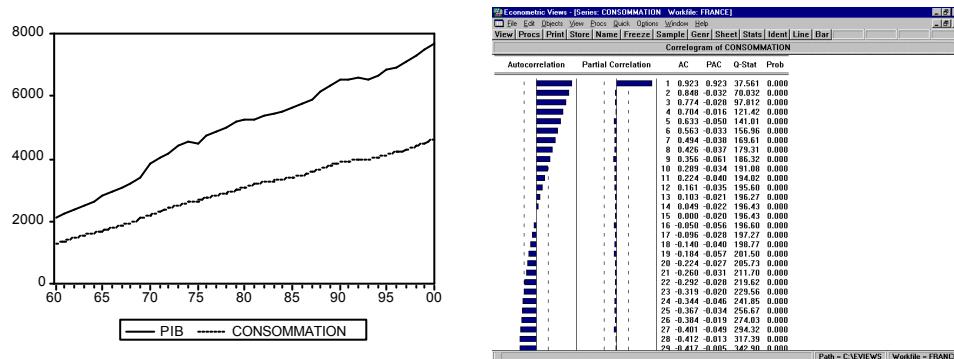
- | | |
|----|---|
| 1. | Pas de tendance déterministe pour (Y_t) , équations de cointégration sans constantes, |
| 2. | Pas de tendance déterministe pour (Y_t) , équations de cointégration avec constantes, |
| 3. | Tendance déterministe pour (Y_t) , équations de cointégration avec constantes, |
| 4. | Tendance déterministe pour (Y_t) , équations de cointégration avec tendances linéaires, |
| 5. | Tendance quadratique pour (Y_t) , équations de cointégration avec tendances linéaires. |

Les modèles testés s'écrivent alors

- | | |
|----|--|
| 1. | $\Pi Y_{t-1} = \alpha \beta' Y_{t-1}$ |
| 2. | $\Pi Y_{t-1} = \alpha (\beta' Y_{t-1} + \mu)$ |
| 3. | $\Pi Y_{t-1} = \alpha (\beta' Y_{t-1} + \mu) + \lambda$ |
| 4. | $\Pi Y_{t-1} = \alpha (\beta' Y_{t-1} + \mu + \alpha t) + \lambda$ |
| 5. | $\Pi Y_{t-1} = \alpha (\beta' Y_{t-1} + \mu + \alpha t) + \lambda + \beta t$ |

Il est d'ailleurs possible sous Eviews de rajouter des composantes exogènes. On considère alors des modèles de la forme $Y_t = A_1 Y_{t-1} + \dots + A_p Y_{t-p} + BX_t + \varepsilon_t$.

Exemple 6 Pour reprendre l'exemple introductif, les PIB français et l'indice de la consommation sont des séries non-stationnaires, et cointégrées,



comme cela peut se voir sur les sorties des tests de Johansen

```
Included observations: 39
Test assumption: No deterministic trend in the data
Series: PIB CONSUMMATION
Lags interval: 1 to 1
Eigenvalue      Likelihood      5 Percent      1 Percent      Hypothesized
          Ratio          Critical Value    Critical Value  No. of CE(s)
0.266578       16.71568        12.53          16.31           None **
0.111813       4.624333        3.84          6.51           At most 1 *
* (**) denotes rejection of the hypothesis at 5%(1%) significance level
L.R. test indicates 2 cointegrating equation(s) at 5% significance level

Unnormalized Cointegrating Coefficients:
          PIB      CONSUMMATION
0.000644     -0.000961
-0.001586      0.002665

Normalized Cointegrating Coefficients: 1 Cointegrating Equation(s)
          PIB      CONSUMMATION
1.000000     -1.492961
                           (0.12258)

Log likelihood -434.1513
```

Il est toutefois possible, sur la version 4.1, d'obtenir des sorties plus complètes sur ce test,

Included observations: 39 after adjusting endpoints Trend assumption: Quadratic deterministic trend Series: CONSUMMATION PIB Lags interval (in first differences): 1 to 1	Unrestricted Cointegrating Coefficients (normalized by b'S11*b=1):			
Unrestricted Cointegration Rank Test				
Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	5 Percent Critical Value	1 Percent Critical Value
None	0.196115	13.52735	18.17	23.46
At most 1 *	0.120636	5.013688	3.74	6.40
*(**) denotes rejection of the hypothesis at the 5%(1%) level				
Trace test indicates no cointegration at both 5% and 1% levels				
Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	5 Percent Critical Value	1 Percent Critical Value
None	0.196115	8.513665	16.87	21.47
At most 1 *	0.120636	5.013688	3.74	6.40
*(**) denotes rejection of the hypothesis at the 5%(1%) level				
Max-eigenvalue test indicates no cointegration at both 5% and 1% levels				
Unrestricted Adjustment Coefficients (alpha):				
D(CONSUMMATION)	8.415089	-6.857180		
D(PIB)	-0.123872	-26.79345		
1 Cointegrating Equation(s):	Log likelihood	-390.8848		
Normalized cointegrating coefficients (std.err. in parentheses)				
CONSUMMATION	PIB			
1.00000	-0.348888 (0.03825)			
Adjustment coefficients (std.err. in parentheses)				
D(CONSUMMATION)	-0.401329 (0.21329)			
D(PIB)	0.005908 (0.63095)			

La lecture de la sortie se fait de la façon suivante : il s'agit ici d'un test de type ADF où les deux séries sont supposées avoir une tendance et une constante

Sample: 1960 2000					
Included observations: 39					
Series: CONSUMMATION PIB					
Lags interval: 1 to 1					
Data Trend:	None	None	Linear	Linear	Quadratic
Rank or No. of CEs	No Intercept No Trend	Intercept No Trend	Intercept No Trend	Intercept Trend	Intercept Trend
Selected (5% level) Number of Co integrating Relations by Model (columns)					
Trace	2	1	0	0	0
Max-Eig	2	1	0	0	0
Log Likelihood by Rank (rows) and Model (columns)					
0	-407.9952	-407.9952	-395.3806	-395.3806	-395.1417
1	-401.9496	-394.8963	-392.3676	-391.0583	-390.8848
2	-399.6374	-392.0803	-392.0803	-388.3780	-388.3780
Akaike Information Criteria by Rank (rows) and Model (columns)					
0	21.12796	21.12796	20.58362*	20.58362*	20.67393
1	21.02305	20.71263	20.63423	20.61837	20.66076
2	21.10961	20.82463	20.82463	20.73733	20.73733
Schwarz Criteria by Rank (rows) and Model (columns)					
0	21.29858	21.29858	20.83955*	20.83955*	21.01518
1	21.36430	21.09653	21.06079	21.08758	21.17263
2	21.62148	21.42181	21.42181	21.41982	21.41982

1.3 La modélisation VAR(p)

1.3.1 La représentation VAR(p)

La représentation VAR à k variables et à p décalages VAR(p) s'écrit sous forme matricielle,

$$Y_t = A_0 + A_1 Y_{t-1} + \dots + A_p Y_{t-p} + \varepsilon_t, \quad (6)$$

avec

$$Y_t = \begin{bmatrix} Y_t^1 \\ Y_t^2 \\ \vdots \\ Y_t^k \end{bmatrix} \quad A_i = \begin{bmatrix} a_{1i}^1 & a_{1i}^2 & \cdots & a_{1i}^k \\ a_{2i}^1 & a_{2i}^2 & \cdots & a_{2i}^k \\ \vdots & \vdots & \ddots & \vdots \\ a_{ki}^1 & a_{ki}^2 & \cdots & a_{ki}^k \end{bmatrix} \quad A_0 = \begin{bmatrix} a_1^0 \\ a_2^0 \\ \vdots \\ a_k^0 \end{bmatrix} \quad \text{et} \quad \varepsilon_t = \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \\ \vdots \\ \varepsilon_t^k \end{bmatrix}.$$

La matrice de covariance des erreurs $\Sigma = \mathbb{E}(\varepsilon_t \varepsilon_t')$ est ici inconnue. De la même façon que dans le cas univarié, on pourra noter

$$A(L) Y_t = A_0 + \varepsilon_t,$$

où A est un polynôme matriciel ($k \times k$), $A(z) = \mathbb{I} - A_1z - A_2z^2 - \dots - A_pz^p$, et on appera **polynôme caractéristique** le polynôme $\det(\mathbb{I} - A_1z - A_2z^2 - \dots - A_pz^p)$.

Remarque 5 L'écriture VAR présentée ici suppose que la i ème équation, définissant Y_t^i ne fasse intervenir aucun Y_t^j . Ceci n'est pas forcément toujours le cas : considérons les deux processus stationnaires (X_t) et (Y_t) définis par les relations suivantes

$$\begin{cases} X_t = \alpha + \beta_1 X_{t-1} + \dots + \beta_p X_{t-p} - \delta Y_t + \gamma_1 Y_{t-1} + \dots + \gamma_q Y_{t-q} + \varepsilon_t \\ Y_t = a + b_1 Y_{t-1} + \dots + b_r Y_{t-p} - d X_t + c_1 X_{t-1} + \dots + c_s X_{t-s} + \eta_t, \end{cases}$$

qui peut se réécrire sous forme matricielle $DZ_t = A + \Phi_1 Z_{t-1} + \dots + \Phi_n Z_{t-n} + u_t$ où $n = \max\{p, q, r, s\}$ avec les notations matricielles

$$Z_t = \begin{bmatrix} X_t \\ Y_t \end{bmatrix}, \quad u_t = \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}, \quad D = \begin{bmatrix} 1 & \delta \\ d & 1 \end{bmatrix}, \quad A = \begin{bmatrix} \alpha \\ a \end{bmatrix} \text{ et pour } i = 1, \dots, n \quad A_i = \begin{bmatrix} \beta_i & \gamma_i \\ b_i & c_i \end{bmatrix},$$

avec la convention, par exemple, $\beta_i = 0$ pour $p < i \leq n$. On supposera que les bruits (ε_t) et (η_t) sont non-corrélos. Il est alors possible d'écrire l'équation sous forme "réduite", obtenue en multipliant par D^{-1} :

$$Z_t = \Psi + \Psi_1 Z_{t-1} + \dots + \Psi_n Z_{t-n} + v_t \text{ où } v_t = B^{-1} u_t.$$

On constate que les innovations (v_t) sont alors fonctions des innovations de la forme structurelle, (ε_t) et (η_t) , et peuvent être corrélés :

$$v_t = \left(\frac{\varepsilon_t - \delta \eta_t}{1 - \delta d}, \frac{\eta_t - d \varepsilon_t}{1 - \delta d} \right)'.$$

Si chacunes des composantes sont i.i.d. les variances sont alors respectivement $[\sigma_\varepsilon^2 + \delta \sigma_\eta^2] / [1 - \delta d]^2$ et $[\sigma_\eta^2 + d \sigma_\varepsilon^2] / [1 - \delta d]$. On constate que l'on peut alors toujours se ramener à des processus de la forme (6).

Définition 4 Le processus Y_t est **stationnaire** (au second ordre) si

- (i) $\mathbb{E}(Y_t) = \mu$ pour tout t
- (ii) $V(Y_t)$ est finie et constante
- (iii) $\text{cov}(Y_t, Y_{t+k}) = \mathbb{E}([Y_t - \mu][Y_{t+k} - \mu]) = \Gamma_k$ pour tout t

Propriété 2 Un processus VAR(p) est stationnaire si le polynôme caractéristique (défini à partir du déterminant $\det(\mathbb{I} - A_1z - A_2z^2 - \dots - A_pz^p)$) a ses racines à l'extérieur du cercle unité.

Preuve. Hamilton (1994) page 259 ■

Exemple 7 Le processus bivarié défini par

$$\begin{bmatrix} Y_t^1 \\ Y_t^2 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 0.7 & 0.4 \\ 0.2 & 0.3 \end{bmatrix} \begin{bmatrix} Y_{t-1}^1 \\ Y_{t-1}^2 \end{bmatrix} + \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{bmatrix},$$

a pour polynôme caractéristique

$$\det \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.7 & 0.4 \\ 0.2 & 0.3 \end{bmatrix} z \right) = \begin{vmatrix} 1 - 0.7z & -0.4z \\ -0.2z & 1 - 0.3z \end{vmatrix} = 1 - z + 0.13z^2,$$

qui admet pour racines $z_1 = 0.84$ et $z_2 = -0.15$: le processus n'est pas stationnaire.

Exemple 8 Le processus défini par

$$\begin{bmatrix} Y_t^1 \\ Y_t^2 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 0.2 & 0.4 \\ 0.7 & 0.3 \end{bmatrix} \begin{bmatrix} Y_{t-1}^1 \\ Y_{t-1}^2 \end{bmatrix} + \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{bmatrix},$$

a pour polynôme caractéristique

$$\det \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.2 & 0.4 \\ 0.7 & 0.3 \end{bmatrix} z \right) = \begin{vmatrix} 1 - 0.2z & -0.4z \\ -0.7z & 1 - 0.3z \end{vmatrix} = 1 - 0.5z - 0.22z^2,$$

qui admet pour racines conjuguées $z_1 = 0.25 - 0.79i$ et $z_2 = 0.25 + 0.79i$: le processus n'est pas stationnaire.

Exemple 9 Le processus défini par

$$\begin{bmatrix} Y_t^1 \\ Y_t^2 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 0.2 & 0.7 \\ 0.3 & 0.4 \end{bmatrix} \begin{bmatrix} Y_{t-1}^1 \\ Y_{t-1}^2 \end{bmatrix} + \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{bmatrix},$$

a pour polynôme caractéristique

$$\det \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.2 & 0.7 \\ 0.3 & 0.4 \end{bmatrix} z \right) = \begin{vmatrix} 1 - 0.2z & -0.7z \\ -0.3z & 1 - 0.4z \end{vmatrix} = 1 - 0.6z + 0.13z^2,$$

qui admet pour racines $z_1 = 1.30$ et $z_2 = -5.91$: le processus est stationnaire.

Les processus $VAR(p)$ peuvent se mettre sous forme $VAR(1)$ en utilisant l'écriture suivante. Soit $Y_t = (Y_t^1, \dots, Y_t^n)^T \in \mathbb{R}^n$ satisfaisant la représentation $VAR(p)$ suivante : pour tout t ,

$$Y_t = \Gamma + A_1 Y_{t-1} + \dots + A_p Y_{t-p} + \varepsilon_t \text{ soit } A(L) Y_t = \Gamma + \varepsilon_t.$$

L'espérance de ce processus est alors $\mathbb{E}(X_t) = \mathbb{E}(\Phi(L)^{-1}[\Gamma + \varepsilon_t]) = \Phi(1)^{-1}\mu = \mu$ pour tout t . On peut alors réécrire $\Phi(L)[Y_t - \mu] = \varepsilon_t$. Considérons alors

$$Z_t = \begin{pmatrix} Y_t - \mu \\ Y_{t-1} - \mu \\ \vdots \\ Y_{t-p+1} - \mu \end{pmatrix}_{(np,1)} \quad A_{(np,np)} = \begin{pmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ \mathbb{I}_n & 0_n & & 0_n & 0_n \\ 0_n & \mathbb{I}_n & & 0_n & 0_n \\ & & \ddots & & \\ 0_n & 0_n & & \mathbb{I}_n & 0_n \end{pmatrix} \text{ et } \eta_t = \begin{pmatrix} \varepsilon_t \\ 0_n \\ \vdots \\ 0_n \end{pmatrix}_{(np,1)}.$$

On peut alors noter que le processus $VAR(p)$ (Y_t) peut se réécrire sous la forme d'un processus transformé Z_t satisfaisant une représentation $VAR(1)$:

$$Z_t = AZ_{t-1} + \eta_t.$$

Exemple 10 Considérons le processus bivarié défini par

$$\begin{bmatrix} Y_t^1 \\ Y_t^2 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} + \begin{bmatrix} 0.2 & 0.7 \\ 0.3 & 0.4 \end{bmatrix} \begin{bmatrix} Y_{t-1}^1 \\ Y_{t-1}^2 \end{bmatrix} - \begin{bmatrix} 0.4 & 0.6 \\ 0.1 & 0.8 \end{bmatrix} \begin{bmatrix} Y_{t-2}^1 \\ Y_{t-2}^2 \end{bmatrix} + \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{bmatrix},$$

suivant un modèle $VAR(2)$. Le polynôme autorégressif $A(L)$ est alors donné par

$$\begin{aligned} A(L) &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} -0.2 & -0.7 \\ -0.3 & -0.4 \end{bmatrix} L + \begin{bmatrix} 0.4 & 0.6 \\ 0.1 & 0.8 \end{bmatrix} L^2 \\ &= \begin{bmatrix} 1 - 0.2L + 0.4L^2 & -0.7L + 0.6L^2 \\ -0.3L + 0.1L^2 & 1 - 0.4L + 0.8L^2 \end{bmatrix}. \end{aligned}$$

L'espérance μ du processus (Y_t) est alors donnée par

$$\mu = A(L)^{-1} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 2.59 \\ 1.80 \end{bmatrix}.$$

On a alors

$$\begin{bmatrix} Y_t^1 - 2.59 \\ Y_t^2 - 1.80 \\ Y_{t-1}^1 - 2.59 \\ Y_{t-1}^2 - 1.80 \end{bmatrix} = \begin{bmatrix} -0.2 & -0.7 & 0.4 & 0.6 \\ -0.3 & -0.4 & 0.1 & 0.8 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} Y_{t-1}^1 - 2.59 \\ Y_{t-1}^2 - 1.80 \\ Y_{t-2}^1 - 2.59 \\ Y_{t-2}^2 - 1.80 \end{bmatrix} + \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \\ 0 \\ 0 \end{bmatrix}.$$

1.3.2 Les modèles ARMAX ou VARMA

Les modèles $ARMAX$ sont une généralisation des processus $VAR(p)$, de la même façon que les processus $ARMA(p,q)$ sont une généralisation des processus $AR(p)$:

$$Y_t = A_0 + A_1 Y_{t-1} + \dots + A_p Y_{t-p} + \varepsilon_t + B_1 \varepsilon_{t-1} + \dots + B_q \varepsilon_{t-q},$$

où A_i et B_j sont des matrices $k \times k$. Il est possible de distinguer les processus VMA (obtenus quand $p = 0$) : moyennes mobiles multivariées.

- les processus VAR sont toujours inversibles, et sont stationnaires si les racines du polynôme caractéristique sont à l'extérieur du disque unité

- les processus VMA sont toujours stationnaires, et sont inversibles si les racines du polynôme caractéristique sont à l'extérieur du disque unité

- les conditions d'inversibilité et de stationnarité des processus $ARMA$ dépendent des parties VAR et VMA du processus.

Remarque 6 Il est possible de montrer que les processus $VARMA$ peuvent se mettre sous forme $VAR(1)$. Soit Y_t suivant un modèle $VARMA(p, q)$, tel que

$$Y_t = A_1 Y_{t-1} + \dots + A_p Y_{t-p} + \varepsilon_t - M_1 \varepsilon_{t-1} - \dots - M_q \varepsilon_{t-q}$$

et notons

$$Z_t = \begin{bmatrix} Y_t \\ \vdots \\ Y_{t-p+1} \\ \varepsilon_t \\ \vdots \\ \varepsilon_{t-q-1} \end{bmatrix} \text{ vecteur de taille } K(p+q) \text{ et } U_t = \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \\ \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \text{ de taille } Kp + Kq.$$

On posera alors

$$B = \begin{bmatrix} B_{11} & 0 \\ B_{21} & B_{22} \end{bmatrix} \text{ de taille } K(p+q) \times K(p+q),$$

où les blocs B_{ij} sont définis par

$$B_{11} = \begin{bmatrix} A_1 & \cdots & A_{p-1} & A_p \\ \mathbb{I}_K & & & 0 \\ & \ddots & & \vdots \\ & & \mathbb{I}_K & 0 \end{bmatrix} \text{ de taille } Kp \times Kp,$$

$$B_{12} = \begin{bmatrix} M_1 & \cdots & M_{q-1} & M_q \\ 0 & & & 0 \\ & \ddots & & \vdots \\ & & 0 & 0 \end{bmatrix} \text{ de taille } Kp \times Kq,$$

$$B_{22} = \begin{bmatrix} 0 & \cdots & 0 & 0 \\ \mathbb{I}_K & & & 0 \\ & \ddots & & \vdots \\ & & \mathbb{I}_K & 0 \end{bmatrix} \text{ de taille } Kq \times Kq.$$

Alors, le processus (Z_t) suit un modèle $VAR(1)$, $Z_t = BZ_{t-1} + U_t$.

1.3.3 Estimation des paramètres d'un modèle VAR

Chacun des paramètres peut être obtenue soit par *mco* (*moindres carrés ordinaires*), soit par maximum de vraisemblance. Pour un modèle VAR stationnaire, la stationnarité de la série va entraîner la convergence et la normalité asymptotique des estimateurs obtenus par *mco*, ce qui permet de mener des tests sur les paramètres du modèle, ou de donner des intervalles de confiance pour les prévisions. Toutefois, comme nous l'avons déjà dit dans le cas univarié, les variables économiques sont souvent intégrées (*d'ordre 1 ou plus*). Dans ce cas, l'écriture (6) est toujours valable, mais le déterminant du polynôme caractéristique admet des racines de module 1. Les coefficients du modèles peuvent toujours être estimés par des *mco* et les estimateurs obtenus sont toujours convergents (*en fait, ils sont même super-convergents puisqu'ils convergent à la vitesse $1/T$ et non pas $1/\sqrt{T}$*). Cependant, ces estimateurs ne sont pas asymptotiquement normaux, et l'on peut plus, dans ce cadre, mener les tests usuels sur les paramètres du modèle, ni déterminer d'intervalle de confiance pour les prévisions.

Cependant, lorsque les variables sont non-stationnaires et cointégrées, les résultats de Engle et Granger (1987) montrent que la bonne spécification du modèle consiste à utiliser une forme à correction d'erreur (*développé dans la partie (1.2.1)*), qui permet de se ramener à une écriture ne faisant intervenir que des variables stationnaires, et dans lesquels il est possible d'effectuer des tests sur les paramètres du modèle.

1.3.4 Autocovariances et autocorrélations

La k -ième autocovariance croisée entre la i -ième et la j -ième variable est donnée par $\gamma_{ij}(h) = \mathbb{E} \left[(Y_t^i - \bar{Y}^i)(Y_{t-h}^j - \bar{Y}^j) \right]$.

Remarque 7 La propriété de symétrie valable dans le cas univarié n'est plus vérifiée ici : pour $h \neq 0$, $\gamma_{ij}(h) \neq \gamma_{ij}(-h)$. En effet, si Y_t dépend fortement de X_{t-1} , ce ne signifie pas que Y_t dépende, de la même façon, de X_{t+1} . Et de même $\gamma_{ij}(h) \neq \gamma_{ji}(h)$

La fonction d'autocorrélation est alors

$$\rho_{ij}(h) = \frac{\gamma_{ij}(h)}{\sqrt{\gamma_i(0)\gamma_j(0)}}.$$

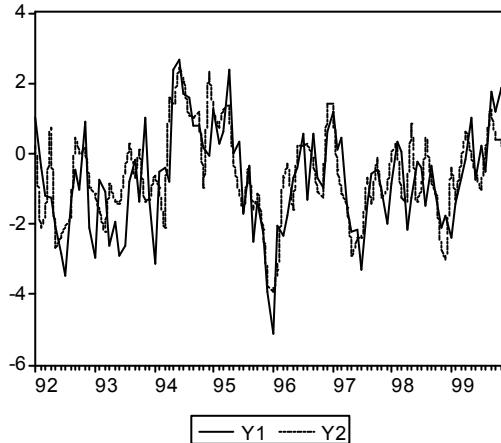
Aussi, on peut définir, pour tout retard h la matrice d'autocorrélation

$$\Xi(h) = \begin{bmatrix} \rho_{11}(h) & \rho_{12}(h) & \cdots & \rho_{1n}(h) \\ \rho_{21}(h) & \rho_{22}(h) & \cdots & \rho_{2n}(h) \\ \vdots & \vdots & & \vdots \\ \rho_{n1}(h) & \rho_{n2}(h) & \cdots & \rho_{nn}(h) \end{bmatrix}.$$

On peut alors noter que $\Xi(h) = \Xi(-h)'$, et que les éléments sur la diagonale correspondent aux autocorrélations usuelles.

1.3.5 Application sur un exemple

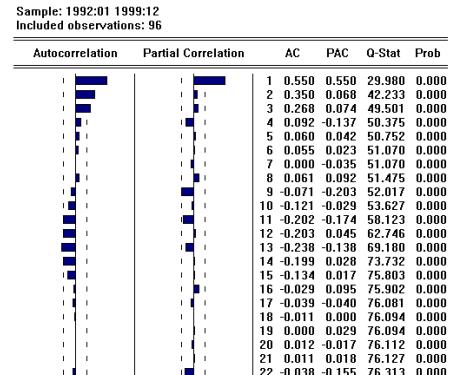
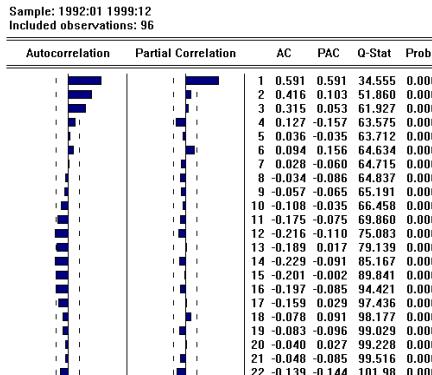
Considérons les séries suivantes, Y_1 et Y_2 ,



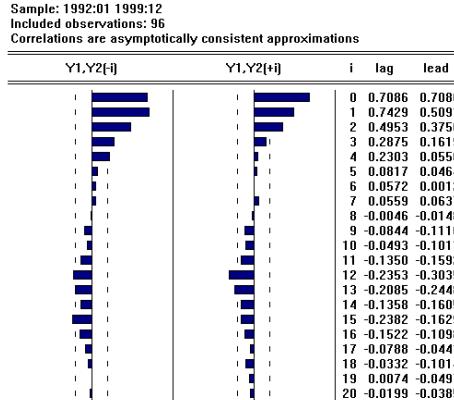
Les observations sont les suivantes,

	Y_1	Y_2		Y_1	Y_2		Y_1	Y_2
1992 - 01	0.999	0.202	1994 - 09	0.806	1.036	1997 - 05	-2.22	-2.913
1992 - 02	-0.415	-2.135	1994 - 10	0.79	1.188	1997 - 06	-2.171	-2.405
1992 - 03	-1.22	-1.734	1994 - 11	0.155	-0.952	1997 - 07	-3.32	-2.394
1992 - 04	-1.251	0.768	1994 - 12	-0.062	2.341	1997 - 08	-1.397	-0.704
1992 - 05	-2.056	-2.648	1995 - 01	1.259	1.138	1997 - 09	-0.555	-1.461
1992 - 06	-2.847	-2.387	1995 - 02	0.263	0.753	1997 - 10	-0.438	-0.134
1992 - 07	-3.446	-2.104	1995 - 03	0.642	1.257	1997 - 11	-1.033	-1.271
1992 - 08	-1.316	-1.869	1995 - 04	2.391	1.385	1997 - 12	-1.972	-1.024
1992 - 09	-0.436	0.443	1995 - 05	-0.032	-0.253	1998 - 01	-0.848	0.048
1992 - 10	-1.033	-0.01	1995 - 06	0.316	-0.994	1998 - 02	0.337	0.3
1992 - 11	0.889	0.157	1995 - 07	-1.717	-1.68	1998 - 03	0.074	-1.264
1992 - 12	-2.096	-0.949	1995 - 08	-0.329	-0.398	1998 - 04	-2.177	-1.406
1993 - 01	-2.963	-1.135	1995 - 09	-2.52	-1.584	1998 - 05	-1.263	0.855
1993 - 02	-0.714	-1.672	1995 - 10	-1.36	-1.163	1998 - 06	-0.246	-1.389
1993 - 03	-1.109	-2.234	1995 - 11	-2.264	-2.163	1998 - 07	-0.391	-1.063
1993 - 04	-2.626	-0.845	1995 - 12	-3.935	-3.727	1998 - 08	-1.471	0.434
1993 - 05	-1.924	-1.35	1996 - 01	-5.133	-3.934	1998 - 09	-0.363	-0.776
1993 - 06	-2.899	-1.449	1996 - 02	-2.051	-3.298	1998 - 10	-1.15	-1.124
1993 - 07	-2.627	-0.523	1996 - 03	-2.321	-0.705	1998 - 11	-2.115	-2.718
1993 - 08	-0.824	0.281	1996 - 04	-1.767	-0.308	1998 - 12	-1.75	-3.002
1993 - 09	-0.16	-0.695	1996 - 05	-0.696	-1.596	1999 - 01	-2.367	-0.387
1993 - 10	-1.395	0.117	1996 - 06	-0.393	0.226	1999 - 02	-1.398	-1.295
1993 - 11	1.019	-1.347	1996 - 07	0.586	0.203	1999 - 03	-0.7	-0.258
1993 - 12	-1.303	-1.329	1996 - 08	-1.334	0.278	1999 - 04	-0.147	0.663
1994 - 01	-3.123	-0.623	1996 - 09	0.559	-0.143	1999 - 05	1.01	0.000
1994 - 02	-0.533	-0.975	1996 - 10	-0.696	-1.111	1999 - 06	-0.728	-0.535
1994 - 03	-0.427	-2.129	1996 - 11	-0.979	-1.249	1999 - 07	0.222	-1.054
1994 - 04	-0.793	1.596	1996 - 12	0.615	1.395	1999 - 08	-0.496	0.129
1994 - 05	2.38	1.407	1997 - 01	1.173	1.409	1999 - 09	1.734	1.292
1994 - 06	2.66	2.467	1997 - 02	0.115	-0.464	1999 - 10	1.217	0.395
1994 - 07	1.726	2.019	1997 - 03	0.427	-1.125	1999 - 11	1.899	0.407
1994 - 08	1.596	1.063	1997 - 04	-1.618	-1.52	1999 - 12	0.373	-0.227

L'analyse univariée des séries Y_1 et Y_2 donne les autocorrélogrammes suivants,



L'autocorrélogramme croisé permet de représenter $\rho_{Y_2 \rightarrow Y_1}(h) = \text{corr}(Y_t^1, Y_{t-h}^2)$ et $\rho_{Y_1 \rightarrow Y_2}(h) = \text{corr}(Y_t^2, Y_{t-h}^1)$,



Estimation d'un modèle VAR(1) avec constante Estimons le modèle suivant sur notre échantillon,

$$\begin{bmatrix} Y_t^1 \\ Y_t^2 \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} Y_{t-1}^1 \\ Y_{t-1}^2 \end{bmatrix} + \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{bmatrix},$$

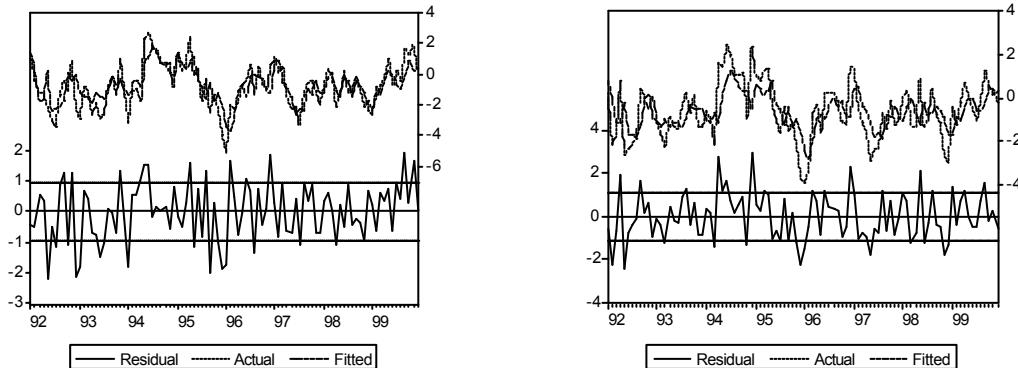
qui peut se réécrire

$$\begin{cases} Y_t^1 = a_1 + a_{11}Y_{t-1}^1 + a_{12}Y_{t-1}^2 + \varepsilon_t^1 \\ Y_t^2 = a_2 + a_{21}Y_{t-1}^1 + a_{22}Y_{t-1}^2 + \varepsilon_t^2. \end{cases}$$

Pour estimer les paramètres de ce modèle, nous pouvons utiliser les *mco* sur chacune des équations indépendamment :

LS // Dependent Variable is Y1				LS // Dependent Variable is Y2					
Sample: 1992:01 1999:12				Sample: 1992:01 1999:12					
Included observations: 96				Included observations: 96					
Variable	Coefficient	Std. Error	T-Statistic	Variable	Coefficient	Std. Error	T-Statistic		
C	-0.219299	0.111804	-1.961460	0.0528	C	-0.239672	0.126545	-1.893965	0.0613
Y1(-1)	0.132401	0.096128	1.377344	0.1717	Y1(-1)	0.219594	0.108802	2.018300	0.0464
Y2(-1)	0.712173	0.104954	6.785548	0.0000	Y2(-1)	0.369461	0.118792	3.110149	0.0025
R-squared	0.574829	Mean dependent var	-0.741438		R-squared	0.333872	Mean dependent var	-0.620939	
Adjusted R-squared	0.565686	S.D. dependent var	1.481991		Adjusted R-squared	0.319547	S.D. dependent var	1.340096	
S.E. of regression	0.976670	Akaike info criterion	-0.016462		S.E. of regression	1.105439	Akaike info criterion	0.231237	
Sum squared resid	88.71123	Schwarz criterion	0.063674		Sum squared resid	113.6457	Schwarz criterion	0.311373	
Log likelihood	-132.4279	F-statistic	62.86777		Log likelihood	-144.3175	F-statistic	23.30639	
Durbin-Watson stat	2.099930	Prob(F-statistic)	0.000000		Durbin-Watson stat	1.910206	Prob(F-statistic)	0.000000	

dont les sorties graphiques donnent respectivement



Les résidus estimés, $\hat{\varepsilon}_t^1$ et $\hat{\varepsilon}_t^2$ sont alors estimés (*représentés ci-dessus*). Leur matrice de variance-covariance est alors

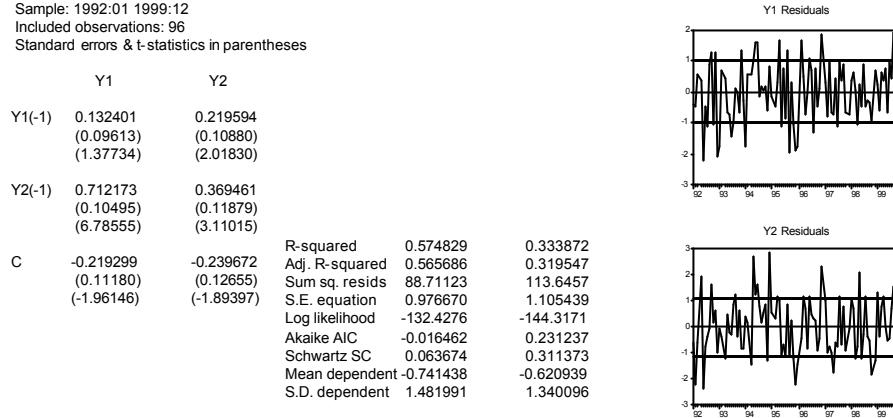
$$\hat{\Sigma} = \begin{bmatrix} 0.924075 & 0.545083 \\ 0.545083 & 1.183809 \end{bmatrix},$$

d'où les expressions des critères AIC et SC ,

$$AIC(1) = \ln |\widehat{\Sigma}| + 2\frac{k^2 1}{n} = \ln 0.796051 + 2\frac{2^2}{96} = -0.144759,$$

$$SC(1) = \ln |\widehat{\Sigma}| + \frac{k^2 1 \ln n}{n} = \ln 0.796051 + \frac{2^2 \ln 96}{96} = -0.037911.$$

Cette estimation peut être faite directement sous *EViews*, de façon globale



Toutefois, si l'on considère les deux régressions faite au début, on peut noter que la constante n'est pas significative (*à un seuil de 5%*) : on peut tester un modèle $VAR(1)$ sans constante.

Estimation d'un modèle $VAR(1)$ sans constante Estimons le modèle suivant sur notre échantillon,

$$\begin{bmatrix} Y_t^1 \\ Y_t^2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} Y_{t-1}^1 \\ Y_{t-1}^2 \end{bmatrix} + \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{bmatrix},$$

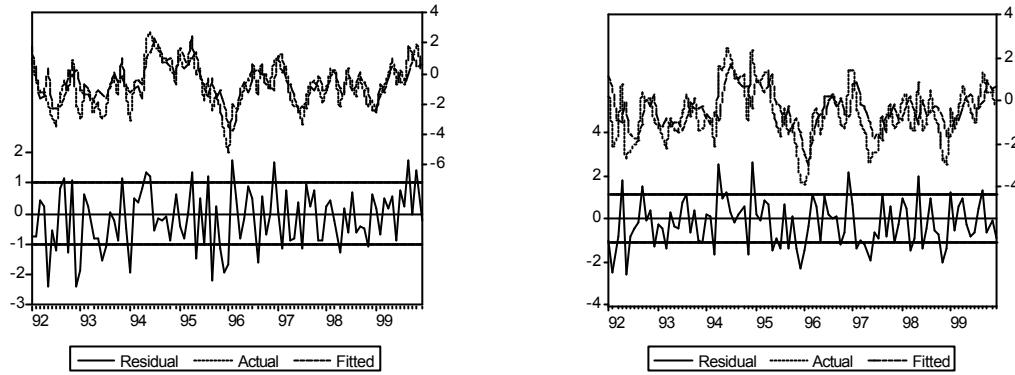
qui peut se réécrire

$$\begin{cases} Y_t^1 = a_1 + a_{11}Y_{t-1}^1 + a_{12}Y_{t-1}^2 + \varepsilon_t^1 \\ Y_t^2 = a_2 + a_{21}Y_{t-1}^1 + a_{22}Y_{t-1}^2 + \varepsilon_t^2. \end{cases}$$

Pour estimer les paramètres de ce modèle, nous pouvons utiliser les *mco* sur chacune des équations indépendamment :

LS // Dependent Variable is Y1				LS // Dependent Variable is Y2					
Sample: 1992:01 1999:12 Included observations: 96				Sample: 1992:01 1999:12 Included observations: 96					
Variable	Coefficient	Std. Error	T-Statistic	Prob.	Variable	Std. Error	T-Statistic	Prob.	
Y1(-1)	0.175336	0.095009	1.845459	0.0681	Y1(-1)	0.266518	0.107391	2.481746	0.0148
Y2(-1)	0.734900	0.105881	6.940826	0.0000	Y2(-1)	0.394300	0.119680	3.294631	0.0014
R-squared	0.557240	Mean dependent var		-0.741438	R-squared	0.308179	Mean dependent var		-0.620939
Adjusted R-squared	0.552530	S.D. dependent var		1.481991	Adjusted R-squared	0.300819	S.D. dependent var		1.340096
S.E. of regression	0.991352	Akaike info criterion		0.003241	S.E. of regression	1.120548	Akaike info criterion		0.248250
Sum squared resid	92.38113	Schwarz criterion		0.056665	Sum squared resid	118.0291	Schwarz criterion		0.301674
Log likelihood	-134.3737	F-statistic		118.3047	Log likelihood	-146.1341	F-statistic		41.87322
Durbin-Watson stat	2.133717	Durbin-Watson stat		0.000000	Durbin-Watson stat	1.929628	Prob(F-statistic)		0.000000

dont les sorties graphiques donnent respectivement



Les résidus estimés, $\hat{\varepsilon}_t^1$ et $\hat{\varepsilon}_t^2$ sont alors estimés (*représentés ci-dessus*). Leur matrice de variance-covariance est alors

$$\Sigma = \begin{bmatrix} 0.931916 & 0.553652 \\ 0.553652 & 1.193174 \end{bmatrix},$$

d'où les expressions des critères AIC et SC ,

$$AIC(1) = \ln |\Sigma| + 2 \frac{k^2 1}{n} = \ln 0.805407 + 2 \frac{2^2}{96} = -0.216407,$$

$$SC(1) = \ln |\Sigma| + \frac{k^2 1 \ln n}{n} = \ln 0.805407 + \frac{2^2 \ln 96}{96} = -0.0262263.$$

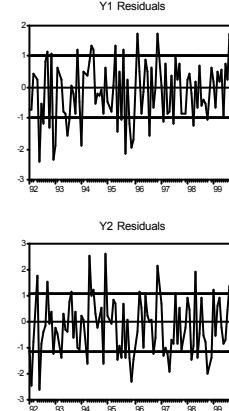
Cette estimation peut être faite directement sous EViews, de façon globale

```
Sample: 1992:01 1999:12
Included observations: 96
Standard errors & t-statistics in parentheses

Y1          Y2
Y1(-1)    0.175336    0.266518
           (0.09501)   (0.10739)
           (1.84546)   (2.48175)

Y2(-1)    0.734900    0.394300
           (0.10588)   (0.11968)
           (6.94083)   (3.29463)

R-squared      0.557240    0.308179
Adj. R-squared  0.552530    0.300819
Sum sq. resids 92.38113   118.0291
S.E. equation   0.991352   1.120548
Log likelihood  -134.3733  -146.1337
Akaike AIC     0.003241   0.248250
Schwartz SC    0.056665   0.301674
Mean dependent -0.741438  -0.620939
S.D. dependent 1.481991   1.340096
```



Comparaison de différents modèles $VAR(p)$ Pour différents modèles $VAR(p)$, les critères AIC et SC sont

	$ \Sigma $	AIC^{Y_1}	AIC^{Y_2}	AIC^{tot}	SC^{Y_1}	SC^{Y_2}	SC^{tot}
$VAR(1)$ avec constante	0.796051	-0.016462	0.231237	-0.144759	0.063674	0.311373	-0.037911
$VAR(1)$ sans constante	0.805407	0.003241	0.248250	-0.216407	0.056665	0.301674	-0.026226
$VAR(2)$ avec constante	0.741204	0.018655	0.242400	-0.132813	0.152215	0.375960	0.080883
$VAR(2)$ sans constante	0.749653	0.038084	0.254792	-0.121478	0.144932	0.361640	0.092217
$VAR(3)$ avec constante		0.059019	0.250520		0.246003	0.437504	
$VAR(3)$ sans constante		0.074667	0.276974		0.234939	0.437246	
$VAR(4)$ avec constante		0.092828	0.289736		0.333236	0.530143	
$VAR(4)$ sans constante		0.111325	0.318514		0.325021	0.532210	

La comparaison des critères AIC pousse à choisir un modèle $VAR(1)$ sans constante. Le modèle retenu

$$\begin{bmatrix} Y_t^1 \\ Y_t^2 \end{bmatrix} = \begin{bmatrix} 0.175 & 0.735 \\ 0.267 & 0.394 \end{bmatrix} \begin{bmatrix} Y_{t-1}^1 \\ Y_{t-1}^2 \end{bmatrix} + \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{bmatrix} \text{ où } \widehat{\Sigma} = \begin{bmatrix} 0.931916 & 0.553652 \\ 0.553652 & 1.193174 \end{bmatrix}.$$

1.3.6 Prévision à l'aide des modèles VAR

Expression de la prévision à horizon h Considérons un modèle $VAR(1)$, $Y_t = A_0 + A_1 Y_{t-1} + \varepsilon_t$, estimé à partir de T observations. La prévision à horizon 1 faite en T est alors

$$\hat{Y}_T(1) = \hat{A}_0 + \hat{A}_1 Y_T.$$

La prévision à horizon 2 est alors

$$\hat{Y}_T(2) = \hat{A}_0 + \hat{A}_1 \hat{Y}_T(1) = \hat{A}_0 + \hat{A}_1 [\hat{A}_0 + \hat{A}_1 Y_T] = \hat{A}_0 + \hat{A}_1 \hat{A}_0 + \hat{A}_1^2 Y_T.$$

De façon plus générale, la prévision à horizon h est alors

$$\hat{Y}_T(h) = \hat{A}_0 + \hat{A}_1 \hat{Y}_T(h-1) = [\mathbb{I} + \hat{A}_1 + \dots + \hat{A}_1^{h-1}] \hat{A}_0 + \hat{A}_1^h Y_T.$$

L'espérance de l'erreur de prévision est nulle, et sa variance est donnée par $\Sigma_1 = \Sigma$ à horizon 1. À horizon 2, cette variance est

$$\Sigma_2 = \Sigma_1 + \hat{A}_1 \Sigma \hat{A}'_1 = \Sigma + \hat{A}_1 \Sigma \hat{A}'_1,$$

et à horizon 3, elle vaut

$$\Sigma_2 = \Sigma_2 + \hat{A}_1^2 \Sigma \hat{A}'_1 = \Sigma + \hat{A}_1 \Sigma \hat{A}'_1 + \hat{A}_1^2 \Sigma \hat{A}'_1.$$

De façon plus générale, la variance de la prévision à horizon h est

$$\Sigma_h = \Sigma_{h-1} + \hat{A}_1^{h-1} \Sigma \hat{A}'_1 = \Sigma + \hat{A}_1 \Sigma \hat{A}'_1 + \hat{A}_1^2 \Sigma \hat{A}'_1 + \dots + \hat{A}_1^{h-1} \Sigma \hat{A}'_1.$$

Des formules analogues existent pour les processus $VAR(p)$. La variance de l'erreur de prévision de chacune des k variables est obtenue sur la diagonale des Σ_h : l'intervalle de confiance de la prévision au niveau $1 - \alpha/2$ est donnée par

$$\hat{Y}_T^i(h) \pm t_{\alpha/2} \cdot \sqrt{\Sigma_h^{i,i}},$$

où $t_{\alpha/2}$ est le fractile de la loi normale.

Application sur l'exemple précédent Le modèle retenu était

$$\begin{bmatrix} Y_t^1 \\ Y_t^2 \end{bmatrix} = \begin{bmatrix} 0.175 & 0.735 \\ 0.267 & 0.394 \end{bmatrix} \begin{bmatrix} Y_{t-1}^1 \\ Y_{t-1}^2 \end{bmatrix} + \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{bmatrix} \text{ où } \Sigma = \begin{bmatrix} 0.931916 & 0.553652 \\ 0.553652 & 1.193174 \end{bmatrix}.$$

On obtient alors la prévision, faite à fin décembre 1999, pour le mois de janvier 2000, de la forme

$$\begin{cases} \hat{Y}_{2000-01}^1 = 0.175336 \cdot Y_{1999-12}^1 + 0.7349 \cdot Y_{1999-12}^2 \\ \hat{Y}_{2000-01}^2 = 0.266518 \cdot Y_{1999-12}^1 + 0.3943 \cdot Y_{1999-12}^2. \end{cases}$$

En utilisant cette estimation, on obtient de façon récursive

$$\begin{cases} \hat{Y}_{2000-02}^1 = 0.175336 \cdot \hat{Y}_{2000-01}^1 + 0.7349 \cdot \hat{Y}_{2000-01}^2 \\ \hat{Y}_{2000-02}^2 = 0.266518 \cdot \hat{Y}_{2000-01}^1 + 0.3943 \cdot \hat{Y}_{2000-01}^2, \\ \\ \hat{Y}_{2000-03}^1 = 0.175336 \cdot \hat{Y}_{2000-02}^1 + 0.7349 \cdot \hat{Y}_{2000-02}^2 \\ \hat{Y}_{2000-03}^2 = 0.266518 \cdot \hat{Y}_{2000-02}^1 + 0.3943 \cdot \hat{Y}_{2000-02}^2. \end{cases}$$

La variance de l'erreur de prévision est donnée par Σ_h , soit

$$\Sigma_h = \Sigma = \begin{bmatrix} 0.931916 & 0.553652 \\ 0.553652 & 1.193174 \end{bmatrix}.$$

1.3.7 Tests de bruit blanc des erreurs

De la même façon que pour les modèles AR , il convient de vérifier que les erreurs correspondent à un bruit blanc. Soit (ε_t) le processus d'erreur, et $\Xi(h)$ sa fonction d'autocorrélation. L'hypothèse à tester est alors

$$H_0 : \Xi(1) = \dots = \Xi(h) = 0.$$

La statistique Q de Box & Pierce peut alors se généraliser en multivarié,

$$Q(h) = n \sum_{i=1}^h \text{trace} \left(\widehat{\Xi}(i)' \widehat{\Xi}(0)^{-1} \widehat{\Xi}(i) \widehat{\Xi}(0)^{-1} \right).$$

La statistique de Ljung & Box s'écrit

$$Q'(h) = n^2 \sum_{i=1}^h \frac{1}{n-i} \text{trace} \left(\widehat{\Xi}(i)' \widehat{\Xi}(0)^{-1} \widehat{\Xi}(i) \widehat{\Xi}(0)^{-1} \right).$$

La distribution asymptotique suit une loi du chi-deux.

1.3.8 Test de causalité (*Granger (1969)*) sur des modèles VAR

En dimension 2, considérons deux processus (X_t) et (Y_t) . Le but est de se demander si (X_t) cause (Y_t) , et de voir dans quelle proportion la valeur courante de (Y_t) peut être expliquée par ses valeurs passées, et si en ajoutant des valeurs retardées de (X_t) , l'explication est meilleure.

Considérons un modèle autorégressif,

$$\Phi(L) Z_t = \mu + u_t \text{ soit } \begin{bmatrix} \Phi_X(L) & \Phi_{XY}(L) \\ \Phi_{YX}(L) & \Phi_Y(L) \end{bmatrix} \begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix},$$

où on supposera $\Phi(0) = I$, ce qui permettra d'interpréter le vecteur u_t comme le processus d'innovation.

- (X_t) ne cause pas (Y_t) au sens de Granger si et seulement si les coefficients de Φ_{YX} sont nuls : $\Phi_{YX}(L) = 1$
- (Y_t) ne cause pas (X_t) au sens de Granger si et seulement si les coefficients de Φ_{XY} sont nuls : $\Phi_{XY}(L) = 1$

Le test peut se faire sous *EViews* (*Granger Causality*) : pour cela, on commence par choisir le nombre de retards à introduire, et l'hypothèse nulle testée est que (Y_t) ne cause pas (X_t) dans la première équation, et que (X_t) ne cause pas (Y_t) dans la seconde.

Considérons ici deux vecteurs (*ou variables*) (X_t) et (Y_t) , respectivement de taille n et m , de telle sorte que $n + m = K$, de telle sorte que le processus $(Z_t) = (X_t, Y_t)'$ soit un processus $VAR(p)$:

$$Z_t = M + A_1 Z_{t-1} + A_2 Z_{t-2} + \dots + A_p Z_{t-p} + \varepsilon_t \text{ ou encore } Z_t^* = B Z_{t-1}^* + U_t,$$

en reprenant les notations de la remarque (6), c'est à dire

$$B = \begin{bmatrix} M & A_1 & \dots & A_{p-1} & A_p \end{bmatrix} \text{ de taille } K \times (Kp+1), \quad Z_t^* = \begin{bmatrix} 1 \\ Z_t \\ Z_{t-1} \\ \vdots \\ Z_{t-p+1} \end{bmatrix} \text{ et } U_t = \begin{bmatrix} 0 \\ \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Le test de l'hypothèse de causalité est un test de Wald : on cherche à tester $(H_0) : R\beta = c$ contre $(H_1) : R\beta \neq c$, où C est une matrice $N \times (K^2 p + K)$ de rang N et c est un vecteur \mathbb{R}^N . Supposons que $\widehat{\beta}$, estimateur de β , vérifie une propriété de normalité asymptotique, c'est à dire $\sqrt{T}(\widehat{\beta} - \beta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma^{-1} \Sigma)$, alors $\sqrt{T}(R\widehat{\beta} - R\beta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, R(\Gamma^{-1} \Sigma) R')$ et on a alors

$$T(R\beta - c)' [R(\Gamma^{-1} \Sigma) R']^{-1} (R\beta - c) \xrightarrow{\mathcal{L}} \chi^2(N).$$

En remplaçant les matrices Γ et Σ par leur estimateurs usuels

$$\widehat{\Gamma} = \frac{1}{T} ZZ' \text{ et } \widehat{\Sigma} = \frac{1}{T - Kp - 1} \widehat{u}\widehat{u}',$$

on obtient l'estimateur de Wald

$$\lambda = \left(R\hat{\beta} - c \right)' \left[R \left((ZZ')^{-1} \quad \hat{\Sigma} \right) R' \right]^{-1} \left(R\hat{\beta} - c \right),$$

qui suit asymptotiquement un chi-deux à N degrés de liberté. Toutefois, une alternative également utilisée est de considérer λ/N qui suit asymptotiquement une loi de Fisher à N et $T - Kp - 1$ degrés de liberté.

Dans le cas du test de causalité de Granger le coefficient c est nul, de taille $(m \times n) \times p$, β est le vecteur $\text{vec}(B)$ - c'est à dire la concaténation des vecteurs colonnes de B : le modèle de base s'écrivant de la façon suivante

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} M^X \\ M^Y \end{bmatrix} + \begin{bmatrix} A_1^X & A_1^{XY} \\ A_1^{YX} & A_1^Y \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} A_2^X & A_2^{XY} \\ A_2^{YX} & A_2^Y \end{bmatrix} \begin{bmatrix} X_{t-2} \\ Y_{t-2} \end{bmatrix} + \dots + \begin{bmatrix} A_p^X & A_p^{XY} \\ A_p^{YX} & A_p^Y \end{bmatrix} \begin{bmatrix} X_{t-p} \\ Y_{t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_t^X \\ \varepsilon_t^Y \end{bmatrix},$$

on a

$$\begin{aligned} B &= \begin{bmatrix} M^X & A_1^X & A_1^{XY} & A_2^X & A_2^{XY} & \cdots & A_p^X & A_p^{XY} \\ M^Y & A_1^{YX} & A_1^Y & A_2^{YX} & A_2^Y & \cdots & A_p^{YX} & A_p^Y \end{bmatrix} \\ &= \begin{bmatrix} M_1^X & A_{1,11}^X & \cdots & A_{1,n1}^X & A_{1,11}^{XY} & \cdots & A_{1,m1}^{XY} & A_{p,11}^X & \cdots & A_{p,n1}^X & A_{p,11}^{XY} & \cdots & A_{p,m1}^{XY} \\ \vdots & \vdots & & \vdots \\ M_n^X & A_{1,1n}^X & \cdots & A_{1,nn}^X & A_{1,1n}^{XY} & \cdots & A_{1,mn}^{XY} & A_{p,1n}^X & \cdots & A_{p,nn}^X & A_{p,1n}^{XY} & \cdots & A_{p,mn}^{XY} \\ M_1^Y & A_{1,11}^{YX} & \cdots & A_{1,n1}^{YX} & A_{1,11}^Y & \cdots & A_{1,m1}^Y & A_{p,11}^{YX} & \cdots & A_{p,n1}^{YX} & A_{p,11}^Y & \cdots & A_{p,m1}^Y \\ \vdots & \vdots & & \vdots \\ M_m^Y & A_{1,1m}^{YX} & \cdots & A_{1,nm}^{YX} & A_{1,1m}^Y & \cdots & A_{1,mm}^Y & A_{p,1m}^{YX} & \cdots & A_{p,nm}^{YX} & A_{p,1m}^Y & \cdots & A_{p,mm}^Y \end{bmatrix}, \end{aligned}$$

et donc

$$\beta = [M_1^X \cdots M_n^X M_1^Y \cdots M_m^Y A_{1,11}^X \cdots A_{1,1n}^X A_{1,11}^{YX} \cdots A_{1,1m}^{YX} A_{2,11}^X \cdots A_{p,11}^X \cdots A_{p,mm}^Y]' \in \mathbb{R}^{(n+m)^2 p + n + m}.$$

La matrice R est alors la matrice remplie de 0 et de 1 permettant de tester la nullité des coefficients $A_1^{XY}, \dots, A_p^{XY}$, c'est à dire

$$R = \begin{bmatrix} \text{nullité de } A_{1,12}^{XY} & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ \text{nullité de } A_{1,11}^{XY} & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\ \text{nullité de } A_{1,mn}^{XY} & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ & & & & & & & & & & & & & & & & \vdots \\ \text{nullité de } A_{p,11}^{XY} & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\ \text{nullité de } A_{p,mn}^{XY} & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 \\ & & & n+m & & & n \times (n+m) & & & m \times (n+m) & & & m \times (n+m) & & & & \end{bmatrix} \quad (n \times m)$$

On rejette alors (H_0) (*de non-causalité*) quand la puissance du test est très faible (*inférieure à 5%*), et dans le cas contraire, on ne peut pas rejeter l'hypothèse de non-causalité (puissance supérieure à 5%).

1.3.9 Les modèles VAR(p) : analyse des chocs

(Y_t) suit un modèle VAR d'ordre p si et seulement si

$$Y_t = M + A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + \varepsilon_t,$$

où (ε_t) est un bruit blanc de dimension K , de matrice de variance-covariance Σ .

Propriété 3 Ce processus admet une représentation VAR(1) de la forme $Z_t = \mu + AZ_{t-1} + U_t$ où

$$Z_t = \begin{bmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p+1} \end{bmatrix}, \quad \mu = \begin{bmatrix} M \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad U_t = \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{et} \quad A = \begin{bmatrix} A_1 & \cdots & A_{p-1} & A_p \\ \mathbb{I}_K & & 0 & 0 \\ \ddots & & \ddots & \\ 0 & & \mathbb{I}_K & 0 \end{bmatrix}.$$

Remarque 8 En posant $J = [\mathbb{I}_K, 0, \dots, 0]$, on peut écrire $Y_t = JZ_t$

Sous l'hypothèse de stabilité du processus (racines du polynôme caractéristique à l'extérieur du disque unité) la représentation VMA est la suivante (appelée aussi décomposition de Wald)

$$Y_t = \mu + \sum_{i=0}^{\infty} \Theta_i \varepsilon_{t-i} \text{ où } \begin{cases} \Theta_i = JA^i J' \\ \mu = (\mathbb{I}_K - A_1 - \dots - A_p)^{-1} M \end{cases}. \quad (7)$$

Impact d'une impulsion Nous allons étudier ici l'effet d'une innovation d'une des variables (Y_t^i) sur le système de K variables. Pour cela, on suppose que $Y_t = \mu$ pour $t < 0$ et qu'en $t = 0$, la i ème variable augmente de 1 (soit $\varepsilon_t^j = 0$ pour $j \neq i$ et $\varepsilon_t^i = 1$). L'idée est alors d'étudier comme réagit le système aux dates 1, 2, ..etc, si aucun autre choc ne survient. Nous supposerons pour la suite que le processus (Y_t) est centré, et que l'innovation se fait sur la première variable. Enfin, nous noterons $\delta = (1, 0, 0, \dots, 0)'$ le vecteur d'impulsion.

Remarque 9 L'idée générale de l'analyse des chocs (ou fonction de réponse aux innovations) est de résumer l'information concernant l'évolution d'une composante Y_t^j suite à une impulsion sur Y_t^i , à la date $t = 0$, en supposant que toutes les autres variables sont constantes pour $t \leq 0$.

En utilisant l'écriture VMA(∞), on en déduit

$$\begin{aligned} Y_0 &= \sum_{i=0}^{\infty} \Theta_i \varepsilon_{t-i} = \Theta_0 \delta + 0 + 0 + \dots = \delta \quad (\text{car } \Theta_0 = \mathbb{I}), \\ Y_1 &= \sum_{i=0}^{\infty} \Theta_i \varepsilon_{t-i} = 0 + \Theta_1 \delta + 0 + \dots = \Theta_1 \delta, \end{aligned}$$

et plus généralement, $Y_h = \Theta_h \delta$ où la matrice Θ_h est le coefficient du retard d'ordre h de l'écriture VMA(∞), c'est à dire $\Theta_i = JA^i J'$. Aussi, la réponse à une choc sur l'innovation se lit sur les vecteurs colonnes des matrices de la forme moyenne mobile.

On peut également considérer les réponses accumulées. Pour cela, on considère les matrices définies par $\Psi_i = \mathbb{I} + \Theta_1 + \Theta_2 + \dots + \Theta_h$, avec la matrice limite $\Psi_{\infty} = \mathbb{I} + \Theta_1 + \Theta_2 + \dots + \Theta_h + \dots = (\mathbb{I}_K - A_1 - \dots - A_p)^{-1}$. Un exemple sera traité dans la partie suivante.

Orthogonalisation des innovations et décomposition de la variance Dans l'analyse précédante, nous avons supposé qu'il était possible d'avoir, à la date $t = 0$ une impulsion sur l'une des composantes, c'est à dire que le bruit (ε_t) s'écrivait sous la forme $(0, \dots, 0, 1, 0, \dots, 0)'$.

Supposons que la matrice de variance-covariance soit de la forme $\mathbb{E}(\varepsilon_t \varepsilon_t') = \Sigma \neq \mathbb{I}$. Il peut être intéressant (*dans le cas bivarié, pour simplifier*) d'isoler l'innovation propre au processus (Y_t) non polluée par la réaction de l'innovation (X_t) : on parle alors d'orthogonalisation des innovations. On considère alors la décomposition suivante de la matrice de covariance des innovations :

$$\Sigma = \mathbb{E}(\varepsilon_t \varepsilon_t') = ADA' \text{ où } A = \begin{bmatrix} 1 & 0 \\ * & 1 \end{bmatrix} \text{ et } D = \begin{bmatrix} * & 0 \\ 0 & * \end{bmatrix}.$$

Exemple 11 Dans le cas bivarié, on a par exemple la décomposition suivante

$$\begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho\sigma_1/\sigma_2 & 1 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & (1-\rho^2)\sigma_2^2 \end{bmatrix} \begin{bmatrix} 1 & \rho\sigma_1/\sigma_2 \\ 0 & 1 \end{bmatrix}.$$

On pose alors $\eta_t = A^{-1}\varepsilon_t$: on remarque que les innovations (η_t) sont alors des combinaisons linéaires des innovations du modèle initial (ε_t), qui sont indépendantes (*ou tout du moins non corrélées*) :

$$\mathbb{E}(\eta_t \eta_t') = A^{-1}\mathbb{E}(\varepsilon_t \varepsilon_t')(A^{-1})' = A^{-1}\Sigma(A')^{-1} = A^{-1}ADA'(A')^{-1} = D.$$

Un processus bivarié (Y_t) définit par un modèle VAR(1) de la forme suivante

$$\begin{cases} X_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 Y_{t-1} + \varepsilon_t^X \\ Y_t = \beta_0 + \beta_1 X_{t-1} + \beta_2 Y_{t-1} + \varepsilon_t^Y, \end{cases}$$

devient alors

$$\begin{cases} X_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 Y_{t-1} + \eta_t^X \\ Y_t = \beta_0 + \beta_1 X_{t-1} + \beta_2 Y_{t-1} + \rho \sigma_1 / \sigma_2 \cdot \eta_t^X + \eta_t^Y. \end{cases}$$

L'idée est alors de considérer un choc unitaire non pas sur une composante de (ε_t) mais sur la composante (η_t^Y) : à la date $t = 0$, $(\eta_t) = (0, 1)'$.

En partant de la décomposition des résidus en innovations orthogonales, on peut alors calculer la contribution de chaque innovation à la variance totale de l'erreur de prévision du processus (X_t^i) : c'est cela que l'on appelle **décomposition de la variance**. Considérons le processus (X_t) admettant la représentation $VAR(p)$ suivante

$$A(L)Y_t = Y_t - A_0 - A_1 Y_{t-1} - \dots - A_p Y_{t-p} = M + \varepsilon_t.$$

On suppose que le bruit blanc (ε_t) est de matrice de variance-covariance Σ . On suppose que ce processus est stationnaire, de telle sorte qu'il puisse être représenté sous forme la $VMA(\infty)$ suivante

$$Y_t = \mu + \varepsilon_t + \Theta_1 \varepsilon_{t-1} + \dots + \Theta_i \varepsilon_{t-i} + \dots = \mu + \Theta(L) \varepsilon_t.$$

L'erreur de prévision faite en T à horizon h s'écrit sous la forme

$$\begin{aligned} X_{T+h} - \hat{X}_T(h) &= X_{T+h} - EL(X_{T+h}|X_T, X_{T-1}, \dots, X_1) \\ &= X_{T+h} - EL(X_{T+h}|\varepsilon_T, \varepsilon_{T-1}, \dots, \varepsilon_1) = \varepsilon_{T+h} + \Theta_1 \varepsilon_{T+h-1} + \dots + \Theta_{h-1} \varepsilon_{T+1}. \end{aligned}$$

L'espérance vaut alors 0 (*les bruits blancs étant, par définition, des processus centrés*), et la matrice de variance-covariance est alors donnée par

$$\mathbb{E} \left((X_{T+h} - \hat{X}_T(h)) (X_{T+h} - \hat{X}_T(h))' \right) = \Sigma + \Theta_1 \Sigma \Theta_1' + \dots + \Theta_{h-1} \Sigma \Theta_{h-1}'.$$

Cette erreur de prévision est alors exprimée en fonction de la matrice de variance-covariance $\not\propto$ des résidus (*non diagonale*). Comme précédemment, considérons la transformation $\eta_t = A^{-1} \varepsilon_t$ ou $\varepsilon_t = A \eta_t$, où $\Sigma = ADA'$:

$$\varepsilon_t = \begin{bmatrix} \varepsilon_t^1 \\ \vdots \\ \varepsilon_t^d \end{bmatrix} = [a_1, a_2, \dots, a_n] \begin{bmatrix} \eta_t^1 \\ \vdots \\ \eta_t^d \end{bmatrix} \text{ où } a_i \in \mathbb{R}^d.$$

Dès lors

$$\Sigma = \mathbb{E} (\varepsilon_t \varepsilon_t') = a_1 a_1' V(\eta_t^1) + a_2 a_2' V(\eta_t^2) + \dots + a_d a_d' V(\eta_t^d).$$

En substituant alors cette expression dans la variance de la prévision pour un horizon h , cela permet de réexprimer cette variance en fonction de la variance des innovations orthogonales :

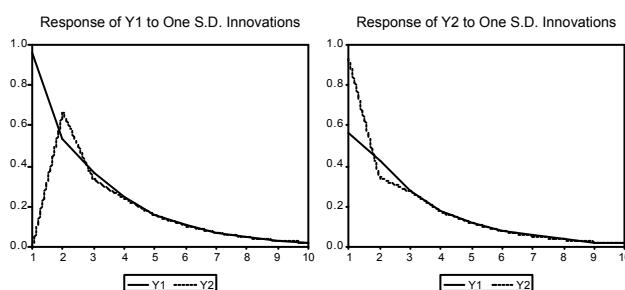
$$\mathbb{E} \left((X_{T+h} - \hat{X}_T(h)) (X_{T+h} - \hat{X}_T(h))' \right) = \sum_{j=1}^d V(\eta_t^j) [\Theta_1(a_1 a_1') \Theta_1' + \dots + \Theta_{h-1}(a_{h-1} a_{h-1}') \Theta_{h-1}'],$$

d'où finalement la définition suivante

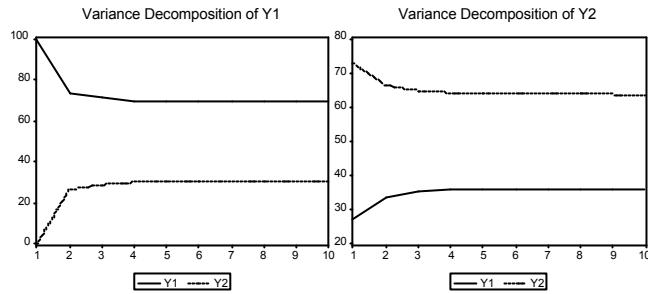
Définition 5 On appelle **contribution d'une innovation pure à la variance totale de la prévision à un horizon h** la quantité

$$V(\eta_t^j) [\Theta_1(a_1 a_1') \Theta_1' + \dots + \Theta_{h-1}(a_{h-1} a_{h-1}') \Theta_{h-1}'].$$

Exemple 12 Sur l'exemple considéré dans la partie précédente, le graphique ci-dessous correspond aux réponses aux fonctions d'impulsion, respectivement sur la variable (Y_1) à gauche et (Y_2) à droite



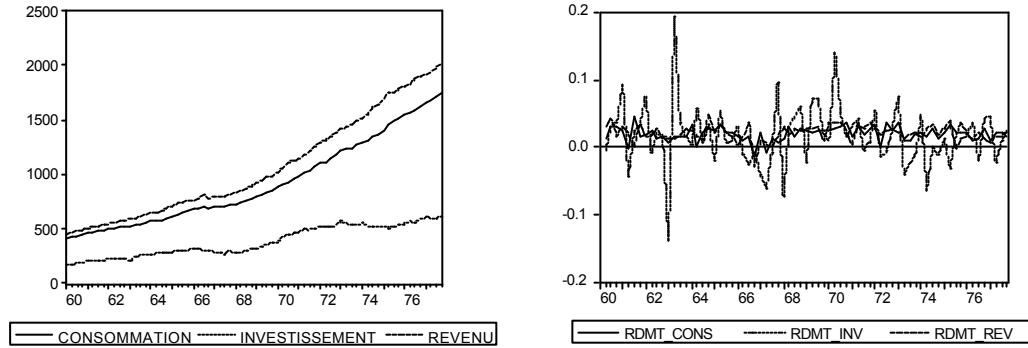
avec la décomposition de la variance ci-dessous



1.4 Application des modèles VAR

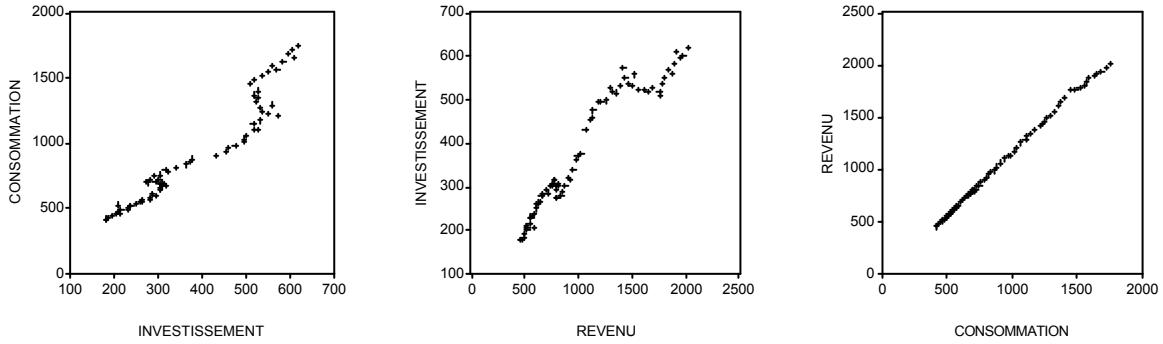
1.4.1 Application : investissement, revenu et consommation

Nous allons considérer ici l'exemple développé par Lütkepohl (1991), *Introduction to Multiple Time Series Analysis*, dans le chapitre 3. La base considérée contient 22 ans de données trimestrielles, avec l'investissement (X), le revenu (Y) et la consommation (Z).



Le graphique ci-dessus à gauche représente, en données brutes, alors que le graphique de droite correspond aux rendements : on posera alors $C_t = \log(Z_t) - \log(Z_{t-1})$, $I_t = \log(X_t) - \log(X_{t-1})$ et $R_t = \log(Y_t) - \log(Y_{t-1})$.

Remarque 10 Un "test" de cointégration peut être mené rapidement, en représentant les scatterplots des différentes variables,



Comme le montrent les graphiques ci-dessus, seul le couple (Z_t, Y_t) semble cointégré.

Le tableau ci-dessous donne différentes statistiques de choix pour des modèles $VAR(p)$ différents⁴ :

Retards	<i>AIC</i>	<i>HQ</i>	<i>SC</i>	<i>FPE</i>
1	-24.50	-24.38	-24.41	2.500
2	-24.59	-24.37	-24.02	2.272
3	-24.41	-24.07	-23.55	2.748
4	-24.36	-23.90	-23.21	2.910

Lütkepohl retenait, en utilisant le critère *AIC* d'Akaike et le *FPE* (*critère que l'on cherche à minimiser*), un modèle $VAR(2)$,

$$Z_t = M + A_1 Z_{t-1} + A_2 Z_{t-2} + \varepsilon_t.$$

L'estimation des paramètres donne les résultats suivants

$$\begin{bmatrix} I_t \\ R_t \\ C_t \end{bmatrix} = \begin{bmatrix} -0.0167 \\ 0.0158 \\ 0.0129 \end{bmatrix} + \begin{bmatrix} -0.3196 & 0.1460 & 0.9612 \\ 0.0439 & -0.1527 & 0.2885 \\ -0.0024 & 0.2248 & -0.2640 \end{bmatrix} \begin{bmatrix} I_{t-1} \\ R_{t-1} \\ C_{t-1} \end{bmatrix} + \begin{bmatrix} -0.1606 & 0.1146 & 0.9344 \\ 0.0500 & 0.0192 & -0.0102 \\ 0.0339 & 0.3549 & -0.0222 \end{bmatrix} \begin{bmatrix} I_{t-2} \\ R_{t-2} \\ C_{t-2} \end{bmatrix} + \begin{bmatrix} \hat{\varepsilon}_t^1 \\ \hat{\varepsilon}_t^2 \\ \hat{\varepsilon}_t^3 \end{bmatrix}, \quad (8)$$

où la matrice de variance-covariance peut être estimée par

$$\hat{\Sigma} = \begin{bmatrix} 212.96 & 7.16 & 12.3 \\ 7.16 & 13.73 & 6.15 \\ 12.3 & 6.15 & 8.92 \end{bmatrix} \times 10^{-5}.$$

La représentation $VMA(\infty)$, permettant de calculer les fonctions d'impulsions, est la suivante

$$\begin{bmatrix} I_t \\ R_t \\ C_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\varepsilon}_t^1 \\ \hat{\varepsilon}_t^2 \\ \hat{\varepsilon}_t^3 \end{bmatrix} + \begin{bmatrix} -0.319 & 0.145 & 0.961 \\ 0.043 & -0.152 & 0.288 \\ -0.002 & 0.224 & -0.263 \end{bmatrix} \begin{bmatrix} \hat{\varepsilon}_{t-1}^1 \\ \hat{\varepsilon}_{t-1}^2 \\ \hat{\varepsilon}_{t-1}^3 \end{bmatrix} + \begin{bmatrix} -0.054 & 0.621 & 0.415 \\ 0.028 & 0.113 & -0.088 \\ 0.045 & 0.260 & 0.109 \end{bmatrix} \begin{bmatrix} \hat{\varepsilon}_{t-2}^1 \\ \hat{\varepsilon}_{t-2}^2 \\ \hat{\varepsilon}_{t-2}^3 \end{bmatrix} + \begin{bmatrix} 0.119 & 0.352 & -0.407 \\ -0.008 & 0.071 & 0.119 \\ 0.000 & -0.098 & 0.090 \end{bmatrix} \begin{bmatrix} \hat{\varepsilon}_{t-3}^1 \\ \hat{\varepsilon}_{t-3}^2 \\ \hat{\varepsilon}_{t-3}^3 \end{bmatrix} + \begin{bmatrix} 0.014 & 0.018 & 0.261 \\ 0.003 & -0.011 & 0.008 \\ 0.005 & 0.845 & -0.015 \end{bmatrix} \begin{bmatrix} \hat{\varepsilon}_{t-4}^1 \\ \hat{\varepsilon}_{t-4}^2 \\ \hat{\varepsilon}_{t-4}^3 \end{bmatrix} + \dots$$

Les fonctions d'autocovariance du processus sont alors données par

$$\Gamma_0 = \begin{bmatrix} 0.00245 & 0.00006 & 0.00014 \\ 0.00006 & 0.00015 & 0.00006 \\ 0.00014 & 0.00006 & 0.00012 \end{bmatrix} \text{ et } \Gamma_1 = \begin{bmatrix} -0.00045 & 0.00007 & 0.00007 \\ 0.00011 & 0.00000 & 0.00003 \\ -0.00002 & 0.00002 & -0.00001 \end{bmatrix},$$

et les fonctions d'autocorrélation sont alors

$$\Upsilon_0 = \begin{bmatrix} 1.0000 & 0.0905 & 0.2578 \\ 0.0905 & 1.0000 & 0.4525 \\ 0.2578 & 0.4525 & 1.0000 \end{bmatrix} \text{ et } \Upsilon_1 = \begin{bmatrix} -0.1981 & 0.1196 & 0.1388 \\ 0.1876 & 0.0128 & 0.2241 \\ -0.0408 & 0.1658 & -0.0875 \end{bmatrix}.$$

⁴où *AIC* est le critère d'Akaike, donné pour m retards, par

$$AIC(m) = \ln(\det(\Sigma_m)) + \frac{2mK^2}{T}$$

où T est le nombre d'observation, m le nombre de retards, K le nombre de variables, et Σ_m est l'estimateur (*du maximum de vraisemblance*) de la matrice de variance-covariance du bruit.

HQ est le critère de Hanna-Quinn, défini par

$$HQ(m) = \ln(\det(\Sigma_m)) + \frac{2\ln(\ln T)}{T} mK^2$$

SC est le critère de Schwarz défini par

$$SC(m) = \ln(\det(\Sigma_m)) + \frac{\ln T}{T} mK^2$$

FPE correspond au critère "final prediction error", définit par

$$FPE(m) = \left[\frac{T+Km+1}{T-Mk-1} \right]^K \det(\Sigma_m)$$

(*FPE* est ici donnée à un facteur multiplicatif près ($\times 10^{11}$)).

Afin de tester la causalité au sens de Granger, par exemple de revenu/consommation vers investissement, on va tester si les matrices A_1 et A_2 sont de la forme

$$A_i = \begin{bmatrix} * & 0 & 0 \\ * & * & * \\ * & * & * \end{bmatrix} \text{ soit } B = [M, A_1, A_2] = \begin{bmatrix} * & * & 0 & 0 & * & 0 & 0 \\ * & * & * & * & * & * & * \\ * & * & * & * & * & * & * \end{bmatrix}.$$

Ces hypothèses peuvent s'écrire sous forme matricielle $Rb = 0$, en considérant que le processus Z_t se subdivise en 2 sous-processus $(C_t, R_t)'$ et (I_t) . On a alors $N = 4$, et $K^2 p + K = 3^2 \times 2 + 3 = 21$: la matrice R (*matrice* 4×21) est alors définie par

$$R = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix},$$

ce qui peut être testé à l'aide de principe de Wald (*test de Student*).

$$\lambda = \frac{1}{4} (R\hat{\beta} - c)' [R((ZZ')^{-1} \hat{\Sigma}) R']^{-1} (R\hat{\beta} - c) = 1.59,$$

alors que le quantile à 95% de la loi de Fisher $F(4, 66)$ vaut 2.5. Aussi, au niveau 5%, on ne peut pas rejeter l'hypothèse de non-causalité des variables revenu et investissement sur la consommation. En revanche, un test de non-causalité instantanée revient à tester si Σ est de la forme

$$\Sigma = \begin{bmatrix} * & 0 & 0 \\ 0 & * & * \\ 0 & * & * \end{bmatrix}.$$

La prévision peut se faire en utilisant l'écriture *VAR* (2), estimée par (8), à compte tenu du fait que les deux derniers points sont

$$Z_{T-1} = \begin{bmatrix} 0.02551 \\ 0.02434 \\ 0.01319 \end{bmatrix} \text{ et } Z_T = \begin{bmatrix} 0.03637 \\ 0.00517 \\ 0.00599 \end{bmatrix}.$$

La prévision optimale se fait alors, de façon récursive,

$$\hat{Z}_T(1) = M + A_1 Z_T + A_2 Z_{T-1} = \begin{bmatrix} -0.011 \\ 0.020 \\ 0.022 \end{bmatrix}, \quad \hat{Z}_T(2) = M + A_1 \hat{Z}_T(1) + A_2 Z_T = \begin{bmatrix} 0.011 \\ 0.020 \\ 0.015 \end{bmatrix} \dots \text{etc.}$$

L'estimateur de la matrice de variance covariance se faire alors, au rang 1, de la façon suivante

$$\hat{\Sigma}_Y(1) = \frac{K + Kp + 1}{T} \hat{\Sigma}_\varepsilon = \frac{73 + 6 + 1}{73} \hat{\Sigma}_\varepsilon = \begin{bmatrix} 23.34 & 0.785 & 1.351 \\ 0.785 & 1.505 & 0.674 \\ 1.351 & 0.674 & 0.978 \end{bmatrix} \times 10^{-4}.$$

Les matrices de variance covariance à des ordres plus élevés se calculent à l'aide de la forme *VMA* (∞), et donnent

$$\hat{\Sigma}_Y(2) = \begin{bmatrix} 25.12 & 0.580 & 1.300 \\ 0.580 & 1.581 & 0.586 \\ 1.300 & 0.586 & 1.009 \end{bmatrix} \times 10^{-4},$$

d'où, finallement, les prévisions suivantes :

$$\text{horizon 1 : } \hat{Z}_T(1) \in \begin{bmatrix} -0.011 \pm 0.095 \\ 0.020 \pm 0.024 \\ 0.022 \pm 0.019 \end{bmatrix} \text{ et horizon 2 : } \hat{Z}_T(2) \in \begin{bmatrix} 0.011 \pm 0.098 \\ 0.020 \pm 0.025 \\ 0.015 \pm 0.020 \end{bmatrix}.$$

Nous allons maintenant étudier les fonctions réponses à une impulsion. Pour cela, rappelons que les deux premières composantes de la forme *VMA* (∞) sont

$$\hat{\Theta}_1 = \begin{bmatrix} -0.319 & 0.145 & 0.961 \\ 0.043 & -0.152 & 0.288 \\ -0.002 & 0.224 & -0.263 \end{bmatrix} = \hat{A}_1 \text{ et } \hat{\Theta}_2 = \begin{bmatrix} -0.054 & 0.621 & 0.415 \\ 0.028 & 0.113 & -0.088 \\ 0.045 & 0.260 & 0.109 \end{bmatrix} = \hat{A}_1 \hat{\Theta}_1 + \hat{A}_2.$$

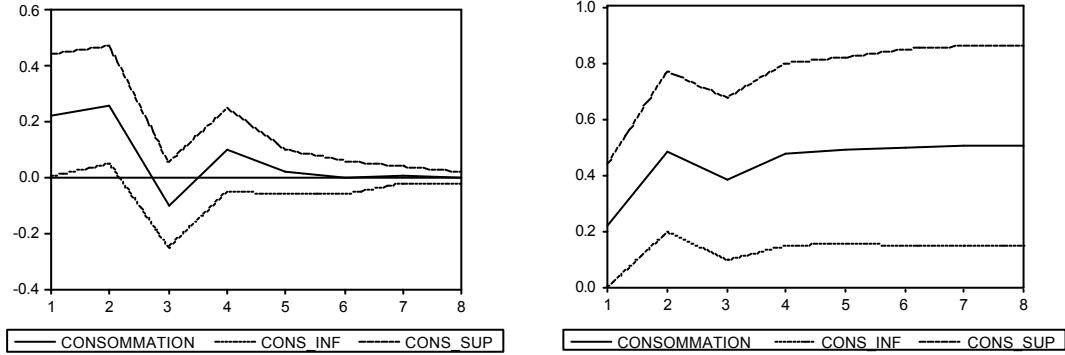
En notant $\widehat{\Theta}_0 = I$, et $\widehat{\Psi}_i = \widehat{\Theta}_0 + \widehat{\Theta}_1 + \dots + \widehat{\Theta}_i$, on a alors les estimations suivantes, avec, entre parenthèses, les intervalles de confiance

i	$\widehat{\Theta}_i$	$\widehat{\Psi}_i$
1	$\begin{bmatrix} -0.320 & 0.146 & 0.961 \\ (0.125) & (0.562) & (0.657) \end{bmatrix}$	$\begin{bmatrix} 0.680 & 0.146 & 0.961 \\ (0.125) & (0.562) & (0.657) \end{bmatrix}$
	$\begin{bmatrix} 0.044 & -0.153 & 0.289 \\ (0.032) & (0.143) & (0.167) \end{bmatrix}$	$\begin{bmatrix} 0.044 & 0.847 & 0.289 \\ (0.032) & (0.143) & (0.167) \end{bmatrix}$
	$\begin{bmatrix} -0.002 & 0.225 & -0.264 \\ (0.025) & (0.115) & (0.134) \end{bmatrix}$	$\begin{bmatrix} -0.002 & 0.225 & 0.736 \\ (0.025) & (0.115) & (0.134) \end{bmatrix}$
	$\begin{bmatrix} -0.054 & 0.262 & 0.416 \\ (0.129) & (0.546) & (0.663) \end{bmatrix}$	$\begin{bmatrix} 0.626 & 0.408 & 1.377 \\ (0.148) & (0.651) & (0.755) \end{bmatrix}$
2	$\begin{bmatrix} 0.029 & 0.114 & -0.088 \\ (0.032) & (0.135) & (0.162) \end{bmatrix}$	$\begin{bmatrix} 0.073 & 0.961 & 0.200 \\ (0.043) & (0.192) & (0.222) \end{bmatrix}$
	$\begin{bmatrix} 0.045 & 0.261 & 0.110 \\ (0.026) & (0.108) & (0.131) \end{bmatrix}$	$\begin{bmatrix} 0.043 & 0.486 & 0.846 \\ (0.033) & (0.144) & (0.167) \end{bmatrix}$
	$\begin{bmatrix} 0.119 & 0.353 & -0.408 \\ (0.084) & (0.384) & (0.476) \end{bmatrix}$	$\begin{bmatrix} 0.745 & 0.761 & 0.969 \\ (0.099) & (0.483) & (0.550) \end{bmatrix}$
3	$\begin{bmatrix} -0.009 & 0.071 & 0.120 \\ (0.016) & (0.078) & (0.094) \end{bmatrix}$	$\begin{bmatrix} 0.064 & 1.033 & 0.320 \\ (0.037) & (0.176) & (0.203) \end{bmatrix}$
	$\begin{bmatrix} -0.001 & -0.098 & 0.091 \\ (0.017) & (0.078) & (0.102) \end{bmatrix}$	$\begin{bmatrix} 0.042 & 0.388 & 0.937 \\ (0.033) & (0.156) & (0.183) \end{bmatrix}$

avec les comportements limites suivants :

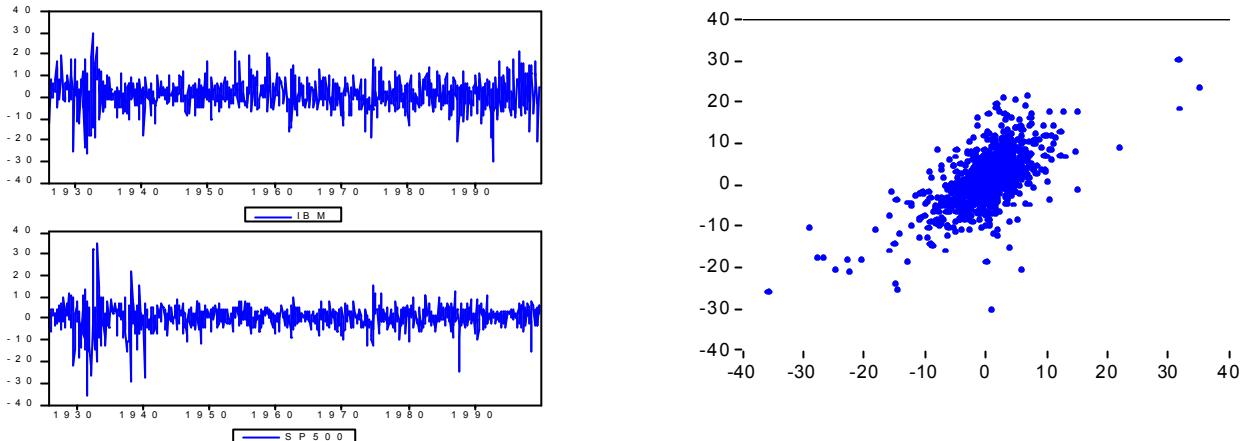
$$\lim_{i \rightarrow \infty} \widehat{\Theta}_i = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ et } \lim_{i \rightarrow \infty} \widehat{\Psi}_i = \begin{bmatrix} 0.756 & 0.866 & 1.295 \\ (0.133) & (0.661) & (0.798) \\ 0.076 & 1.076 & 0.344 \\ (0.048) & (0.236) & (0.285) \\ 0.053 & 0.505 & 0.964 \\ (0.043) & (0.213) & (0.257) \end{bmatrix}.$$

Ceci peut se représenter sur les graphiques ci-dessous, avec l'impact sur la consommation d'un choc de 1 sur le revenu à la date $t = 0$, avec l'impact aux différentes dates à gauche, et l'impact cumulé à droite,



1.4.2 Application des modèles VAR : rendements d'une action et d'un indice

Considérons les deux séries suivantes, le rendement de l'action *IBM*, et le rendement de l'indice *SP500* (*rendement mensuel, entre janvier 1926 et décembre 1999*)



avec les séries (X_t) et (Y_t) à gauche (*i.e. le rendement du titre IBM et de l'indice SP500 respectivement*), et le nuage des points (X_t, Y_t) . Une estimation VAR avec 5 retards donne les résultats suivants

Vector Autoregression Estimates																																																																																																																																																																					
Date: 06/04/03 Time: 14:28																																																																																																																																																																					
Sample(adjusted): 1926:06 1999:12																																																																																																																																																																					
Included observations: 883 after adjusting endpoints																																																																																																																																																																					
Standard errors in () & t-statistics in []																																																																																																																																																																					
<table border="1"> <thead> <tr> <th></th><th>IBM(-1)</th><th>IBM(-2)</th><th>IBM(-3)</th><th>IBM(-4)</th><th>IBM(-5)</th><th></th><th></th><th></th><th></th></tr> </thead> <tbody> <tr> <td>IBM</td><td>0.019012 (0.04354) [0.43664]</td><td>0.096939 (0.04344) [2.23143]</td><td>0.048639 (0.04411) [1.10266]</td><td>-0.010359 (0.04397) [-0.23559]</td><td>-0.073582 (0.04387) [-1.67740]</td><td></td><td></td><td></td><td></td></tr> <tr> <td>SP500</td><td>-0.012372 (0.03658) [-0.33819]</td><td>0.056943 (0.03650) [1.56002]</td><td>-0.001835 (0.03706) [-0.04951]</td><td>0.028003 (0.03695) [0.75794]</td><td>-0.075260 (0.03686) [-2.04191]</td><td></td><td></td><td></td><td></td></tr> <tr> <td colspan="10"><hr/></td></tr> <tr> <td> <table border="1"> <thead> <tr> <th></th><th>SP500(-1)</th><th>SP500(-2)</th><th>SP500(-3)</th><th>SP500(-4)</th><th>SP500(-5)</th><th></th><th></th><th></th><th></th></tr> </thead> <tbody> <tr> <td>IBM</td><td>0.105749 (0.05176) [2.04322]</td><td>-0.138224 (0.05174) [-2.67166]</td><td>-0.113612 (0.05222) [-2.17546]</td><td>-0.016755 (0.05238) [-0.31988]</td><td>0.131304 (0.05238) [2.50697]</td><td></td><td></td><td></td><td></td></tr> <tr> <td>SP500</td><td>0.088680 (0.04349) [2.03927]</td><td>-0.047391 (0.04347) [-1.09018]</td><td>-0.110083 (0.04388) [-2.50873]</td><td>0.017162 (0.04401) [0.38997]</td><td>0.132369 (0.04401) [3.00790]</td><td></td><td></td><td></td><td></td></tr> <tr> <td colspan="10"><hr/></td></tr> <tr> <td> <table border="1"> <thead> <tr> <th></th><th>C</th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></tr> </thead> <tbody> <tr> <td>IBM</td><td>1.167488 (0.24126) [4.83912]</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr> <td>SP500</td><td>0.509100 (0.20271) [2.51144]</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr> <td colspan="10"><hr/></td></tr> <tr> <td colspan="10"> <table border="1"> <thead> <tr> <th></th><th>R-squared</th><th>Adj. R-squared</th><th>Sum sq. resids</th><th>S.E. equation</th><th>F-statistic</th><th>Log likelihood</th><th>Akaike AIC</th><th>Schwarz SC</th><th>Mean dependent</th><th>S.D. dependent</th></tr> </thead> <tbody> <tr> <td>IBM</td><td>0.032500</td><td>0.021404</td><td>38600.50</td><td>6.653317</td><td>2.929166</td><td>-2920.775</td><td>6.640487</td><td>6.700076</td><td>1.251308</td><td>6.725687</td></tr> <tr> <td>SP500</td><td>0.033216</td><td>0.022129</td><td>27250.91</td><td>5.590263</td><td>2.995981</td><td>-2767.054</td><td>6.292309</td><td>6.351897</td><td>0.546273</td><td>5.653163</td></tr> </tbody> </table> </td></tr> </tbody> </table> </td></tr></tbody></table></td></tr></tbody></table>		IBM(-1)	IBM(-2)	IBM(-3)	IBM(-4)	IBM(-5)					IBM	0.019012 (0.04354) [0.43664]	0.096939 (0.04344) [2.23143]	0.048639 (0.04411) [1.10266]	-0.010359 (0.04397) [-0.23559]	-0.073582 (0.04387) [-1.67740]					SP500	-0.012372 (0.03658) [-0.33819]	0.056943 (0.03650) [1.56002]	-0.001835 (0.03706) [-0.04951]	0.028003 (0.03695) [0.75794]	-0.075260 (0.03686) [-2.04191]					<hr/>										<table border="1"> <thead> <tr> <th></th><th>SP500(-1)</th><th>SP500(-2)</th><th>SP500(-3)</th><th>SP500(-4)</th><th>SP500(-5)</th><th></th><th></th><th></th><th></th></tr> </thead> <tbody> <tr> <td>IBM</td><td>0.105749 (0.05176) [2.04322]</td><td>-0.138224 (0.05174) [-2.67166]</td><td>-0.113612 (0.05222) [-2.17546]</td><td>-0.016755 (0.05238) [-0.31988]</td><td>0.131304 (0.05238) [2.50697]</td><td></td><td></td><td></td><td></td></tr> <tr> <td>SP500</td><td>0.088680 (0.04349) [2.03927]</td><td>-0.047391 (0.04347) [-1.09018]</td><td>-0.110083 (0.04388) [-2.50873]</td><td>0.017162 (0.04401) [0.38997]</td><td>0.132369 (0.04401) [3.00790]</td><td></td><td></td><td></td><td></td></tr> <tr> <td colspan="10"><hr/></td></tr> <tr> <td> <table border="1"> <thead> <tr> <th></th><th>C</th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></tr> </thead> <tbody> <tr> <td>IBM</td><td>1.167488 (0.24126) [4.83912]</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr> <td>SP500</td><td>0.509100 (0.20271) [2.51144]</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr> <td colspan="10"><hr/></td></tr> <tr> <td colspan="10"> <table border="1"> <thead> <tr> <th></th><th>R-squared</th><th>Adj. R-squared</th><th>Sum sq. resids</th><th>S.E. equation</th><th>F-statistic</th><th>Log likelihood</th><th>Akaike AIC</th><th>Schwarz SC</th><th>Mean dependent</th><th>S.D. dependent</th></tr> </thead> <tbody> <tr> <td>IBM</td><td>0.032500</td><td>0.021404</td><td>38600.50</td><td>6.653317</td><td>2.929166</td><td>-2920.775</td><td>6.640487</td><td>6.700076</td><td>1.251308</td><td>6.725687</td></tr> <tr> <td>SP500</td><td>0.033216</td><td>0.022129</td><td>27250.91</td><td>5.590263</td><td>2.995981</td><td>-2767.054</td><td>6.292309</td><td>6.351897</td><td>0.546273</td><td>5.653163</td></tr> </tbody> </table> </td></tr> </tbody> </table> </td></tr></tbody></table>		SP500(-1)	SP500(-2)	SP500(-3)	SP500(-4)	SP500(-5)					IBM	0.105749 (0.05176) [2.04322]	-0.138224 (0.05174) [-2.67166]	-0.113612 (0.05222) [-2.17546]	-0.016755 (0.05238) [-0.31988]	0.131304 (0.05238) [2.50697]					SP500	0.088680 (0.04349) [2.03927]	-0.047391 (0.04347) [-1.09018]	-0.110083 (0.04388) [-2.50873]	0.017162 (0.04401) [0.38997]	0.132369 (0.04401) [3.00790]					<hr/>										<table border="1"> <thead> <tr> <th></th><th>C</th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></tr> </thead> <tbody> <tr> <td>IBM</td><td>1.167488 (0.24126) [4.83912]</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr> <td>SP500</td><td>0.509100 (0.20271) [2.51144]</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr> <td colspan="10"><hr/></td></tr> <tr> <td colspan="10"> <table border="1"> <thead> <tr> <th></th><th>R-squared</th><th>Adj. R-squared</th><th>Sum sq. resids</th><th>S.E. equation</th><th>F-statistic</th><th>Log likelihood</th><th>Akaike AIC</th><th>Schwarz SC</th><th>Mean dependent</th><th>S.D. dependent</th></tr> </thead> <tbody> <tr> <td>IBM</td><td>0.032500</td><td>0.021404</td><td>38600.50</td><td>6.653317</td><td>2.929166</td><td>-2920.775</td><td>6.640487</td><td>6.700076</td><td>1.251308</td><td>6.725687</td></tr> <tr> <td>SP500</td><td>0.033216</td><td>0.022129</td><td>27250.91</td><td>5.590263</td><td>2.995981</td><td>-2767.054</td><td>6.292309</td><td>6.351897</td><td>0.546273</td><td>5.653163</td></tr> </tbody> </table> </td></tr> </tbody> </table>		C									IBM	1.167488 (0.24126) [4.83912]									SP500	0.509100 (0.20271) [2.51144]									<hr/>										<table border="1"> <thead> <tr> <th></th><th>R-squared</th><th>Adj. R-squared</th><th>Sum sq. resids</th><th>S.E. equation</th><th>F-statistic</th><th>Log likelihood</th><th>Akaike AIC</th><th>Schwarz SC</th><th>Mean dependent</th><th>S.D. dependent</th></tr> </thead> <tbody> <tr> <td>IBM</td><td>0.032500</td><td>0.021404</td><td>38600.50</td><td>6.653317</td><td>2.929166</td><td>-2920.775</td><td>6.640487</td><td>6.700076</td><td>1.251308</td><td>6.725687</td></tr> <tr> <td>SP500</td><td>0.033216</td><td>0.022129</td><td>27250.91</td><td>5.590263</td><td>2.995981</td><td>-2767.054</td><td>6.292309</td><td>6.351897</td><td>0.546273</td><td>5.653163</td></tr> </tbody> </table>											R-squared	Adj. R-squared	Sum sq. resids	S.E. equation	F-statistic	Log likelihood	Akaike AIC	Schwarz SC	Mean dependent	S.D. dependent	IBM	0.032500	0.021404	38600.50	6.653317	2.929166	-2920.775	6.640487	6.700076	1.251308	6.725687	SP500	0.033216	0.022129	27250.91	5.590263	2.995981	-2767.054	6.292309	6.351897	0.546273	5.653163
	IBM(-1)	IBM(-2)	IBM(-3)	IBM(-4)	IBM(-5)																																																																																																																																																																
IBM	0.019012 (0.04354) [0.43664]	0.096939 (0.04344) [2.23143]	0.048639 (0.04411) [1.10266]	-0.010359 (0.04397) [-0.23559]	-0.073582 (0.04387) [-1.67740]																																																																																																																																																																
SP500	-0.012372 (0.03658) [-0.33819]	0.056943 (0.03650) [1.56002]	-0.001835 (0.03706) [-0.04951]	0.028003 (0.03695) [0.75794]	-0.075260 (0.03686) [-2.04191]																																																																																																																																																																
<hr/>																																																																																																																																																																					
<table border="1"> <thead> <tr> <th></th><th>SP500(-1)</th><th>SP500(-2)</th><th>SP500(-3)</th><th>SP500(-4)</th><th>SP500(-5)</th><th></th><th></th><th></th><th></th></tr> </thead> <tbody> <tr> <td>IBM</td><td>0.105749 (0.05176) [2.04322]</td><td>-0.138224 (0.05174) [-2.67166]</td><td>-0.113612 (0.05222) [-2.17546]</td><td>-0.016755 (0.05238) [-0.31988]</td><td>0.131304 (0.05238) [2.50697]</td><td></td><td></td><td></td><td></td></tr> <tr> <td>SP500</td><td>0.088680 (0.04349) [2.03927]</td><td>-0.047391 (0.04347) [-1.09018]</td><td>-0.110083 (0.04388) [-2.50873]</td><td>0.017162 (0.04401) [0.38997]</td><td>0.132369 (0.04401) [3.00790]</td><td></td><td></td><td></td><td></td></tr> <tr> <td colspan="10"><hr/></td></tr> <tr> <td> <table border="1"> <thead> <tr> <th></th><th>C</th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></tr> </thead> <tbody> <tr> <td>IBM</td><td>1.167488 (0.24126) [4.83912]</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr> <td>SP500</td><td>0.509100 (0.20271) [2.51144]</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr> <td colspan="10"><hr/></td></tr> <tr> <td colspan="10"> <table border="1"> <thead> <tr> <th></th><th>R-squared</th><th>Adj. R-squared</th><th>Sum sq. resids</th><th>S.E. equation</th><th>F-statistic</th><th>Log likelihood</th><th>Akaike AIC</th><th>Schwarz SC</th><th>Mean dependent</th><th>S.D. dependent</th></tr> </thead> <tbody> <tr> <td>IBM</td><td>0.032500</td><td>0.021404</td><td>38600.50</td><td>6.653317</td><td>2.929166</td><td>-2920.775</td><td>6.640487</td><td>6.700076</td><td>1.251308</td><td>6.725687</td></tr> <tr> <td>SP500</td><td>0.033216</td><td>0.022129</td><td>27250.91</td><td>5.590263</td><td>2.995981</td><td>-2767.054</td><td>6.292309</td><td>6.351897</td><td>0.546273</td><td>5.653163</td></tr> </tbody> </table> </td></tr> </tbody> </table> </td></tr></tbody></table>		SP500(-1)	SP500(-2)	SP500(-3)	SP500(-4)	SP500(-5)					IBM	0.105749 (0.05176) [2.04322]	-0.138224 (0.05174) [-2.67166]	-0.113612 (0.05222) [-2.17546]	-0.016755 (0.05238) [-0.31988]	0.131304 (0.05238) [2.50697]					SP500	0.088680 (0.04349) [2.03927]	-0.047391 (0.04347) [-1.09018]	-0.110083 (0.04388) [-2.50873]	0.017162 (0.04401) [0.38997]	0.132369 (0.04401) [3.00790]					<hr/>										<table border="1"> <thead> <tr> <th></th><th>C</th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></tr> </thead> <tbody> <tr> <td>IBM</td><td>1.167488 (0.24126) [4.83912]</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr> <td>SP500</td><td>0.509100 (0.20271) [2.51144]</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr> <td colspan="10"><hr/></td></tr> <tr> <td colspan="10"> <table border="1"> <thead> <tr> <th></th><th>R-squared</th><th>Adj. R-squared</th><th>Sum sq. resids</th><th>S.E. equation</th><th>F-statistic</th><th>Log likelihood</th><th>Akaike AIC</th><th>Schwarz SC</th><th>Mean dependent</th><th>S.D. dependent</th></tr> </thead> <tbody> <tr> <td>IBM</td><td>0.032500</td><td>0.021404</td><td>38600.50</td><td>6.653317</td><td>2.929166</td><td>-2920.775</td><td>6.640487</td><td>6.700076</td><td>1.251308</td><td>6.725687</td></tr> <tr> <td>SP500</td><td>0.033216</td><td>0.022129</td><td>27250.91</td><td>5.590263</td><td>2.995981</td><td>-2767.054</td><td>6.292309</td><td>6.351897</td><td>0.546273</td><td>5.653163</td></tr> </tbody> </table> </td></tr> </tbody> </table>		C									IBM	1.167488 (0.24126) [4.83912]									SP500	0.509100 (0.20271) [2.51144]									<hr/>										<table border="1"> <thead> <tr> <th></th><th>R-squared</th><th>Adj. R-squared</th><th>Sum sq. resids</th><th>S.E. equation</th><th>F-statistic</th><th>Log likelihood</th><th>Akaike AIC</th><th>Schwarz SC</th><th>Mean dependent</th><th>S.D. dependent</th></tr> </thead> <tbody> <tr> <td>IBM</td><td>0.032500</td><td>0.021404</td><td>38600.50</td><td>6.653317</td><td>2.929166</td><td>-2920.775</td><td>6.640487</td><td>6.700076</td><td>1.251308</td><td>6.725687</td></tr> <tr> <td>SP500</td><td>0.033216</td><td>0.022129</td><td>27250.91</td><td>5.590263</td><td>2.995981</td><td>-2767.054</td><td>6.292309</td><td>6.351897</td><td>0.546273</td><td>5.653163</td></tr> </tbody> </table>											R-squared	Adj. R-squared	Sum sq. resids	S.E. equation	F-statistic	Log likelihood	Akaike AIC	Schwarz SC	Mean dependent	S.D. dependent	IBM	0.032500	0.021404	38600.50	6.653317	2.929166	-2920.775	6.640487	6.700076	1.251308	6.725687	SP500	0.033216	0.022129	27250.91	5.590263	2.995981	-2767.054	6.292309	6.351897	0.546273	5.653163																																									
	SP500(-1)	SP500(-2)	SP500(-3)	SP500(-4)	SP500(-5)																																																																																																																																																																
IBM	0.105749 (0.05176) [2.04322]	-0.138224 (0.05174) [-2.67166]	-0.113612 (0.05222) [-2.17546]	-0.016755 (0.05238) [-0.31988]	0.131304 (0.05238) [2.50697]																																																																																																																																																																
SP500	0.088680 (0.04349) [2.03927]	-0.047391 (0.04347) [-1.09018]	-0.110083 (0.04388) [-2.50873]	0.017162 (0.04401) [0.38997]	0.132369 (0.04401) [3.00790]																																																																																																																																																																
<hr/>																																																																																																																																																																					
<table border="1"> <thead> <tr> <th></th><th>C</th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></tr> </thead> <tbody> <tr> <td>IBM</td><td>1.167488 (0.24126) [4.83912]</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr> <td>SP500</td><td>0.509100 (0.20271) [2.51144]</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr> <td colspan="10"><hr/></td></tr> <tr> <td colspan="10"> <table border="1"> <thead> <tr> <th></th><th>R-squared</th><th>Adj. R-squared</th><th>Sum sq. resids</th><th>S.E. equation</th><th>F-statistic</th><th>Log likelihood</th><th>Akaike AIC</th><th>Schwarz SC</th><th>Mean dependent</th><th>S.D. dependent</th></tr> </thead> <tbody> <tr> <td>IBM</td><td>0.032500</td><td>0.021404</td><td>38600.50</td><td>6.653317</td><td>2.929166</td><td>-2920.775</td><td>6.640487</td><td>6.700076</td><td>1.251308</td><td>6.725687</td></tr> <tr> <td>SP500</td><td>0.033216</td><td>0.022129</td><td>27250.91</td><td>5.590263</td><td>2.995981</td><td>-2767.054</td><td>6.292309</td><td>6.351897</td><td>0.546273</td><td>5.653163</td></tr> </tbody> </table> </td></tr> </tbody> </table>		C									IBM	1.167488 (0.24126) [4.83912]									SP500	0.509100 (0.20271) [2.51144]									<hr/>										<table border="1"> <thead> <tr> <th></th><th>R-squared</th><th>Adj. R-squared</th><th>Sum sq. resids</th><th>S.E. equation</th><th>F-statistic</th><th>Log likelihood</th><th>Akaike AIC</th><th>Schwarz SC</th><th>Mean dependent</th><th>S.D. dependent</th></tr> </thead> <tbody> <tr> <td>IBM</td><td>0.032500</td><td>0.021404</td><td>38600.50</td><td>6.653317</td><td>2.929166</td><td>-2920.775</td><td>6.640487</td><td>6.700076</td><td>1.251308</td><td>6.725687</td></tr> <tr> <td>SP500</td><td>0.033216</td><td>0.022129</td><td>27250.91</td><td>5.590263</td><td>2.995981</td><td>-2767.054</td><td>6.292309</td><td>6.351897</td><td>0.546273</td><td>5.653163</td></tr> </tbody> </table>											R-squared	Adj. R-squared	Sum sq. resids	S.E. equation	F-statistic	Log likelihood	Akaike AIC	Schwarz SC	Mean dependent	S.D. dependent	IBM	0.032500	0.021404	38600.50	6.653317	2.929166	-2920.775	6.640487	6.700076	1.251308	6.725687	SP500	0.033216	0.022129	27250.91	5.590263	2.995981	-2767.054	6.292309	6.351897	0.546273	5.653163																																																																																		
	C																																																																																																																																																																				
IBM	1.167488 (0.24126) [4.83912]																																																																																																																																																																				
SP500	0.509100 (0.20271) [2.51144]																																																																																																																																																																				
<hr/>																																																																																																																																																																					
<table border="1"> <thead> <tr> <th></th><th>R-squared</th><th>Adj. R-squared</th><th>Sum sq. resids</th><th>S.E. equation</th><th>F-statistic</th><th>Log likelihood</th><th>Akaike AIC</th><th>Schwarz SC</th><th>Mean dependent</th><th>S.D. dependent</th></tr> </thead> <tbody> <tr> <td>IBM</td><td>0.032500</td><td>0.021404</td><td>38600.50</td><td>6.653317</td><td>2.929166</td><td>-2920.775</td><td>6.640487</td><td>6.700076</td><td>1.251308</td><td>6.725687</td></tr> <tr> <td>SP500</td><td>0.033216</td><td>0.022129</td><td>27250.91</td><td>5.590263</td><td>2.995981</td><td>-2767.054</td><td>6.292309</td><td>6.351897</td><td>0.546273</td><td>5.653163</td></tr> </tbody> </table>											R-squared	Adj. R-squared	Sum sq. resids	S.E. equation	F-statistic	Log likelihood	Akaike AIC	Schwarz SC	Mean dependent	S.D. dependent	IBM	0.032500	0.021404	38600.50	6.653317	2.929166	-2920.775	6.640487	6.700076	1.251308	6.725687	SP500	0.033216	0.022129	27250.91	5.590263	2.995981	-2767.054	6.292309	6.351897	0.546273	5.653163																																																																																																																											
	R-squared	Adj. R-squared	Sum sq. resids	S.E. equation	F-statistic	Log likelihood	Akaike AIC	Schwarz SC	Mean dependent	S.D. dependent																																																																																																																																																											
IBM	0.032500	0.021404	38600.50	6.653317	2.929166	-2920.775	6.640487	6.700076	1.251308	6.725687																																																																																																																																																											
SP500	0.033216	0.022129	27250.91	5.590263	2.995981	-2767.054	6.292309	6.351897	0.546273	5.653163																																																																																																																																																											

c'est à dire que, dans une écriture où l'on noterait $Z_t = (X_t, Y_t)'$,

$$Z_t = \Phi_0 + \Phi_1 Z_{t-1} + \dots + \Phi_5 X_{t-5} + \varepsilon_t,$$

on aurait alors les matrices suivantes,

$$\Phi_0 = \begin{bmatrix} 1.167 \\ 0.509 \end{bmatrix}, \quad \Phi_1 = \begin{bmatrix} 0.019 & 0.105 \\ -0.012 & 0.088 \end{bmatrix}, \quad \Phi_2 = \begin{bmatrix} 0.097 & -0.138 \\ 0.057 & -0.047 \end{bmatrix},$$

$$\Phi_3 = \begin{bmatrix} 0.049 & -0.114 \\ -0.002 & -0.110 \end{bmatrix}, \quad \Phi_4 = \begin{bmatrix} -0.010 & -0.017 \\ 0.028 & 0.017 \end{bmatrix}, \quad \text{et } \Phi_5 = \begin{bmatrix} -0.073 & 0.131 \\ -0.075 & 0.132 \end{bmatrix}.$$

Le schéma ci-dessous indique quelles sont les valeurs significatives, ainsi que leur signe (+ : *significative positive*, - : *significative négative* et . : *non significative*)

$$\Phi_0 = \begin{bmatrix} + \\ + \end{bmatrix}, \quad \Phi_1 = \begin{bmatrix} \cdot & + \\ \cdot & + \end{bmatrix}, \quad \Phi_2 = \begin{bmatrix} + & - \\ \cdot & \cdot \end{bmatrix}, \quad \Phi_3 = \begin{bmatrix} \cdot & - \\ \cdot & - \end{bmatrix}, \quad \Phi_4 = \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}, \quad \text{et } \Phi_5 = \begin{bmatrix} \cdot & + \\ - & + \end{bmatrix}.$$

C'est à dire que les retards à l'ordre 4 ne sont pas significatifs par exemple.

En fait, ce cas est encore plus particulier car un certain nombre de colonnes semblent non-significatives, en particulier la première colonne de Φ_1 ou de Φ_3 (et dans une moindre mesure de Φ_5). Cela signifie que les retards de X_t

n'influencent ni X_t , ni Y_t : en terme d'interprétations, le rendement du cours du titre *IBM* à la date t , ainsi que le cours de l'indice *SP500* à cette même date dépendent des cours passés de l'indice (*en l'occurrence avec 1, 3 et 5 retards*), mais pas vraiment du cours passé du titre *IBM*. Autrement dit, on pourrait être intéressé de tester un modèle particulier, de la forme

$$Z_t = \Phi_0 + \Phi_1 Y_{t-1} + \Phi_3 Y_{t-3} \text{ où ici } \Phi_i \in \mathbb{R}^2,$$

(et Φ_i n'est plus une matrice 2×2). L'estimation de ce modèle contraint donne alors

$$\begin{bmatrix} IBM_t \\ SP500_t \end{bmatrix} = \begin{bmatrix} 1.24 \\ 0.57 \end{bmatrix} + \begin{bmatrix} 0.117 \\ 0.073 \end{bmatrix} SP500_{t-1} - \begin{bmatrix} 0.083 \\ 0.109 \end{bmatrix} SP500_{t-3} + \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}.$$

1.5 Régression linéaire dans un cadre dynamique

Nous allons ici voir les propriétés des modèles économétriques de la forme $Y = X\beta + \varepsilon$, où, contrairement aux modèles économétriques, les observations (*de la variable à expliquer* Y) $Y_1, Y_2, \dots, Y_i, \dots, Y_n$ ne sont pas observées pour des individus différents (*et peuvent alors être supposées "indépendantes"*) mais sont des observations temporelles $Y_1, Y_2, \dots, Y_t, \dots, Y_n$. Nous supposerons ici la fonction de régression dépend des valeurs retardées d'une ou de plusieurs variables dépendantes.

1.5.1 Les retards échelonnés

On pense souvent qu'une variable dépendante (Y_t) dépend de la valeur actuelle et d'un certain nombre de valeurs retardées d'une variable indépendante (X_t). Une modélisation de ce genre consiste à utiliser un modèle à retards échelonnés tel que

$$Y_t = \alpha + \sum_{j=0}^q \beta_j X_{t-j} + \varepsilon_t \text{ où } (\varepsilon_t) \sim BB(0, \sigma^2). \quad (9)$$

Le problème consiste alors à estimer la constante α et les coefficients β_j . Le nombre entier q est ici la longueur du dernier retard. Dans certains cas, imaginer que q est infini peut avoir du sens, mais nous supposerons pour l'instant qu'il prend une valeur finie. La fonction de régression pourrait tout à fait dépendre d'autres variables explicatives, mais nous ne considérerons ici que le cas où intervient une seule variable explicative pour conserver des notations simples.

Le problème avec un tel modèle est que, parce que X_t sera - a priori - fortement corrélé à X_{t-1}, X_{t-2}, \dots etc, et donc les estimations par moindres carrés des coefficients β_j devraient être assez imprécises. Sauf si c'est l'impact à long terme qui nous intéresse. Soit $V(\hat{\beta})$ la matrice de covarinace du vecteur $\hat{\beta}$ estimateur *mco* du vecteur $\beta = (\beta_0, \beta_1, \dots, \beta_q)$. Alors, si $\hat{\gamma}$ désigne la somme des $\hat{\beta}_j$, la variance de $\hat{\gamma}$

$$V(\hat{\gamma}) = \sum_{j=0}^q V(\hat{\beta}_j) + 2 \sum_{j=0}^q \sum_{i=0}^{j-1} cov(\hat{\beta}_i, \hat{\beta}_j).$$

Si X_{t-j} est corrélé positivement à X_{t-k} pour tout $j \neq k$, les termes de covariance dans l'expression ci-dessus seront généralement négatifs. Lorsqu'ils sont importants et négatifs, comme c'est souvent le cas, $V(\hat{\gamma})$ peut être plus petite que la somme des $V(\hat{\beta}_j)$ ou même que chaque $V(\hat{\beta}_j)$. Si c'est le paramètre γ qui nous intéresse plutôt que les β_j individuels, l'approche la plus simple consiste à estimer une version reparamétrisée de (9) par moindres carrés. Cette version s'écrit alors

$$Y_t = \alpha + \gamma X_t + \sum_{j=1}^q \beta_j (X_{t-j} - X_t) + \varepsilon_t \text{ où } \gamma = \beta_0 + \beta_1 + \dots + \beta_q.$$

L'avantage de cette écriture est que l'écart type de $\hat{\gamma}$ est disponible immédiatement dans les résultats de la régression.

La colinéarité peut être un problème important si l'on s'intéresse aux β_j . L'approche la plus connue pour résoudre ce problème consiste à employer les **retards échelonnés polynomiaux**, ou *PDL*. Ces derniers sont quelquefois appelés retards d'Almon, suite à l'article d'Almon (1965).

Dans un polynôme de retards échelonnés, les coefficients β_j de (9) doivent apparaître dans un polynôme de degré d donné. Par, si le polynôme était du second degré, nous aurions

$$\beta_j = P(j) = \alpha_0 + \alpha_1 j + \alpha_2 j^2 \text{ pour tout } j = 0, \dots, q. \quad (10)$$

Si $q > 2$, on peut noter qu'il y aura moins de paramètres α_i que β_j et, par conséquent, cette relation impose $q - 2$ contraintes sur les β_j . L'estimation d'un modèle soumis à des contraintes imposées par un *PDL* est assez immédiate.

Par exemple, pour estimer (9) soumis à (10), nous pouvons remplacer simplement les β_j par $\alpha_0 + \alpha_1 j + \alpha_2 j^2$, c'est à dire

$$Y_t = \alpha + \alpha_0 \sum_{j=0}^q X_{t-j} + \alpha_1 \sum_{j=0}^q jX_{t-j} + \alpha_2 \sum_{j=0}^q j^2 X_{t-j} + \varepsilon_t. \quad (11)$$

Ceci est un exemple de modèle $PDL(q, 2)$. Un modèle $PDL(q, d)$ doit toujours être tel que $d < q$. En fait, les contraintes imposées aux β_j sont simplement des contraintes linéaires. La résolution de (10) montre que

$$\begin{aligned} -\beta_3 + 3\beta_2 - 3\beta_1 + \beta_0 &= 0 \\ -\beta_4 + 3\beta_3 - 3\beta_2 + \beta_1 &= 0 \\ -\beta_5 + 3\beta_4 - 3\beta_3 + \beta_2 &= 0 \\ \dots \text{ et ainsi de suite.} \end{aligned}$$

On peut réécrire ces contraintes sous la forme $R\beta = 0$ où la matrice R serait dans ce cas

$$R = \left[\begin{array}{ccccccccc} 1 & -3 & 3 & -1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 1 & -3 & 3 & -1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -3 & 3 & -1 & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & -3 & 3 & -1 \end{array} \right].$$

Puisque les contraintes sont linéaires, on peut les tester facilement, par exemple à l'aide d'un test de Fisher : le modèle constraint est (11), le modèle non constraint est (9), et le nombre de contraintes dans ce cas est $q - 2$ (*de façon plus générale, pour un modèle $PDL(q, d)$, il y aura $q - j$ contraintes*).

Ainsi, on peut tester un modèle $PDL(q, d)$ contre un modèle $PDL(q, d + 1)$ en utilisant un test de Fisher. La meilleure approche consiste à débuter par une valeur importante de q et en examinant la détérioration de la qualité de l'ajustement du modèle en diminuant sa valeur, sans imposer aucun contrainte sur la forme des retards échelonnés. Une fois que q est déterminé, on peut ensuite déterminer d , en débutant avec une valeur importante et en la réduisant au fur et à mesure. Sargan (1980) a proposé un exemple empirique dans *Some tests of dynamic specification for a single equation*.

Une variante intéressante a été proposée par Shiller (1973) : comme nous l'avons vu, les contraintes imposées par un PDL peuvent toujours s'écrire sous la forme $R\beta = 0$ pour une matrice R de dimension $r \times k$ où $r = q - d$ et k est le nombre d'éléments de β (*généralement supérieur à $q + 1$*). Shiller suggéra qu'au lieu de stipuler que chaque ligne de $R\beta$ soit nulle, il proposa qu'elle soit égale à une variable aléatoire d'espérance nulle et de variance finie. L'un des avantages de cette approche est que d peut être très faible sans pour cela imposer des contraintes trop fortes sur les données. Puisque les estimations n'ont pas besoin de se conformer exactement à la forme du polynôme, $d = 2$ est, dans la pratique, souvent suffisante. Ce genre de contrainte est appelé **contrainte stochastique**, parce qu'elle n'est pas sensées être vérifiée exactement. Les contraintes stochastiques sont très différentes de n'importe quel autre type de contraintes dont nous avons discuté. C'est ainsi que la phase d'estimation est relativement plus compliquée que dans le cas précédent, et repose le plus souvent sur des résultats de statistique bayésienne. Des compléments sur ce point peuvent être trouvés dans Davison et MacKinnon.

1.5.2 Le modèle d'ajustement partiel

Un modèle dynamique simple, et très fréquent dans la littérature, est le **modèle d'ajustement partiel**, dont l'histoire en économie remonte assez loin puisqu'il date de Nerlove (1958). Dans le cadre d'un modèle de politique économique, supposons que le niveau désiré d'une variable économique (Y_t) quelconque, soit (Y_t^*) , qui est supposé être relié à un vecteur de variables explicatives exogènes (X_t) par la relation

$$Y_t^* = X_t \beta^* + \varepsilon_t = (X_t^1, \dots, X_t^d) \begin{pmatrix} \beta_1^* \\ \vdots \\ \beta_d^* \end{pmatrix} + \varepsilon_t$$

Supposons que la variable (Y_t) s'ajuste vers (Y_t^*) , la valeur réalisée, à l'aide d'une relation de la forme

$$Y_t - Y_{t-1} = (1 - \delta)(Y_t^* - Y_{t-1}) + \eta_t.$$

Ces deux équations permettent d'obtenir la relation suivante

$$\begin{aligned} Y_t &= Y_{t-1} - (1 - \delta)Y_{t-1} + (1 - \delta)X_t \beta^* + (1 - \delta)\varepsilon_t + \eta_t \\ &= X_t \beta + \delta Y_{t-1} + u_t, \end{aligned}$$

où $\beta = (1 - \delta)\beta^*$ et $u_t = (1 - \delta)\varepsilon_t + \eta_t$. Cet ajustement n'est toutefois pertinent que si $0 < \delta < 1$ et si δ n'est pas trop proche de 1 (*sinon la vitesse d'ajustement est trop faible*) De façon itérative, on peut montrer que ce modèle se réécrit

$$Y_t = \sum_{j=0}^{+\infty} \delta^j [X_{t-j}\beta + u_{t-j}] = \left[\sum_{j=0}^{+\infty} \delta^j X_{t-j} \right] \beta + \left[\sum_{j=0}^{+\infty} \delta^j u_{t-j} \right]. \quad (12)$$

Ainsi ce modèle corrige une défaillance majeure que nous avions déjà remarquée dans les modèles à retards échelonnés: (Y_t) dépend maintenant autant des valeurs retardées de l'aléa (ε_t) que des valeurs retardées des variables exogènes (X_t). Notons que la solution de (12) repose sur l'hypothèse que $|\delta| < 1$, qui est une condition de stationnarité pour ce modèle.

Le modèle d'ajustement partiel n'est qu'un des nombreux modèles économiques que l'on peut utiliser pour justifier la prise en compte d'un ou de plusieurs retards des variables dépendantes dans la fonction de régression. Dhrymes (1971) et Hendry, Pagan, et Sargan (1984) discutent de nombreux autres modèles. Si nous ne discuterons pas ici ces modèles, nous nous concentrerons ici sur quelques résultats d'ordre général qui peuvent survenir lorsque l'on tente de spécifier et d'estimer des modèles de régression dynamiques.

Un problème qui se manifeste chaque fois que la matrice X contient des variables dépendantes retardées est que l'estimation *mco* par moindres carrés ne produisent pas des estimations sans biais. Ce problème survient parce que X est une matrice stochastique, dont certains éléments sont corrélés à quelques éléments de u . Ainsi on peut en particulier noter que

$$\mathbb{E} \left((X'X)^{-1} X'u \right) \neq (X'X)^{-1} X' \cdot \mathbb{E}(u).$$

Ceci peut se noter sur l'exemple très simple des modèles $AR(p)$: supposons que (Y_t) soit un processus $AR(1)$

$$Y_t = \beta Y_{t-1} + u_t \text{ où } |\beta| < 1 \text{ et } (u_t).$$

L'estimation *mco* de β est donné par

$$\hat{\beta} = \frac{\sum_{t=2}^n Y_t Y_{t-1}}{\sum_{t=2}^n Y_{t-1}^2} = \frac{\beta \sum_{t=2}^n Y_{t-1}^2 + \sum_{t=2}^n u_t Y_{t-1}}{\sum_{t=2}^n Y_{t-1}^2} = \beta + \frac{\sum_{t=2}^n u_t Y_{t-1}}{\sum_{t=2}^n Y_{t-1}^2}. \quad (13)$$

Le second terme dans l'expression la plus à droite de (13) n'est pas d'espérance nulle, parce que le numérateur et le dénominateur ne sont pas indépendants. Son espérance est d'ailleurs assez difficile à déterminer. La conclusion est alors la suivante

A retenir 1 *Dans tous les modèles pour lesquels il y a des variables dépendantes retardées, l'estimateur mco est biaisé*

Par contre, évidemment, l'estimateur $\hat{\beta}$ obtenu par *mco* est convergent. Si l'on divise à la fois le numérateur et le dénominateur du terme aléatoire du membre le plus à droite de (13) par n et si l'on prend les limites en probabilité, on obtient

$$\underset{n \rightarrow \infty}{\text{plim}} \hat{\beta} = \beta + \frac{\underset{n \rightarrow \infty}{\text{plim}} \sum_{t=2}^n u_t Y_{t-1}}{\underset{n \rightarrow \infty}{\text{plim}} \sum_{t=2}^n Y_{t-1}^2} = \beta.$$

1.5.3 Les modèles autorégressifs à retards échelonnés ($ADL(p,q)$)

Ces modèles sont présentés ici dans le cas simple où (Y_t) est la variable à expliquer et (X_t) la variable explicative (*ce type de modèles peut toutefois se généraliser à un ensemble plus vaste de variables explicatives*). La forme générale s'écrit

$$Y_t = \alpha + \sum_{i=1}^p \beta_i Y_{t-i} + \sum_{j=1}^q \gamma_j X_{t-j} + \varepsilon_t \text{ où } \varepsilon_t \sim BB(0, \sigma^2),$$

ou bien, en utilisant des polynômes d'opérateurs retard, $B(L)Y_t = \alpha + \Gamma(L)X_t + \varepsilon_t$. Un cas particulier, mais relativement utilisé est le modèle $ADL(1,1)$

$$Y_t = \alpha + \beta_1 Y_{t-1} + \gamma_0 X_t + \gamma_1 X_{t-1} + \varepsilon_t. \quad (14)$$

Nous étudierons ici le modèle $ADL(1,1)$, la plupart des résultats étant aisément généralisables aux modèles $ADL(p,q)$. Le cas où $\beta_1 = \gamma_1 = 0$ correspond au modèle général de régression linéaire (*statique, tel qu'il est étudié en cours d'économétrie*), le cas $\gamma_0 = \gamma_1 = 0$ correspond au modèle $AR(1)$ tel qu'il est étudié au début de ce cours, le cas où $\beta_1 = 1$ et $\gamma_1 = -\gamma_0$ correspond au modèle en différences premières...etc. Le modèle $ADL(1,1)$ fournit ainsi une alternative naturelle contre laquelle on peut tester n'importe lequel de ces cas particuliers.

Nous allons étudier ici la façon dont (X_t) influence (Y_t) à long terme, dans un modèle *ADL* (1, 1). Sans l'aléa (ε_t), (X_t) et (Y_t) convergeraient vers des valeurs de long terme (*stables*) X^* et Y^* telles que $Y^* = \alpha + \beta_1 Y^* + \gamma_0 X^* + \gamma_1 X^*$, c'est à dire que

$$Y^* = \frac{\alpha}{1 - \beta_1} + \frac{\gamma_0 + \gamma_1}{1 - \beta_1} X^* = \frac{\alpha}{1 - \beta_1} + \lambda X^*,$$

où λ peut alors être vue comme une élasticité. Toutefois, ce résultat n'est pertinent que si $|\beta_1| < 1$, ce qui est une condition de stabilité pour ce modèle.

Une des caractéristiques intéressantes et importantes des modèles *ADL* est que l'on peut les écrire de différentes façons : par exemple, (14) peut être écrit sous les formes suivantes (*toutes équivalentes*) :

- (1) $\Delta Y_t = \alpha + (\beta - 1) Y_{t-1} + \gamma_0 X_t + \gamma_1 X_{t-1} + \varepsilon_t,$
- (2) $\Delta Y_t = \alpha + (\beta - 1) Y_{t-1} + \gamma_0 \Delta X_t + (\gamma_0 + \gamma_1) X_{t-1} + \varepsilon_t,$
- (3) $\Delta Y_t = \alpha + (\beta - 1) Y_{t-1} - \gamma_1 \Delta X_t + (\gamma_0 + \gamma_1) X_t + \varepsilon_t,$
- (4) $\Delta Y_t = \alpha + (\beta - 1) (Y_{t-1} - X_{t-1}) + \gamma_0 \Delta X_t + (\gamma_0 + \gamma_1 + \beta_1 - 1) X_{t-1} + \varepsilon_t,$
- (5) $\Delta Y_t = \alpha + (\beta - 1) (Y_{t-1} - \lambda X_{t-1}) + \gamma_0 \Delta X_t + \varepsilon_t,$

où Δ est l'opérateur des différences premières, $\Delta = (1 - L)$. Si l'on s'intéresse à la somme des γ_i , les estimations et les écarts types s'obtiennent directement à partir de l'estimation par *mco* de (2) ou (3), et si l'on s'intéresse à λ , elles peuvent être obtenues par une estimation par moindres carrés non-linéaires de (5). La plus intéressante de ces spécifications est sans doute (5), dans laquelle le modèle est écrit sous la forme que l'on appelle **forme à correction d'erreur**. Le paramètre λ apparaît directement dans cette forme du modèle. Et bien que la forme à correction d'erreur soit non linéaire, l'estimation peut se faire facilement parce que le modèle est simplement un modèle linéaire soumis à une contrainte non linéaire. La différence entre Y_{t-1} et λX_{t-1} mesure l'importance de la défaillance de la relation d'équilibre de long terme entre (X_t) et (Y_t) . On appelle alors souvent le terme $(\beta - 1)(Y_{t-1} - \lambda X_{t-1})$ qui apparaît dans (5) terme de correction d'erreur, et le modèle est alors appelé **modèle à correction d'erreur**, ou *ECM*.

Ce qu'il convient de retenir sur cette partie, c'est que lorsque l'on tente de spécifier des modèles de régression dynamiques, il existe, en général, plusieurs manières a priori de le faire.

1.6 Compléments : modèles multivariés sous *SAS*

Il est possible de modéliser des séries temporelles sous *SAS* à l'aide de la procédure **VARMAX** (*Vector AutoRegressive and Moving-Average processes with eXogenous regressors*). La synthaxe de cette procédure est la suivante,

```
PROC VARMAX options;
  BY variables ;
  CAUSAL group1 = (variables) group2 = (variables) ;
  COINTEG rank = number < options > ;
  ID variable interval= value < option > ;
  MODEL dependent variables < = regressors >;
  OUTPUT < options >;
  RESTRICT restrictions ;
  TEST restrictions ;
```

Les ordres autorégressifs et moyenne-mobile sont déterminés ici, en utilisant les critères usuels (*AIC*, *HQ* ou *encore SBC*). Cette procédure propose également, dans le cas de données non-stationnaires, un certain nombre de tests : des tests de racine unité (*ADF*), de cointégration (*Johansen*), des tests de causalité (*Granger*). Enfin, un grand nombre d'options sont possibles (*erreurs ARCH*, *composantes saisonnières*, *introduction de variables exogènes*)

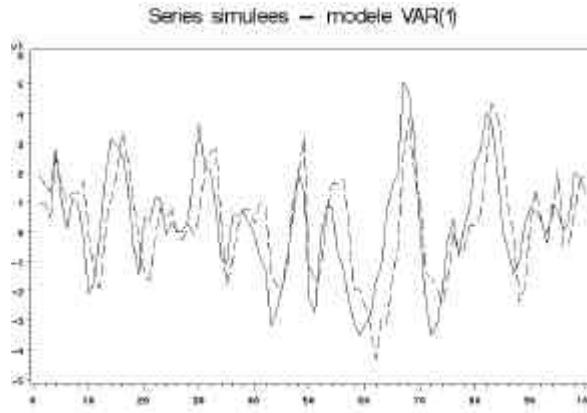
Les tests de cointégration, par exemple, reprennent la même forme que sous Eviews, i.e. pour un modèle de la forme (5), à savoir sans variable exogène.

1.6.1 Estimation d'un modèle *VAR*

Considérons ici le modèle *VAR* (1) simulé suivant, $Z_t = AZ_{t-1} + u_t$ où

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 1.2 & -0.5 \\ 0.6 & 0.3 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \text{ où } \Sigma = V(u_t) = \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.25 \end{pmatrix}$$

que nous simulons sur 100 observations,



Le programme permettant de tester les paramètres du modèle *VAR* sont les suivants, où les séries sont X et Y ,

```
PROC VARMAX data=donnees;
  MODEL X Y / p=1 noint lagmax=3;
  run ;
```

Parmi les sorties, on notera le deux sorties présentées ci-dessous

The VARMAX Procedure							Covariance Matrix for the Innovation		
Variable	Type	NoMissN	Mean	StdDev	Min	Max	Variable	y1	y2
y1	DEP	100	-0.17054	2.05299	-4.80463	3.94386	y1	0.85478	0.60495
y2	DEP	100	-0.15409	2.02123	-4.63446	3.69658	y2	0.60495	1.28770
Information Criteria									
Type of Model VAR(1)									
Estimation Method Least Squares Estimation									
AR Coefficient Estimates									
Lag	Variable		y1		y2		Residual Cross-Covariance Matrices		
1	y1		1.20790		-0.52848		0	y1	0.83413
	y2		0.58047		0.29561		1	y2	0.59108
							1	y1	0.09129
							2	y2	0.13094
							2	y1	-0.13532
								y2	-0.11822
Model Parameter Estimates									
Equation	Parameter	Estimate	Std Error	T Ratio	Prob> T	Variable	0	y1	0.59108
y1(t)	AR1_1_1	1.20790	0.06766	17.85	0.0001	y1(t-1)		y2	1.26089
	AR1_1_2	-0.52848	0.06925	-7.63	0.0001	y2(t-1)			0.16000
y2(t)	AR1_2_1	0.58047	0.08304	6.99	0.0001	y1(t-1)			0.03236
	AR1_2_2	0.29561	0.08500	3.48	0.0008	y2(t-1)			-0.15877
							2	y1	-0.05082

où apparaît la matrice autorégressive à l'ordre 1, relativement proche des valeurs théorique (*du modèle simulé*),

$$\hat{\Phi} = \begin{bmatrix} 1.20790 & -0.52848 \\ 0.58047 & 0.29561 \end{bmatrix} \text{ relativement proche de } \Phi = \begin{bmatrix} 1.2 & -0.5 \\ 0.6 & 0.3 \end{bmatrix},$$

dont les précisions (*écart-type de l'estimateur par exemple*) sont donnés dans le tableau en bas à gauche. La matrice de variance-covariance du bruit est elle aussi relativement bien estimée,

$$\hat{\Sigma} = \begin{bmatrix} 0.85478 & 0.60495 \\ 0.60495 & 1.28770 \end{bmatrix} \text{ relativement proche de } \Sigma = \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.25 \end{bmatrix}.$$

SAS donne aussi des informations permettant de faire par la suite du choix de modèle, avec les critère *AIC* corrigé, de Hannan-Quinn, ou *AIC* et *SBC*. La sortie ci-dessous permet de vérifier si l'on n'a pas oublié de composante autorégressive dans la régression, puisque SAS teste un modèle de la forme $Z_t = \Phi_1 Z_{t-1} + \Phi_2 Z_{t-2} + \Phi_3 Z_{t-3}$: comme

le montre le petit tableau en bas à droite, seule la première composante est significative.

The VARMAX Procedure									
Residual Cross-Covariance Matrices									
Lag	Variable	y1	y2	Portmanteau Test for Residual Cross Correlations					
3	y1	-0.06943	-0.02261	To Lag	Chi-Square				
	y2	-0.12722	-0.06607	DF	Prob>ChiSq				
				2	6.74				
				3	8.40				
					0.3957				
Residual Cross-Correlation Matrices									
Lag	Variable	y1	y2	The VARMAX Procedure					
0	y1	1.00000	0.57636	Univariate Model Diagnostic Checks					
	y2	0.57636	1.00000	Variable	R-square	StdDev	F Value	Prob>F	
1	y1	0.10945	0.07859	y1	0.8003	0.9245	388.84	<.0001	
	y2	0.12768	0.02566	y2	0.6907	1.1348	216.63	<.0001	
2	y1	-0.16223	-0.15481	Univariate Model Diagnostic Checks					
	y2	-0.11528	-0.04030	Variable	DW(1)	Normality ChiSq	Prob>ChiSq	ARCH1 F Value	Prob>F
3	y1	-0.07604	-0.02204	y1	1.77	0.85	0.6541	1.39	0.2408
	y2	-0.12405	-0.05240	y2	1.93	0.05	0.9742	0.60	0.4417
				Schematic Representation of Residual Cross Correlations					
				Variable					
				Lag	0	1	2	3	
				y1	++	
				y2	++	
					+ is > 2*std error, - is < -2*std error, . is between				

SAS donne alors des prévisions sur un horizon plus ou moins long, sous hypothèse de modèle $VAR(1)$,

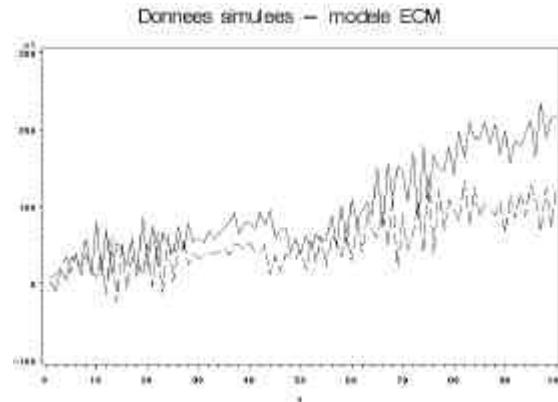
Univariate Model Diagnostic Checks									
Variable	AR1		AR1-2		AR1-3		AR1-4		Prob>F
	F Value	Prob>F							
y1	1.17	0.2816	2.07	0.1324	1.42	0.2425	1.55	0.1944	
y2	0.06	0.7995	0.13	0.8755	0.13	0.9443	0.25	0.9069	
Forecasts									
Variable Obs Forecast Standard Error 95% Confidence Limits									
y1 101 0.5676 0.9245 -1.2444 2.3797									
102 1.2602 1.2997 -1.2872 3.8075									
103 1.5178 1.5904 -1.5993 4.6350									
104 1.4456 1.7990 -2.0805 4.9716									
105 1.1958 1.995 -2.6139 4.4556									
y2 101 1.0971 1.5348 -1.12 1.1370									
102 0.00814 1.3760 2.6905 2.7068									
103 0.7339 1.5418 -2.2881 3.7558									
104 1.0980 1.6965 -2.2270 4.4230									
105 1.1637 1.8290 -2.4211 4.7484									

1.6.2 Estimation d'un modèle à correction d'erreur

Les sorties ci-dessous correspondent à l'estimation d'un modèle à correction d'erreur, noté $VECM$ sous SAS. Considérons le modèles $VAR(2)$ non stationnaire suivant

$$Y_t = \begin{pmatrix} Y_{1,t} \\ Y_{2,t} \end{pmatrix} = \begin{pmatrix} -0.2 & 0.1 \\ 0.5 & 0.2 \end{pmatrix} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} + \begin{pmatrix} 0.8 & 0.7 \\ -0.4 & 0.6 \end{pmatrix} \begin{pmatrix} Y_{1,t-2} \\ Y_{2,t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix},$$

où le bruit blanc admet pour matrice de variance-covariance $\Sigma = 100\mathbb{I}_2$ dont une simulation (*à partir de la valeur initiale $Y_{1,0} = Y_{2,0} = 0$*) a l'allure suivante



Remarque 11 Ce processus est en effet non-stationnaire : On retrouve d'ailleurs sur le graphique que ces deux séries sont cointégrées.

L'écriture *VECM* de ce processus a la forme suivante

$$\Delta Y_t = \begin{pmatrix} -0.4 \\ 0.1 \end{pmatrix} \begin{pmatrix} 1 & -2 \end{pmatrix} Y_{t-1} - \begin{pmatrix} 0.8 & 0.7 \\ -0.4 & 0.6 \end{pmatrix} \Delta Y_{t-1} + \varepsilon_t.$$

Considérons alors le programme *SAS* suivant

```
PROC VARMAX data=donnes;
  MODEL X Y / p=2 noint dftest cointtest=(johansen);
  run ;
```

The VARMAX Procedure										Long-Run Parameter BETA Estimates											
										Variable	Dummy 1	Dummy 2									
		Number of Observations		100						y1	-0.04867	0.03109									
		Number of Pairwise Missing		0						y2	0.09518	-0.04155									
										Adjustment Coefficient ALPHA Estimates											
		y1		DEP		100		96.42914		Mean	62.79930	233.29867									
		y2		DEP		54.56026		37.03117		StdDev	-24.92182	135.77617									
										Min	-11.71118	-1.65									
										Max	233.29867	233.29867									
Dickey-Fuller Unit Root Tests										Variable	Dummy 1	Dummy 2									
										y1	9.59171	0.25525									
										y2	-2.19184	1.07837									
Cointegration Rank Test										The VARMAX Procedure											
		H_0:		H_1:		Eigenvalue		Trace		Critical Value	Drift InECM	Type of Model									
		Rank=r		Rank>r		0.5086		70.73		12.21	NOINT	Least Squares Estimation									
		0		1		0.0111		1.09		4.14	Constant	VAR(2)									
AR Coefficient Estimates										Lag	Variable	y1	y2								
										1	y1	-0.20804	0.16002								
										2	y1	0.52041	0.19141								
										y2	0.74917	0.74233									
										y2	-0.38021	0.55517									

la sortie ci-dessus à gauche montre que chacune des composantes est effectivement intégrée d'ordre 1, à l'aide des tests *ADF*. Le test de cointégration de Johansen nous conduit à rejeter l'hypothèse de rang $r = 0$ (*première ligne du tableau* : $70.73 > 12.21$) et à accepter l'hypothèse de rang $r = 1$ (*seconde ligne du tableau* : $1.09 < 4.14$). On peut d'ailleurs noter que l'option **Noint** dit qu'il n'y a pas de constante dans la forme *ECM*, mais qu'il y a un terme constant dans l'écriture à long terme. : en conclusion, il existe une relation de cointégration entre Y_1 et Y_2 . La sortie présenté ci-dessus à droite donne également une estimation des termes constants α et β : $\hat{\beta} = (-1.07837 \quad -2.19184)'$ correspondant à la première colonne de la matrice **BETA** et $\hat{\alpha} = (-0.04867 \quad 0.09518)$ correspondant à la seconde ligne de la matrice **ALPHA**. On retrouve également une estimation des matrices autorégressive,

$$\hat{\Phi}_1 = \begin{bmatrix} -0.208 & 0.160 \\ 0.504 & 0.191 \end{bmatrix} \text{ et } \hat{\Phi}_2 = \begin{bmatrix} 0.749 & 0.742 \\ -0.380 & 0.555 \end{bmatrix}, \text{ avec } \Phi_1 = \begin{pmatrix} -0.2 & 0.1 \\ 0.5 & 0.2 \end{pmatrix} \text{ et } \Phi_2 = \begin{pmatrix} 0.8 & 0.7 \\ -0.4 & 0.6 \end{pmatrix},$$

De plus, on obtient, sur la sortie ci-dessous à gauche, la matrice de variance covariance des résidus,

$$\hat{\Sigma} = \begin{bmatrix} 98.72 & 4.432 \\ 4.432 & 113.02 \end{bmatrix} \text{ relativement proche de } \Sigma = \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix}$$

On obtient alors des sorties relativement proches de celles décrites dans le paragraphe précédent.

Model Parameter Estimates										Portmanteau Test for Residual Cross Correlations					
Equation	Parameter	Estimate	Std Error	T Ratio	Prob> T	Variable	To Lag	Chi-Square	DF	Prob>					
y1(t)	AR1_1_1	-0.20804	0.03375	-6.16	0.0001	y1(t-1)	3	9.95	4	0.0410					
	AR1_1_2	-0.16002	0.03766	2.36	0.0201	y2(t-1)	4	13.90	8	0.0848					
	AR2_1_1	0.74917	0.05161	14.52	0.0001	y1(t-2)	5	18.06	12	0.1137					
	AR2_1_2	0.74233	0.05102	14.55	0.0001	y2(t-2)	6	22.49	16	0.1281					
y2(t)	AR1_2_1	0.52041	0.03612	14.41	0.0001	y1(t-1)	7	23.04	20	0.2867					
	AR1_2_2	0.19141	0.07240	2.64	0.0096	y2(t-1)	8	24.70	24	0.4222					
	AR2_2_1	-0.38021	0.05522	-6.88	0.0001	y1(t-2)	9	27.06	28	0.5147					
	AR2_2_2	0.55517	0.05459	10.17	0.0001	y2(t-2)	10	29.42	32	0.5979					
							11	30.35	36	0.7338					
							12	33.07	40	0.7729					
Covariance Matrix for the Innovation										The VARMAX Procedure					
Variable		y1		y2			Univariate Model Diagnostic Checks								
y1		98.71098		4.43251			Variable	R-square	StdDev	F Value	Prob>F				
y2		4.43251		113.02058			y1	0.9750	9.9358	1219.97	<.0001				
							y2	0.9199	10.6311	359.87	<.0001				
Information Criteria										Univariate Model Diagnostic Checks					
AIC(Corrected AIC)							Normality								
HQ(Hannan-Quinn Criterion)							Variable	DW(1)	ChiSq	Prob>	ARCH1				
AIC(Akaike Information Criterion)							y1	1.93	2.97	0.2264	0.16	0.6931			
SC(Schwarz Bayesian Criterion)							y2	1.77	0.25	0.8808	0.00	0.9694			
FPE(Final Prediction Error Criterion)															
12066.5															

Univariate Model Diagnostic Checks										Model Parameter Estimates																					
Variable	F Value	Prob>F	F Value	Prob>F	F Value	Prob>F	F Value	Prob>F	Equation	Parameter	Estimate	Std Error	T Ratio	Prob> T	Variable																
y1	0.02	0.8797	0.74	0.4815	0.70	0.5563	0.61	0.6554	y1(t)	AR1_1_1	-0.20804	0.03375	-6.16	0.0001	y1(t-1)																
y2	0.91	0.3420	2.63	0.0773	1.73	0.1655	1.36	0.2544		AR1_1_2	0.16002	0.06766	2.36	0.0201	y2(t-1)																
The VARMAX Procedure																															
Type of Model	VAR(2)		Covariance Matrix for the Innovation										Information Criteria																		
Estimation Method	Least Squares Estimation		Variable	y1		y2		y1	98.71998	4.42821		y2	4.43251	113.02058		AICC(Corrected AIC)	9.404962														
AR Coefficient Estimates																															
Lag	Variable	y1	y2	Variable	y1	y2	Variable	y1	y2	Variable	y1	y2	Variable	y1	y2	HQIC(Hannan-Quinn Criterion)	9.483367														
1	y1	-0.26804	0.16002	1	0.52041	0.19141	1	0.74917	0.74233	1	0.5311	0.1612	1	0.49141	0.19141	AIC(Akaike Information Criterion)	9.398015														
2	y2	-0.38021	0.65517													SBC(Schwarz Bayesian Criterion)	9.609033														
																FPEG(Final Prediction Error Criterion)	12065.5														
Estimating the VECM(2) model with one cointegrating relationship																															
The VARMAX Procedure																															
Type of Model	VECM(2)		AR Coefficient Estimates										DIF_Lag	Variable	y1	y2															
Estimation Method	Maximum Likelihood Estimation		1	1	y1	-0.7432	-0.74621	y2	0.40493	-0.57157																					
Cointegrated Rank	1		Model Parameter Estimates																												
Long-Run Parameter BETA Estimates given RANK=1																															
Variable	Dummy 1		Equation	Parameter	Estimate	Std Error	T Ratio	Prob> T	Variable																						
y1	1.00000	D_y1(t)	AR1_1_1	-0.46680	0.04786	-9.75	0.0001	y1(t-1)	0.91295	0.09359	9.75	0.0001	y2(t-1)																		
y2	-1.98575		AR1_1_2	-0.7432	0.04526	-16.42	0.0001	D_y1(t-1)	AR2_1_1	-0.7432	0.04769	-15.65	0.0001	D_y2(t-1)																	
Adjustment Coefficient ALPHA Estimates given RANK=1																															
Variable	Dummy 1		D_y2(t)	AR1_2_1	0.10667	0.05146	2.07	0.0409	y1(t-1)	AR1_2_2	-0.20862	0.10064	-2.07	0.0409	y2(t-1)																
y1	-0.46680		AR2_2_1	0.40493	0.04867	8.32	0.0001	D_y1(t-1)	AR2_2_2	-0.57157	0.05128	-11.15	0.0001	D_y2(t-1)																	
y2	0.10667																														
Covariance Matrix for the Innovation																															
Variable	y1		Variable	y2		Variable	y1		Variable	y2		Variable	y1		Variable	y2		Variable	y1		Variable	y2		Variable	y1		Variable	y2			
y1	94.75575		94.75575	4.52684		4.52684	0	1	2	3	4	5	6	7	8	9	10	11	12												
y2	4.52684			109.57038			y1	+												
Information Criteria																				+ is > 2*std error, - is < -2*std error, . is between											
AICC(Corrected AIC)	9.37221		HQIC(Hannan-Quinn Criterion)	9.432357		AIC(Akaike Information Criterion)	9.368343		SBC(Schwarz Bayesian Criterion)	9.526606		FPEG(Final Prediction Error Criterion)	11712.14		Portmanteau Test for Residual Cross Correlations																
Residual Cross-Covariance Matrices																				To	Chi-Square	DF	ChiSq								
Lag	Variable	y1	y2	Variable/Lag	0	1	2	3	4	5	6	7	8	9	10	11	12														
0	y1	94.66850	5.04865	1.70467	1.115994	1.25695	-11.01656	-0.72040	-6.74860	-7.66607	5.04511	15.74197	-0.16002	0.92414	0.73321	-0.06990		3	10.55	4	0.0321										
1	y2	5.04865	106.4947	1.115994	-1.115994	-1.115994	-4.97710	-25.06356	6.14494	-4.25547	8.60918	-9.33422	8.31450	-17.81379	6.74612	1.81515		4	14.43	8	0.0713										
2	y1	1.70467	1.115994	1.25695	1.25695	7.13978	-4.97710	-25.06356	6.14494	-4.25547	8.60918	-9.33422	8.31450	-17.81379	6.74612	1.81515		5	18.69	12	0.0963										
3	y2	1.115994	-1.115994	-1.115994	-1.115994	-7.13978	-4.97710	-25.06356	6.14494	-4.25547	8.60918	-9.33422	8.31450	-17.81379	6.74612	1.81515		6	23.51	16	0.1008										
4	y1	1.25695	-1.115994	-4.97710	-4.97710	-11.01656	-0.72040	-25.06356	-6.14494	-4.25547	-8.60918	-9.33422	-8.31450	-17.81379	-6.74612	-1.81515		7	24.06	20	0.2399										
5	y2	7.13978	-4.97710	-25.06356	-25.06356	-11.01656	-0.72040	-25.06356	-6.14494	-4.25547	-8.60918	-9.33422	-8.31450	-17.81379	-6.74612	-1.81515		8	25.92	24	0.4444										
6	y1	-0.72040	-25.06356	-25.06356	-25.06356	-11.01656	-4.97710	-25.06356	-6.14494	-4.25547	-8.60918	-9.33422	-8.31450	-17.81379	-6.74612	-1.81515		9	29.15	28	0.4565										
7	y2	-6.14494	-4.25547	-6.14494	-6.14494	-4.25547	-0.72040	-6.14494	-4.25547	-8.60918	-9.33422	-8.31450	-17.81379	-6.74612	-1.81515		10	30.93	32	0.5208											
8	y1	-4.25547	-8.60918	-8.60918	-8.60918	-9.33422	-4.25547	-8.60918	-8.60918	-8.60918	-9.33422	-8.31450	-17.81379	-6.74612	-1.81515		11	31.95	36	0.6616											
9	y2	-8.60918	-9.33422	-9.33422	-9.33422	-8.60918	-4.25547	-8.60918	-8.60918	-8.60918	-9.33422	-8.31450	-17.81379	-6.74612	-1.81515		12	34.63	40	0.7103											
Univariate Model Diagnostic Checks																				Variable	R-square	StdDev	F Value	Prob>F							
													y1	0.9391	9.7343	482.78	<.0001														
													y2	0.9409	10.4676	498.42	<.0001														

Univariate Model Diagnostic Checks							
Variable	DW(1)	Normality	Prob>ChiSq	ARCH1	F Value	Prob>F	
y1	1.93	2.14	0.3437	0.20	0.6536		
y2	1.83	0.46	0.7951	0.03	0.8545		

Estimating the VECM(2) model with one cointegrating relationship							
The VARMAX Procedure							
Univariate Model Diagnostic Checks							
Variable	AR1	F Value	Prob>F	AR1-2	F Value	Prob>F	AR1-4
y1	0.03	0.8541		0.69	0.5049	0.64	0.5881
y2	0.44	0.5068		3.11	0.0491	2.01	0.1181
							F Value
							Prob>F

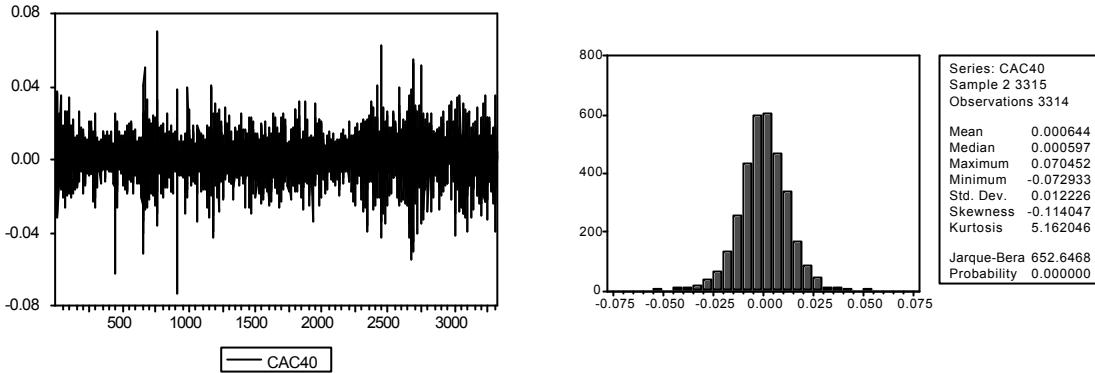
Infinite Order AR Representation							
Lag	Variable	y1	y2				
1	y1	-0.21013	0.16674				
	y2	0.51160	0.21980				
2	y1	0.74332	0.74621				
	y2	-0.40493	0.57157				

1.7 Conseils bibliographiques

- GOURIEROUX,C. & MONFORT,A. (1995). Séries temporelles et modèles dynamiques. Economica.
- LUTKEPOHL,H. (1991). Introduction to multiple time series analysis. Springer
- DAVISON,R. & MacKINNON (1993) Estimation and Inference in Econometrics. Oxford University Press, New York

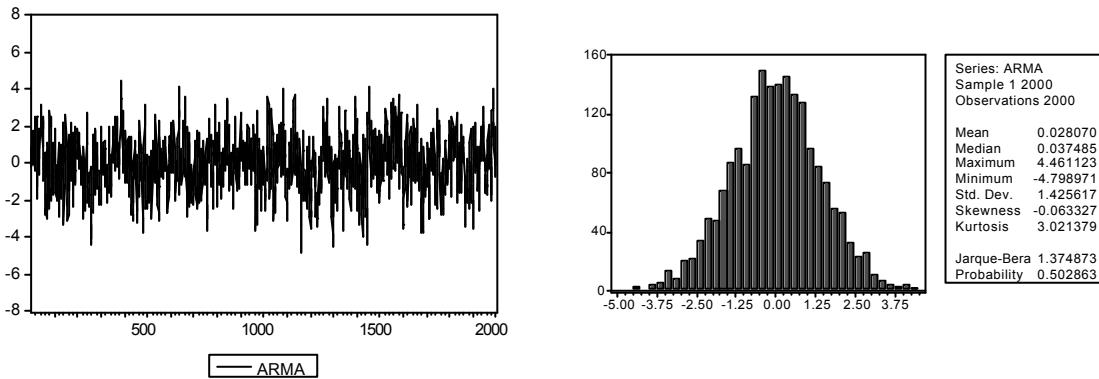
2 Les modèles *ARCH* - Autorégressifs Conditionnellement Hétérosclématiques

Les modèles *ARCH* ont été introduits par Engle en 1982, puis généralisés en modèles *GARCH* en 1986 par Bollerslev. L'analogie entre les modèles *ARCH* en temps discret et les modèles de diffusion en temps continu a été établie par Nelsen en 1990. Ce analogie a permis, en particulier, le développement des modèles à volatilité stochastique. Un des apports des modèles *ARCH* était de mieux s'ajuster aux données (*en particulier aux données financières*) que ne le faisaient les modèles *ARMA*. Considérons la série suivante, correspondant au rendement de l'indice CAC40



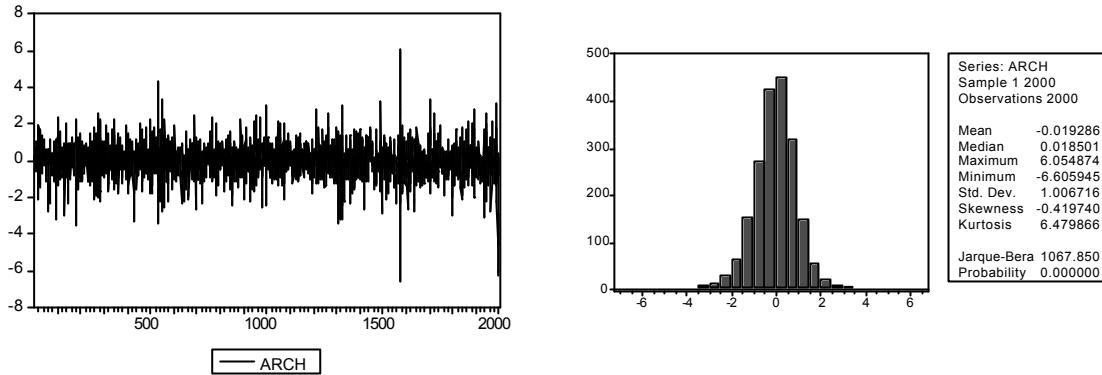
avec à droite, l'histogramme, correspondant à la distribution marginale de r_t . La première constatation que l'on peut faire est que la série n'est pas gaussienne, et, en particulier, a des queues de distribution trop épaisses (*kurtosis supérieure à 3*). La seconde est que la série est stationnaire, mais que l'on observe quand même des périodes de forte variation, ou de forte volatilité.

Si l'on considère un processus *AR*, avec un bruit blanc gaussien, il est relativement difficile de modéliser ce genre de comportement.



En effet, le processus *AR*(1) est un processus gaussien : les queues de distribution sont moins épaisses que les queues observées sur la variation de l'indice CAC40, et on n'observe pas de 'période de haute volatilité'. Les modèles *ARCH*

(simulé ci-dessous) permettent, eux, de mieux prendre en compte ce genre de comportement :



Les queues de distribution peuvent être plus épaisses que celles des lois normales (*kurtosis* de 6.5 avec les paramètres choisis), et on observe, comme sur les données empiriques, des zones de forte volatilité.

Ces modèles non-linéaires permettent d'avoir des propriétés que n'avaient pas les modèles linéaires (ARMA), et permettent ainsi de modéliser de façon plus réaliste un grand nombre de données, en particulier les données financières.

2.1 Notions de stationnarité et notions de linéarité

Définition 6 Stationnarité au second ordre - Un processus (Y_t) 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 : $\text{cov}(X_t, X_s) = \gamma(|t-s|)$,

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

Définition 7 Stationnarité au sens fort - Un processus (Y_t) est dit stationnaire au sens fort si quel que soit n , t_1, \dots, t_n et h , on a l'égalité entre les lois jointes

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

Théorème 1 Théorème de Wold - Soit (Y_t) un processus stationnaire au second ordre. Alors (Y_t) peut s'écrire

$$Y_t = c_0 + \sum_{i=0}^{\infty} \theta_i \varepsilon_{t-i} \text{ où } \varepsilon_t \text{ est un bruit blanc au second ordre.}$$

L'écriture est ici plus simple que dans le théorème (??) puisque l'on impose plus ici de contraintes pour avoir unicité de la décomposition.

Théorème 2 Théorème de Volterra - Soit (Y_t) un processus stationnaire au sens fort. Alors (Y_t) peut s'écrire

$$Y_t = c_0 + \sum_{i=0}^{\infty} \theta_i \varepsilon_{t-i} + \sum_{i,j=0}^{\infty} \theta_{ij} \varepsilon_{t-i} \varepsilon_{t-j} + \sum_{i,j,k=0}^{\infty} \theta_{ijk} \varepsilon_{t-i} \varepsilon_{t-j} \varepsilon_{t-k} + \dots \text{ où } \varepsilon_t \text{ est un bruit blanc fort gaussien.}$$

2.2 Présentation des modèles ARCH

Soit (X_t) un processus AR(1), tel que $X_t = \alpha X_{t-1} + \varepsilon_t$ où $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, alors

$$V(X_t) = \frac{1}{1-\alpha^2} \sigma^2 \text{ et } V(X_t | X_{t-1}) = \sigma^2,$$

c'est à dire que la variance, et la variance conditionnelle, ne dépendent pas du temps.

Soit (X_t) un processus ARCH(1), tel que $X_t = \varepsilon_t \sqrt{\alpha_0 + \alpha_1 X_{t-1}^2}$ où $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, alors

$$V(X_t | X_{t-1}) = [\alpha_0 + \alpha_1 X_{t-1}^2] \sigma^2.$$

Définition 8 On dit que le processus (X_t) suit un processus ARCH (p) s'il est défini par une équation de la forme

$$X_t = \varepsilon_t \sqrt{h_t} \text{ où } h_t = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2,$$

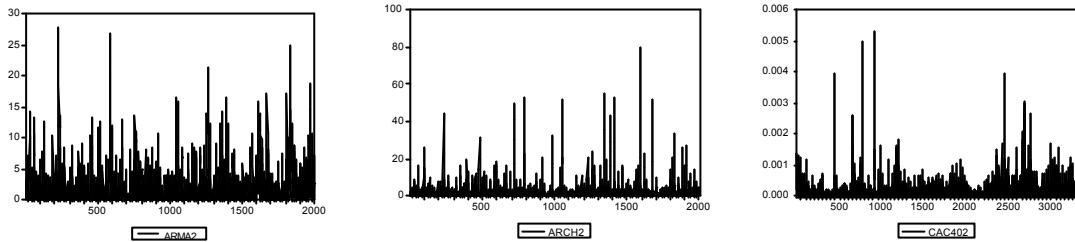
où (ε_t) est un bruit blanc gaussien, centré, de variance σ^2 , soit $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.

Ceci peut aussi s'écrire

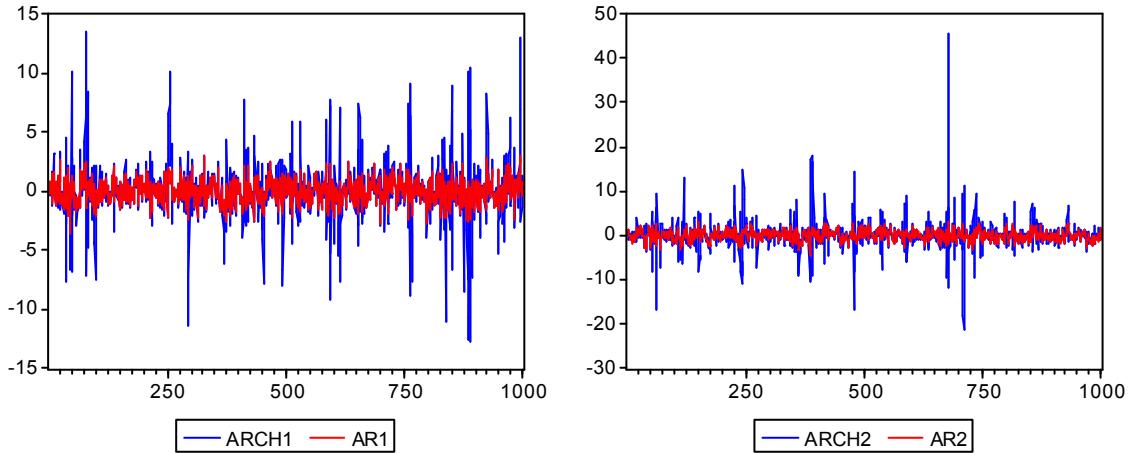
$$X_t^2 = \varepsilon_t^2 \left[\alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 \right].$$

Ce n'est plus le processus (X_t) que l'on cherche à modéliser, mais le processus (X_t^2) .

Exemple 13 Les graphiques montrent l'évolution des processus (X_t^2) dans le cas d'un modèle ARMA à gauche, d'un modèle ARCH (1) au centre, et du rendement de l'indice CAC40 à droite



Exemple 14 Les graphiques ci-dessous permettent de comparer un processus AR (1) et un processus ARCH (1)



Remarque 12 Les variables X_t sont non-corrélatées, mais ne sont pas indépendantes.

2.2.1 Processus ARCH (1)

Définition 9 Le processus (X_t) est un processus ARCH (1) si $X_t = \varepsilon_t \sqrt{\alpha_0 + \alpha_1 X_{t-1}^2}$ où (ε_t) est un bruit blanc gaussien, $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$. On notera généralement $h_t = \alpha_0 + \alpha_1 X_{t-1}^2$, et donc $X_t = \sqrt{h_t} \cdot \varepsilon_t$.

Dans le cas où (ε_t) est un bruit blanc fort (ε_t i.i.d.) gaussien, alors (ε_t) peut être vu comme une différence de martingale. Pour que le processus (X_t) soit stationnaire au second ordre, la variance marginale de (X_t) doit être constante. Or $V(X_t) = \alpha_0 + \alpha_1 V(X_{t-1})$, c'est à dire que l'on doit avoir $\alpha_1 < 1$, et donc $V(X_t) = \alpha_0 / (1 - \alpha_1)$.

Propriété 4 (X_t) est stationnaire au second ordre si $\alpha_0 > 0$ et $0 \leq \alpha_1 < 1$.

On peut alors noter que pour tout $h \geq 1$,

$$\mathbb{E} \left(X_t | \underline{X_{t-h}} \right) = 0,$$

puisque $\mathbb{E} \left(X_t | \underline{X_{t-h}} \right) = \mathbb{E} \left(\mathbb{E} \left(X_t | \underline{X_{t-1}} \right) | \underline{X_{t-h}} \right) = \mathbb{E} \left(0 | \underline{X_{t-h}} \right) = 0$. De plus,

$$V \left(X_t | \underline{X_{t-h}} \right) = \alpha_0 \frac{1 - \alpha_1^h}{1 - \alpha_1} + \alpha_1 X_{t-h}^2 \text{ et } V(X_t) = \frac{\alpha_0}{1 - \alpha_1}. \quad (15)$$

Aussi,

$$V \left(X_t | \underline{X_{t-h}} \right) - V(X_t) = \alpha_1^h [X_{t-h}^2 - V(X_{t-h})].$$

Comme nous l'avions vu en introduction, les modèles *ARCH* permettent d'avoir des processus avec des queues de distribution plus épaisse que les processus *ARMA*. La kurtosis conditionnelle est obtenue à l'aide de la relation

$$\mathbb{E} \left(X_t^4 | \underline{X_{t-1}} \right) = 3 [\alpha_0 + \alpha_1 X_{t-1}^2]^2,$$

et la kurtosis à la date t est

$$\mathbb{E} (X_t^4) = 3 \left[\alpha_0^2 + \frac{2\alpha_1\alpha_0^2}{1 - \alpha_1} + \alpha_1^2 \mathbb{E} (X_{t-1}^4) \right] = \frac{3\alpha_0^2 (1 + \alpha_1)}{(1 - 3\alpha_1^2) (1 - \alpha_1)}.$$

La condition d'existence du moment d'ordre 4 est $3\alpha_1^2 < 1$, et on en déduit alors l'expression de la kurtosis

$$\kappa = \frac{\mathbb{E} (X_t^4)}{\mathbb{E} (X_t^2)^2} = 3 \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} > 3 \text{ (cas de la loi normale).}$$

Les queues de la distribution marginale d'un processus *ARCH* (1) sont donc plus épaisse que pour un processus gaussien (*on parlera de distribution leptokurtique*).

2.2.2 Processus *ARCH* (p)

Les modèles *ARCH* (p) sont des extensions des modèles *ARCH* (1). Les modèles *ARCH* (1) étaient

$$X_t = \varepsilon_t h_t \text{ où } h_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 \text{ et } \varepsilon_t \sim \mathcal{N}(0, \sigma^2),$$

et les modèles *ARCH* (p) font intervenir plusieurs retards,

$$X_t = \varepsilon_t h_t \text{ où } h_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 \text{ et } \varepsilon_t \sim \mathcal{N}(0, \sigma^2).$$

La volatilité de la date t est alors fonction des carrés des écarts à la moyenne observés dans le passé proche. Si les coefficients α_i sont tous positifs (*et assez grands*), il y a une persistance des niveaux de volatilité : on observe des périodes de forte volatilité suivies de périodes de faible volatilité. On peut alors écrire

$$\mathbb{E} \left(X_t | \underline{X_{t-1}} \right) = 0, \text{ et } \mathbb{E} \left(X_t^2 | \underline{X_{t-1}} \right) = h_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2.$$

2.2.3 Processus *GARCH* (p, q)

Ces modèles ont été introduits par Bollerslev en 1986, inspirés de la démarche de Box et Jenkins, avec une dynamique autorégressive,

$$X_t = \varepsilon_t h_t \text{ où } h_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j}^2 \text{ et } \varepsilon_t \sim \mathcal{N}(0, \sigma^2). \quad (16)$$

On peut noter tout d'abord que, sous cette forme, les coefficients p et q ne sont pas analogues à ceux des modèles *ARMA* : en particulier, q correspond au caractère autorégressif du processus (h_t^2). On a alors

$$\mathbb{E} \left(X_t | \underline{X_{t-1}} \right) = 0 \text{ et } \mathbb{E} \left(X_t^2 | \underline{X_{t-1}} \right) = h_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j}^2.$$

Propriété 5 Si le processus $GARCH(p, q)$ est stationnaire au second ordre, alors nécessairement

$$\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1, \quad (17)$$

et dans ce cas, la variance de X_t est

$$V(X_t) = \frac{\alpha_0}{1 - \sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j}.$$

Dans le cas où (17) est saturée, i.e. $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j = 1$, on dira alors que le processus $GARCH(p, q)$ est intégré, et on parlera de processus $IGARCH(p, q)$. Cette dénomination peut se justifier par l'existence d'une racine unité dans la composante autorégressive de (16). Toutefois, cette analogie avec l'extension des modèles $ARMA$ aux modèles $ARIMA$ peut être trompeuse : un processus $ARIMA$ n'est pas stationnaire (*au second ordre ou au sens fort*) alors qu'il existe une solution stationnaire (*au sens fort*) d'un modèle $IGARCH$ (*qui admettra une variance infinie d'après la propriété ci-dessus*).

Dans le cas d'un processus $GARCH(1, 1)$, il est possible de montrer que la kurtosis de X_t est de la forme

$$\kappa(X_t) = \frac{1 - (\alpha_1 + \beta_1)^2}{1 - (\alpha_1 + \beta_1)^2 - \alpha_1^2(\mu_4 - 1)} \kappa(\varepsilon_t) \text{ où } \mu_4 = \mathbb{E}(\varepsilon_t^4).$$

2.3 Modèles avec erreurs $ARCH$

Nous allons considérer un processus (X_t) dont la dynamique est régie par un processus de type $AR(1)$:

$$X_t = \rho X_{t-1} + \varepsilon_t$$

Dans la partie sur les processus $ARIMA$, nous avions étudier le cas où (ε_t) correspondait à un bruit blanc, c'est à dire non-autocorrélé avec son passé.

2.3.1 Erreurs $ARCH(1)$

Nous allons regarder les propriétés de (Y_t) , processus autorégressif, dans le cas où (ε_t) n'est plus un bruit blanc, en particulier lorsque (ε_t) est un processus $AR(1)$, $\varepsilon_t = \theta \varepsilon_{t-1} + \eta_t$, et lorsque (ε_t) est un processus $ARCH(1)$, $\varepsilon_t = \eta_t \sqrt{1 + \alpha \varepsilon_{t-1}^2}$. Le premier cas correspond aux modèles $ARMA$, et le second correspond aux modèles autorégressifs à erreurs $ARCH$.

Considérons le modèle

$$Y_t = \mu + \phi Y_{t-1} + \varepsilon_t \quad \text{où } |\phi| < 1 \text{ et } \varepsilon_t \text{ satisfait} \quad \begin{cases} (1) : \mathbb{E}(\varepsilon_t | \underline{\varepsilon_{t-1}}) = 0 \\ (2) : V(\varepsilon_t | \underline{\varepsilon_{t-1}}) = c + \alpha \varepsilon_{t-1}^2. \end{cases}$$

L'espérance conditionnelle de (Y_t) est alors

$$\mathbb{E}(Y_t | \underline{Y_{t-h}}) = \mu + \phi \mathbb{E}(Y_{t-1} | \underline{Y_{t-h}}) = \mu \frac{1 - \phi^h}{1 - \phi} + \phi Y_{t-h},$$

et sa variance conditionnelle, en utilisant (15)

$$V(Y_t | \underline{Y_{t-1}}) = \frac{c}{1 - \alpha} \left[\frac{1 - \phi^{2h}}{1 - \phi^2} - \alpha \frac{\phi^h - \phi^{2h}}{\alpha - \phi^2} \right] + \alpha \frac{\phi^h - \phi^{2h}}{\alpha - \phi^2} \varepsilon_{t-h}^2,$$

qui n'est pas constante (*puisque elle dépend de ε_{t-h}*).

Dans le cas d'erreurs $ARCH$, la condition (1) : $\mathbb{E}(\varepsilon_t | \underline{\varepsilon_{t-1}}) = 0$ signifie que (ε_t) est un bruit blanc faible qui satisfait la condition de différence de martingale, et dont la variance conditionnelle $V(\varepsilon_t | \underline{\varepsilon_{t-1}})$ suit un processus $ARCH(1)$: $\varepsilon_t^2 = c + \alpha \varepsilon_{t-1}^2 + \eta_t$, où (η_t) est un bruit blanc gaussien. Les propriétés du processus d'innovation (ε_t) sont alors les suivantes,

Propriété 6 (i) Le processus (ε_t) est orthogonal aux valeurs passées, quel que soit le retard : $\mathbb{E}(\varepsilon_t | \varepsilon_{t-h}) = 0$ pour tout h

(ii) La propriété d'orthogonalité implique que les corrélations conditionnelles sont nulles : $\text{cov}(\varepsilon_t, \varepsilon_{t+h} | \varepsilon_{t-h}) = 0$.

Preuve. Cette seconde propriété s'obtient de la façon suivante

$$\begin{aligned} \text{cov}(\varepsilon_t, \varepsilon_{t+k} | \varepsilon_{t-h}) &= \mathbb{E}(\varepsilon_t \varepsilon_{t+k} | \varepsilon_{t-h}) - \mathbb{E}(\varepsilon_t | \varepsilon_{t-h}) \mathbb{E}(\varepsilon_{t+k} | \varepsilon_{t-h}) = \mathbb{E}(\varepsilon_t \varepsilon_{t+k} | \varepsilon_{t-h}) \\ &= \mathbb{E}((\varepsilon_t \varepsilon_{t+k} | \varepsilon_{t+k-1}) | \varepsilon_{t-h}) = \mathbb{E}(\varepsilon_t \mathbb{E}(\varepsilon_{t+k} | \varepsilon_{t+k-1}) | \varepsilon_{t-h}) = 0. \end{aligned}$$

■

Il y a donc absence de corrélation entre les valeurs présentes et futures du processus, quels que soient les retards h et k . Mais si la variance conditionnelle de ε_t n'est pas constante, la variance (*non conditionnelle*) est constante :

$$\varepsilon_t^2 = c + \alpha \varepsilon_{t-1}^2 + \eta_t = \dots + c(1 + \alpha + \alpha^2 + \dots + \alpha^{h-1}) + \alpha^h \varepsilon_{t-h}^2 + \eta_t + \alpha \eta_{t-1} + \dots + \alpha^{h-1} \eta_{t-h+1},$$

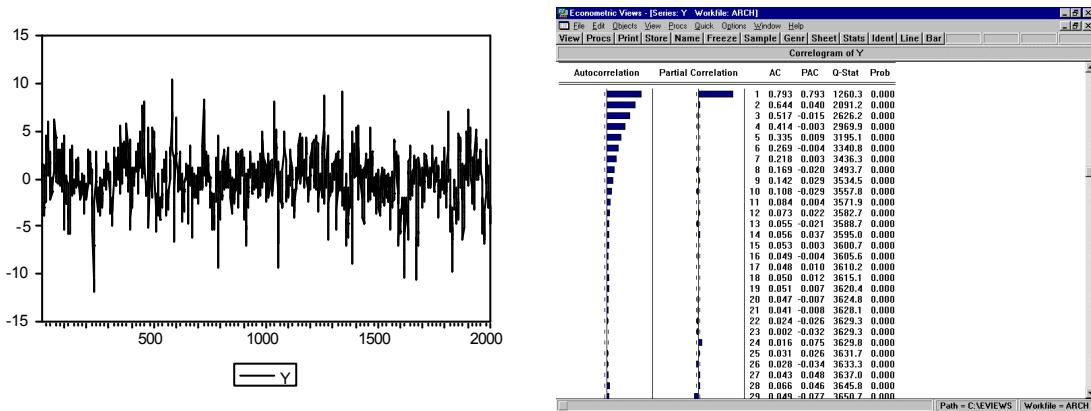
et donc

$$V(\varepsilon_t | \varepsilon_{t-h}) = \mathbb{E}(\varepsilon_t^2 | \varepsilon_{t-h}) = c \frac{1 - \alpha^h}{1 - \alpha} + \alpha^h \varepsilon_{t-h}^2,$$

alors que

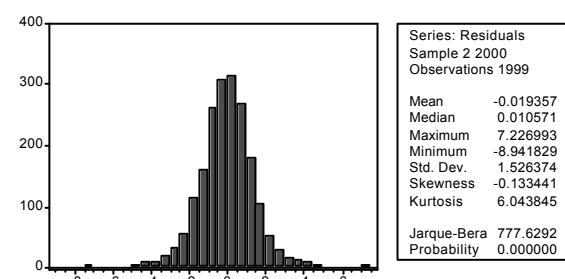
$$\sigma^2 = V(\varepsilon_t) = \mathbb{E}(V(\varepsilon_t | \varepsilon_{t-h})) = \frac{c}{1 - \alpha}.$$

Exemple 15 Le graphique ci-dessous correspond à la simulation d'un tel processus, avec à droite son autocorrélogramme,



L'autocorrélogramme suggère de tester un modèle autorégressif d'ordre 1 sur Y . Toutefois, si l'on étudie la distribution des résidus du modèle $Y_t = \phi Y_{t-1} + \varepsilon_t$, l'hypothèse de normalité est clairement rejetée

LS // Dependent Variable is Y				
Sample: 2 2000				
Included observations: 1999 after adjusting endpoints				
Convergence achieved after 2 iterations				
Variable	Coefficient	Std. Error	T-Statistic	Prob.
AR(1)	0.794042	0.013613	58.33134	0.0000
R-squared	0.629583	Mean dependent var	-0.087891	
Adjusted R-squared	0.629583	S.D. dependent var	2.508131	
S.E. of regression	1.526497	Akaike info criteron	0.846451	
Sum squared resid	4655.723	Schwartz criteron	0.849252	
Log likelihood	-3681.486	Durbin-Watson stat	2.064924	
Inverted AR Roots	.79			



L'autocorrélogramme ne permet pas de rejeter l'hypothèse de bruit blanc, mais l'autocorrélogramme ne permet de ne mesurer qu'une dépendance linéaire entre ε_t et ε_{t-h} . L'idée peut alors être de tester le caractère ARCH des résidus

obtenus, pour expliquer cette forte kurtosis,

ARCH Test:				
F-statistic	444.8826	Probability	0.000000	
Obs*R-squared	364.1615	Probability	0.000000	
Test Equation:				
LS // Dependent Variable is RESID^2				
Sample: 3 2000				
Included observations: 1998 after adjusting endpoints				
Variable	Coefficient	Std. Error	T-Statistic	Prob.
C	1.335464	0.116004	11.51218	0.0000
RESID^2(-1)	0.426916	0.020240	21.09224	0.0000
R-squared	0.182263	Mean dependent var	2.330146	
Adjusted R-squared	0.181853	S.D. dependent var	5.237595	
S.E. of regression	4.737481	Akaike info criterion	3.112012	
Sum squared resid	44797.68	Schwarz criterion	3.117617	
Log likelihood	-5941.939	F-statistic	444.8826	
Durbin-Watson stat	1.985069	Prob(F-statistic)	0.000000	

Ce test est alors clairement significatif, et l'on valide l'hypothèse de modèle ARCH pour les résidus. Le modèle est alors

$$Y_t = 0.79404 Y_{t-1} + \varepsilon_t \text{ où } \varepsilon_t = \eta_t \sqrt{1.335464 + 0.42691 \varepsilon_{t-1}^2}$$

et où (η_t) est un bruit blanc gaussien.

2.3.2 Erreurs ARCH (p)

On considère ici un modèle plus général, sur le processus (Y_t) , de la forme

$$Y_t = AX_t + \varepsilon_t \text{ où } |\phi| < 1 \text{ et } (\varepsilon_t) \text{ satisfait } \varepsilon_t = h_t \times \eta_t \text{ avec } (\eta_t) \text{ bruit blanc,}$$

et où (X_t) est un vecteur multivarié (*modèle de régression classique : variables exogènes*, ou composé de retards de Y_t (*modèle ARMA*))

$$h_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_p \varepsilon_{t-p}^2 = \alpha_0 + A(L) \varepsilon_t^2,$$

en posant $A(L) = \alpha_1 L + \alpha_2 L^2 + \dots + \alpha_p L^p$.

Nous allons tester ici l'hypothèse $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_p = 0$, contre l'hypothèse alternative où il existe i tel que $\alpha_i \neq 0$. La procédure pour tester cette absence d'effet ARCH est fondée soit sur un test de Fisher, soit sur un test LM du multiplicateur de Lagrange. La mise en place de ces tests se fait de la façon suivante,

- (1) calcul des résidus/erreurs du modèles de régression : $\hat{\varepsilon}_t$,
- (2) calcul du carré des erreurs $\hat{\varepsilon}_t^2$,
- (3) à p fixé, on effectue la régression linéaire de $\hat{\varepsilon}_t^2$ sur son passé $\hat{\varepsilon}_{t-1}^2, \dots, \hat{\varepsilon}_{t-p}^2$:

$$\hat{\varepsilon}_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \hat{\varepsilon}_{t-i}^2,$$

(4) calcul de la statistique LM, $LM = nR^2$ où n est le nombre d'observations utilisées lors de la régression, et R^2 le coefficient de la régression. Si $LM > \chi^2(p)$, on rejette H_0 : l'erreur est alors modélisée par un processus ARCH(p).

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

Comme nous l'avions vu dans la partie (?????), les tests de Dickey-Fuller permet de tester l'hypothèse de présence ou non de racine unité : le test de régression s'écrit alors

$$X_t = \phi X_{t-1} + \eta_t \text{ dont l'estimation est } X_t = \hat{\phi} X_{t-1} + \hat{\eta}_t, \quad (18)$$

(ou, de façon plus générale encore $X_t = \phi X_{t-1} + \alpha + \beta t + \eta_t$). Il est alors possible de montrer que sous l'hypothèse $(H_0) : \phi = 1$, dans le cadre du modèle (18)⁵ la statistique de Student du test s'écrit

$$\hat{t}_{\phi=1} = \frac{\hat{\phi} - 1}{\hat{\sigma}(\hat{\phi})} \text{ où } \hat{\phi} - 1 = \frac{\sum X_{t-1} \eta_t}{\sum X_{t-1}^2}, \quad s^2 = \frac{1}{T-1} \sum [X_t - \hat{\phi} X_{t-1}]^2 \text{ et } \hat{\sigma}(\hat{\phi})^2 = \frac{s^2}{\sum X_{t-1}^2},$$

⁵Dans le cas plus compliqué avec tendance et constante (i.e. $X_t = \phi X_{t-1} + \alpha + \beta t + \eta_t$), la distribution asymptotique est alors un peu plus compliquée : cf article de Stock (1994) dans le Handbook of Econometrics.

avec $\widehat{\sigma}(\widehat{\phi})$ écart type (*par moindre carrés*) de l'estimateur de ϕ , et sa distribution est donnée par

$$\widehat{t}_{\phi=1} = \frac{\widehat{\phi} - 1}{\widehat{\sigma}(\widehat{\phi})} \xrightarrow{\mathcal{L}} \frac{\int_0^1 W_t dW_t}{\left[\int_0^1 W_t^2 dt\right]^{1/2}} \neq \mathcal{N}(0, 1) \text{ où } (W_t) \text{ est un brownien standard sur } [0, 1].$$

Il est possible de montrer que ces tests, bien que développés en théorie sur des modèles de type *AR* restent asymptotiquement robustes en présence d'effet *ARCH* sur les résidus⁶.

2.4 Estimation et prévision

2.4.1 Estimation des paramètres d'un modèles *ARCH*

L'estimation peut se faire en utilisant des techniques inspirées du maximum de vraisemblance. La log-vraisemblance à la date t est de la forme

$$\mathcal{L}_t = \text{constante} - \frac{1}{2} \log h_t^2 - \frac{1}{2} \varepsilon_t^2 h_t^{-2},$$

et la log-vraisemblance totale du modèle s'écrit

$$\mathcal{L} = \text{constante} - \frac{1}{2} \sum_{t=1}^n \log h_t^2 - \frac{1}{2} \sum_{t=1}^n \varepsilon_t^2 h_t^{-2}.$$

La méthode pour estimer les paramètres est alors la suivante

- (1) calcul des résidus/erreurs du modèles de régression : $\widehat{\varepsilon}_t$,
- (2) calcul du carré des erreurs $\widehat{\varepsilon}_t^2$, et on effectue la régression linéaire de $\widehat{\varepsilon}_t^2$ sur son passé $\widehat{\varepsilon}_{t-1}^2, \dots, \widehat{\varepsilon}_{t-p}^2$:

$$\widehat{\varepsilon}_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \widehat{\varepsilon}_{t-i}^2 + u_t,$$

(3) on considère que la variance de l'erreur peut être approchée par h_t^2 , et on estime de nouveau les paramètres du modèle par moindres carrés généralisés, avec comme facteur de pondération $\omega = 1/h_t$ et $= \text{diag}(h_t^2)$: $\widehat{\alpha} = (\widehat{\varepsilon}' \widehat{\varepsilon})^{-1} \widehat{\varepsilon}' h$ et

$$h_t^2 = \widehat{\alpha}_0 + \sum_{i=1}^p \widehat{\alpha}_i \widehat{\varepsilon}_{t-i}^2 + u_t.$$

Par rapport aux modèles linéaire, on a une grande différence quant à la variance de l'erreur du modèle, qui va être fonction de la variance résiduelle $\widehat{\sigma}$, alors que pour les modèles *ARCH*, elle sera fonction de h_t^2 : la variance de l'erreur n'est alors plus constante (*ce qui va influencer, par exemple, les intervalles de confiance lors des prévisions*).

Plus rigoureusement, l'estimation des paramètre dans les modèles *ARCH* se fait en utilisant l'estimateur du pseudo maximum de vraisemblance, sous l'hypothèse de loi conditionnelle normale. Aussi, so $l_t(y, \theta)$ désigne la vraisemblance de Y_t , conditionnelle au passé, la vraisemblance de Y_1, \dots, Y_T , conditionnelle à Y_0 est alors

$$L(y, \theta) = \prod_{t=1}^T l_t(y, \theta) \text{ et on pose } \widehat{\theta} = \arg \max \log L(y, \theta).$$

En fait, cet estimateur est convergent, même si la loi conditionnelle n'est pas normale. Cet estimateur est également asymptotiquement normal

2.4.2 La procédure de Diebold (1987) : test d'autocorrélation en présence d'effet *ARCH*

Considérons une série (X_t) de moyenne nulle. L'autocorrélation empirique d'ordre h est donnée par

$$\widehat{\rho}(h) = \frac{\widehat{\gamma}(h)}{\widehat{\gamma}(0)} \text{ où } \widehat{\gamma}(h) = \frac{1}{T-h} \sum_{t=h+1}^T X_t X_{t-h}.$$

⁶On pourra consulter en particulier Kim et Schmidt (1993) Unit root with conditional heteroskedasticity, dans Journal of Econometrics.

Sous l'hypothèse (H_0) où (X_t) est un bruit blanc, $\widehat{\rho}(h)$ est asymptotiquement normalement distribuée, de moyenne nulle et de variance $V(\widehat{\rho}(h))$ équivalente à $1/T$. Cependant en présence d'hétéroscédasticité, la variance de l'autocorrélation n'est plus en $1/T$ mais est égale à

$$V(\widehat{\rho}_X(h)) = \frac{1}{T} \left[1 + \frac{\gamma_{X^2}(h)}{\sigma^4} \right],$$

où $\gamma_{X^2}(h)$ est l'autocovariance du processus (X_t^2) . Aussi, l'intervalle de confiance de Bartlett ($\pm 1.96/\sqrt{T}$ à 95%) s'avère trop étroit en présence d'effet *ARCH*. Et comme ces deux paramètres, $\gamma_{X^2}(h)$ et σ^2 , sont facilement estimables, une estimation de la variance des autocorrelations peut être obtenue comme

$$\widehat{V}(\widehat{\rho}_X(h)) = \frac{1}{T} \left[1 + \frac{\widehat{\gamma}_{X^2}(h)}{\widehat{\sigma}^4} \right],$$

et l'intervalle de confiance corrigé des effets *ARCH* devient, à 95%, $\pm 1.96\sqrt{\widehat{V}(\widehat{\rho}_X(h))}$. La statistique de Box-Pierce corrigée est alors donnée par

$$Q'(k) = T \sum_{h=1}^k \left[\frac{\sigma^4}{\sigma^2 + \widehat{\gamma}_{X^2}(h)} \right] \widehat{\rho}_X^2(h),$$

qui est asymptotiquement distribuée comme une loi du chi-deux à k degrés de liberté.

2.4.3 Prévision et intervalle de confiance

Nous allons ici étudier la prévision d'une variable, modélisée par un modèle *ARMA* dont les erreurs suivent un bruit blanc ou un modèle *ARCH*. Considérons par un modèle *AR(1)*

$$X_t = \mu + \rho X_{t-1} + \varepsilon_t.$$

Compte tenu de cette écriture, à la date $t+1$, X_{t+1} est donnée par $X_{t+1} = \mu + \rho X_t + \varepsilon_{t+1}$, dont la prévision, faite à la date t est

$$\widehat{X}_t(1) = EL(X_{t+1}|X_t, X_{t-1}, \dots, \varepsilon_t, \varepsilon_{t-1}, \dots) = \mu + \rho X_t + EL(\varepsilon_{t+1}|\varepsilon_t, \varepsilon_{t-1}, \dots).$$

Dans le cas où (ε_t) est un bruit blanc, ε_{t+1} est orthogonal à son passé, et donc $EL(\varepsilon_{t+1}|\varepsilon_t, \varepsilon_{t-1}, \dots) = 0$ et donc $\widehat{X}_t(1) = \mu + \rho X_t$. De façon itérative, on aura, pour tout $h > 0$,

$$\widehat{X}_t(h) = \mu + \rho \widehat{X}_t(h-1) = \mu \frac{1-\rho^h}{1-\rho} + \rho^h X_t,$$

dont l'erreur de prévision est $e_h = X_{t+h} - \widehat{X}_t(h) = \varepsilon_{t+h} + \rho \varepsilon_{t+h-1} + \dots + \rho^{h-1} \varepsilon_{t+1}$. On peut noter que $\widehat{X}_t(h) \rightarrow \mu / (1 - \rho)$ quand $h \rightarrow \infty$.

L'intervalle de prévision, au seuil α , est de la forme $[\underline{X}(h), \overline{X}(h)]$ de telle sorte que

$$\mathbb{P}(e_h \in [\underline{X}(h), \overline{X}(h)]) = \mathbb{P}\left(X_{t+h} \in [\widehat{X}_t(h) + \underline{X}(h), \widehat{X}_t(h) + \overline{X}(h)]\right) = 1 - \alpha.$$

Dans le cas où (ε_t) est un bruit blanc fort gaussien, alors pour un horizon de 1, on peut noter aisément que $e_1 = X_{t+1} - \widehat{X}_t(1) = \varepsilon_{t+1} \sim \mathcal{N}(0, \sigma^2)$, et donc, un intervalle de confiance à 95% pour la valeur en $t+1$ sera

$$[\widehat{X}_t(1) - 1.96\sigma, \widehat{X}_t(1) + 1.96\sigma],$$

et plus généralement, $e_h = X_{t+h} - \widehat{X}_t(h) = \varepsilon_{t+h} + \rho \varepsilon_{t+h-1} + \dots + \rho^{h-1} \varepsilon_{t+1} \sim \mathcal{N}(0, [1 + \rho^2 + \dots + \rho^{2(h-1)}] \sigma^2)$ et donc, un intervalle de confiance à 95% pour la valeur en $t+h$ sera

$$[\widehat{X}_t(h) - 1.96\sigma\sqrt{1 + \rho^2 + \dots + \rho^{2(h-1)}}, \widehat{X}_t(h) + 1.96\sigma\sqrt{1 + \rho^2 + \dots + \rho^{2(h-1)}}].$$

Modèle AR(1) – ARCH(1) Supposons que (ε_t) ne soit plus un bruit blanc (*indépendant de son passé*) mais que le bruit soit ARCH(1), c'est à dire que $\varepsilon_t = \eta_t \sqrt{h_t}$ où $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$, où on supposera que (η_t) est un bruit blanc que l'on supposera de variance unitaire, indépendant du passé de (ε_t) , c'est à dire η_t indépendant de $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots$. On notera tout d'abord que $\mathbb{E}(\varepsilon_t) = 0$ pour tout t , que la corrélation entre deux dates est donnée par la relation

$$\mathbb{E}(\varepsilon_t \varepsilon_s) = \mathbb{E}\left(\eta_t \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2} \cdot \eta_s \sqrt{\alpha_0 + \alpha_1 \varepsilon_{s-1}^2}\right) = \mathbb{E}(\eta_t) \mathbb{E}\left(\sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2} \cdot \eta_s \sqrt{\alpha_0 + \alpha_1 \varepsilon_{s-1}^2}\right) = 0 \text{ pour } s < t,$$

et que la variance à la date t vérifie

$$\sigma^2 = \mathbb{E}(\varepsilon_t^2) = \mathbb{E}(\eta_t^2 [\alpha_0 + \alpha_1 \varepsilon_{t-1}^2]) = \mathbb{E}(\eta_t^2) \mathbb{E}(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2) = \alpha_0 + \alpha_1 \mathbb{E}(\varepsilon_{t-1}^2) = \alpha_0 + \alpha_1 \sigma^2,$$

c'est à dire $\sigma^2 = \alpha_0 / (1 - \alpha_1)$. Pour obtenir un intervalle de confiance pour X_{t+h} , rajoutons une hypothèse supplémentaire de normalité du bruit (η_t). Dans ce cas, l'intervalle de confiance, à 95%, pour la valeur à la date $t+1$ est donné par

$$\left[\widehat{X}_t(1) - 1.96\sigma \sqrt{\alpha_0 + \alpha_1 \varepsilon_t^2}, \widehat{X}_t(1) + 1.96\sigma \sqrt{\alpha_0 + \alpha_1 \varepsilon_t^2} \right].$$

Aussi, dans le cas d'un modèle *ARCH*, la taille de l'intervalle de confiance dépend de la valeur ε_t^2 .

Modèle $AR(1)-GARCH(1,1)$ On considère ici un modèle $AR(1)-GARCH(1,1)$, i.e. un processus (X_t) vérifiant une relation de type $AR(1)$ -sans constante - et (ε_t) suit un modèle $GARCH(1,1)$,

$$\begin{cases} X_t = \rho X_{t-1} + \varepsilon_t \\ \varepsilon_t = \eta_t \sqrt{h_t} \\ h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1} \end{cases} \quad \text{ou} \quad \begin{cases} X_t = \rho X_{t-1} + \varepsilon_t \\ \varepsilon_t = \eta_t \sigma_t \\ \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \end{cases}$$

où $\omega > 0$, $\alpha, \beta \geq 0$ et $\alpha + \beta \leq 1$ et $|\rho| < 1$. Alors, pour tout $h \geq 0$,

$$X_{t+h} = \varepsilon_{t+h} + \rho \varepsilon_{t+h-1} + \dots + \rho^h \varepsilon_t + \rho^{h+1} X_{t-1},$$

de telle sorte que $\mathbb{E}(X_{t+h}|X_{t-1}) = \rho^{h+1} X_{t-1}$ (ou même, plus généralement, $\mathbb{E}(X_{t+h}|\underline{X_{t-1}}) = \rho^{h+1} X_{t-1}$). De plus,

$$V(X_{t+h}|\underline{X_{t-1}}) = \sum_{i=0}^h \rho^{2(h-i)} V(\varepsilon_{t+i}|\underline{\varepsilon_{t-1}}),$$

où $V(\varepsilon_t|\underline{\varepsilon_{t-1}}) = \sigma_t^2$ et pour $i \geq 1$,

$$\begin{aligned} V(\varepsilon_{t+i}|\underline{\varepsilon_{t-1}}) &= \mathbb{E}(\sigma_{t+i}^2|\underline{\varepsilon_{t-1}}) = \omega + (\alpha + \beta) \mathbb{E}(\sigma_{t+i-1}^2|\underline{\varepsilon_{t-1}}) \\ &= \omega \left(1 + \dots + (\alpha + \beta)^{i-1} \right) + (\alpha + \beta)^i \sigma_t^2 = \omega \frac{1 - (\alpha + \beta)^i}{1 - (\alpha + \beta)} + (\alpha + \beta)^i \sigma_t^2, \end{aligned}$$

de telle sorte que

$$\begin{aligned} V(X_{t+h}|\underline{X_{t-1}}) &= \frac{\omega (1 - \rho^{2(h+1)})}{(1 - (\alpha + \beta))(1 - \rho^2)} + \left[\sigma_t^2 - \frac{\omega}{1 - (\alpha + \beta)} \right] \frac{\rho^{2(h+1)} - (\alpha + \beta)^{h+1}}{\rho^2 - (\alpha + \beta)} \text{ si } \rho^2 \neq \alpha + \beta \\ &= \frac{\omega (1 - \rho^{2(h+1)})}{(1 - \rho^2)^2} + \left[\sigma_t^2 - \frac{\omega}{1 - (\alpha + \beta)} \right] (h+1) \rho^{2h} \text{ si } \rho^2 = \alpha + \beta. \end{aligned}$$

On peut d'ailleurs noter que si $h \rightarrow \infty$, on retrouve la variance non conditionnelle de X_t , i.e.

$$\lim_{h \rightarrow \infty} V(X_{t+h}|\underline{X_{t-1}}) = V(X_t) = \frac{V(\varepsilon_t)}{1 - \rho^2}.$$

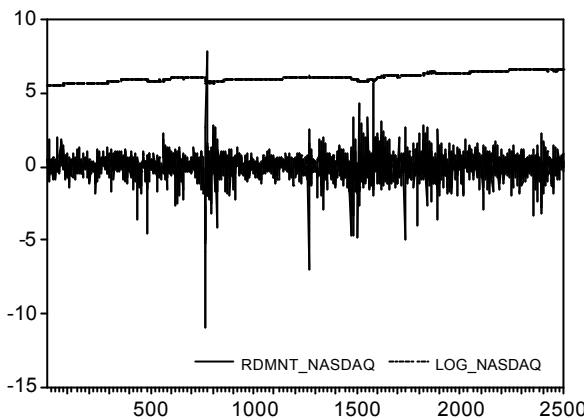
Exemple 16 Les graphiques ci-dessous donnent les intervalles de confiance à 95%, à horizon 1, 2 et 5 respectivement, pour des processus *GARCH(1,1)* simulé, où $\omega = 1$, $\alpha = 0.1$ et $\beta = 0.8$ (avec un bruit η_t gaussien, $\eta_t \sim \mathcal{N}(0, 1)$)

Les graphiques ci-dessous correspondent au cas $\omega = 1$, $\alpha = 0.6$ et $\beta = 0.2$ (avec un bruit η_t gaussien, $\eta_t \sim \mathcal{N}(0, 1)$)

Enfin, pour rappels, les graphiques ci-dessous correspondent à des intervalles de confiance associés à des bruits blancs, pour les mêmes horizons

2.5 Modèles *ARCH* et finance

La série temporelle ci-dessous représente l'indice *NASDAQ*, avec le rendement de l'indice au jour le jour, et le logarithme de l'indice (*en haut*). De même que pour la série du *CAC40*, on retrouve là aussi des périodes de forte volatilité sur le marché, comportement qui ne peut être pris en compte à l'aide de modèles *ARMA*,



Notons (p_t) le prix d'un titre à la date t , et (r_t) le logarithme du rendement, i.e. $r_t = \log(p_t) - \log(p_{t-1}) = \log(1 + R_t)$ où R_t correspond à la variation relative des prix, i.e. $R_t = [R_t - R_{t-1}] / R_t$. Ces deux séries sont d'ailleurs sans unité, ce qui facilite la comparaison entre elles. Les propriétés suivantes ont été notées sur la plupart des séries financières,

(1) Les processus (p_t) sont non-stationnaires : les trajectoires de prix sont généralement proches d'une marche aléatoire sans terme constant. Et, en revanche, les séries des rendements ont des trajectoires compatibles avec la stationnarité au second ordre.

(2) Autocorrélations des carrés des variations des prix : les faibles autocorrélations de la série (r_t) la rendent proche d'un bruit blanc⁷ (*sauf pour des rendements définis sur des périodes très courtes, de l'ordre d'une vingtaine de minutes*) Néanmoins, la série (r_t^2) est souvent fortement corrélée. Ce qui n'est pas compatible avec une hypothèse de bruit blanc.

(3) Queue de distribution épaisse : les distributions empiriques des séries des rendements, on s'aperçoit que l'hypothèse de normalité est rejetée. En particulier, au niveau des queues, la décroissance est généralement plus faible que dans le cas gaussien, i.e. en $\exp(-x^2/2)$. On parle alors de distribution **leptokurtique**.

(4) **Volatility clustering** (*regroupement des extrêmes*) : empiriquement, de fortes valeurs, ou de fortes variations, tendent à être suivies par d'autres grandes variations.

(5) Queue épaisse conditionnelles : souvent, même une fois corrigées de la volatility clustering (*en utilisant des modèles de type GARCH, par exemple*), les séries ont des résidus possédant encore des queues épaisse (*même si cela est moins important que la distribution non conditionnelle, i.e. la distribution des rendements*).

(6) **Effet de levier** : cette propriété, notée par Brock en 1976 repose sur l'observation du fait qu'il existe une asymétrie entre l'impact des valeurs passées positives et des valeurs passées négatives. Les baisses de cours tendent à provoquer un accroissement de la volatilité supérieure à celui induit par une hausse du cours, de même amplitude.

(7) **Saisonnalité** : la volatilité tend à augmenter lorsque les marchés ferment (*week end ou jours fériés*). On parle alors d'accumulation d'information.

(8) **Asymétrie perte/gain** : ormis dans le cas des taux de change, il y a généralement asymétrie de la distribution : il y a plus de mouvements forts à la baisse qu'à la hausse.

Toutes ces propriétés sont difficiles à obtenir à l'aide de modèles *ARMA* classiques.

Les modèles *ARCH* sont apparus à une époque où se développaient les modèles en temps continu, en particulier pour la valorisation de produits dérivés. Mais si cette valorisation a donné de bons résultats théoriques, l'estimation, à partir de données discrètes s'est avérée plus difficile : les modèles les plus simples (*brownien géométrique, modèle d'Ornstein-Uhlenbeck ou modèle de Cox-Ingersoll-Ross*) étaient généralement très mal spécifiés. Les modèles *ARCH* en revanche se prêtaient bien à l'estimation, et fournissaient de bons ajustements, même si la valorisation des produits dérivés a posé quelques problèmes (*le temps discret impliquant une incomplétude du marché*). Toutefois, ces modèles

⁷L'hypothèse usuelle en théorie de la finance consistait effectivement à supposer que les processus de rendements sont i.i.d. et de variance finie. Cette hypothèse d'indépendance n'est souvent retenue qu'en première approximation. Un hypothèse plus forte est aussi souvent faite sur le caractère gaussien de ces rendements. Ces hypothèses se traduisent, dans le cas continu, à des modèles de la forme $dR_t = \mu dt + \sigma dW_t$ où (W_t) est un mouvement brownien standard.

avaient quelques faiblesses, d'où le grands nombre de modèles généralisant encore davantage ces modèles : *IGARCH*, *TARCH*, *QARCH*, *SWARCH*, *ACD – GARCH*...etc.

2.5.1 Liens entre temps continu et temps discret

Il est possible de passer du temps continu en temps discret en considérant des '*petites variations*' de temps. Ce passage est en particulier utile si l'on souhaite faire de l'estimation de modèle. Soit (X_t) un processus d'Ito en temps continu, i.e.

$$dX_t = \mu(X_t, t) dt + \sigma(X_t, t) dW_t \text{ où } (W_t) \text{ est un brownien standard,} \quad (19)$$

et soit $g(x, t)$ une application différentiable en x et en t , alors, si $Y_t = g(X_t, t)$, le lemme d'Ito permet d'écrire

$$dY_t = \left(\frac{\partial g}{\partial x} \mu(X_t, t) + \frac{\partial g}{\partial t} + \frac{1}{2} \frac{\partial^2 g}{\partial x^2} \sigma^2(X_t, t) \right) dt + \frac{\partial g}{\partial x} \sigma(X_t, t) dW_t.$$

La version discrète d'un processus d'Ito est

$$\Delta X_t = X_{t+1} - X_t = \mu(X_t, t) + \sigma(X_t, t) \varepsilon_t.$$

Supposons que (X_t) suit un brownien géométrique, i.e. μ et σ sont des constantes dans (19), où (X_t) correspond, par exemple, au prix d'une action , où à un niveau d'indice. Soit alors (Y_t) le log-rendement associé, i.e. $Y_t = \log X_t - \log X_{t-1}$ pour $t = 1, \dots, T$ (*en supposant* (X_t) *observé à la date 0*), soit $X_t = X_{t-1} \exp(Y_t)$. Nous noterons Δ le temps entre les deux observations, exprimé en année⁸. Puisque (X_t) suit un brownien géométrique, alors à chaque date, Y_t est distribué suivant une loi gaussienne, de moyenne $(\mu - \sigma^2/2)\Delta$ et de variance $\sigma^2\Delta$, et $\text{cov}(Y_t, Y_{t+h}) = 0$ pour $h \neq 0$. Notons $\mu_Y = (\mu - \sigma^2/2)\Delta$ et $\sigma_Y^2 = \sigma^2\Delta$, \bar{Y} et S_Y les estimateurs empiriques de μ_Y et de σ_Y , i.e.

$$\bar{Y} = \frac{1}{T} \sum_{i=1}^T Y_i \text{ et } S_Y = \sqrt{\frac{1}{T-1} \sum_{i=1}^T (Y_i - \bar{Y})^2}$$

(qui sont alors des estimateurs consistants de μ_Y et de σ). On peut alors estimer μ et σ par

$$\hat{\mu} = \frac{\bar{Y}}{\Delta} + \frac{S_Y^2}{2\Delta} \text{ et } \hat{\sigma} = \frac{S_Y}{\sqrt{\Delta}}$$

On peut également noter que les rendements sont supposés gaussiens, et en particulier, on aura alors, pour $h > 0$

$$\log X_{t+h} - \log X_t \sim \mathcal{N} \left(\left[\mu - \frac{\sigma^2}{2} \right] h, \sigma^2 h \right).$$

Exemple 17 Considérons le rendement de l'action IBM au cours de l'année 1998 [A INSERER]

L'autocorrélogramme montre que la série des log-rendements est non-autocorrélées (cf statistique de Ljung Box par exemple). Supposons que le prix de l'action soit modélisé par un brownien géométrique, i.e. $dX_t = \mu dt + \sigma dW_t$, alors, en utilisant le fait que $\bar{Y} = 0.002276$ et $S_Y = 0.01915$ (cf statistiques ci-dessous), les paramètres μ et σ peuvent être estimés par

$$\hat{\mu} = \frac{\bar{Y}}{\Delta} + \frac{S_Y^2}{2\Delta} \approx 0.6198 \text{ et } \hat{\sigma} = \frac{S_Y}{\sqrt{\Delta}} \approx 0.3040$$

C'est à dire que le rendement espéré annuel est de 61.98%, avec une volatilité de 30.4%, en 1998. On peut toutefois noter que l'hypothèse de normalité n'est pas valide.

⁸On prendra $\Delta = 1/12$ pour des données mensuelles, ou $\Delta = 1/250$ pour des données journalières (si l'on ne prend en compte que les jours ouvrés).

2.5.2 Modèles avec variance stochastique/déterministe en temps continu

Dans le modèle initial proposé par Black et Scholes, le processus de prix des actifs (*risqués*) est modélisé par un processus d'Ito, i.e.

$$dS_t = \mu S_t dt + \sigma S_t dB_t \text{ où } B_t \text{ est un mouvement brownien standard,}$$

et où μ et σ sont des constantes. Toutefois, cette hypothèse de constance de σ s'est vite avérée, en pratique, inexacte⁹.

Une alternative a été de supposer que cette volatilité pouvait être, elle même, stochastique, dont la dynamique pourrait également être modélisée par un processus stochastique,

$$\begin{cases} dS_t = S_t [\mu_t dt + \sigma_t dW_t] \\ d\sigma_t = b(t, \sigma_t) dt + \Sigma(t, \sigma_t) dB_t. \end{cases}$$

2.5.3 Modèles avec variance stochastique/déterministe en temps discret

Les modèles introduits dans la littérature économique se présentaient généralement sous forme multiplicative, $\varepsilon_t = \sigma_t \eta_t$ où (η_t) est un processus i.i.d. et (σ_t) une suite de variables aléatoires telles que

- σ_t soit mesurable par rapport à la filtration engendrée par le passé de ε_t - et éventuellement le présent - et par le passé d'un processus sous-jacent inobservables (v_t),

- σ_t et η_t soient des variables indépendantes,

- $\sigma_t > 0$ pour tout t .

La variable aléatoire σ_t est appelée **volatilité** de ε_t . Il est possible de noter que $\mathbb{E}(\varepsilon_t) = \mathbb{E}(\sigma_t \eta_t) = 0$ et $\text{cov}(\varepsilon_t, \varepsilon_{t-h}) = 0$ pour tout $h > 0$: (ε_t) est alors un bruit blanc, au sens faible. Mais pas au sens fort. Plusieurs types de modèles existent alors suivant la spécification retenue pour (σ_t)

(1) les processus **conditionnellement hétérosclélastiques**, de type *GARCH* pour lesquels la filtration \mathcal{F}_t est engendrée par le passé de ε_t . La volatilité est alors une fonction déterministe du passé de ε_t . Les modèles *GARCH* standard sont caractérisés par une volatilité fonction affine des valeurs passées de ε_t^2 .

(2) les processus à **volatilité stochastique**, pour lesquels la filtration \mathcal{F}_t est engendrée par le passé de v_t . La volatilité est alors un processus. Le modèle le plus simple et le plus utilisé consiste à supposer que le processus $(\log \sigma_t)$ suit un *AR(1)*, i.e. $\log \sigma_t = \omega + \rho \log \sigma_{t-1} + v_t$.

(3) les processus à **changement de régime**, pour lesquels on suppose alors généralement que le processus sous-jacent (v_t) est modélisable par une chaîne de Markov à espace d'état fini.

Exemple 18 ARCH - L'idée est ici de modéliser la volatilité à l'aide d'un processus *ARCH*. Si l'on note S_i le prix d'un actif à la date i , où $i = 0, 1, \dots, n$, et u_i le log-rendement associé, $u_i = \log(S_i) - \log(S_{i-1})$ pour $i = 1, \dots, n$, on suppose que l'on peut écrire la volatilité σ_n^2 sous la forme

$$\sigma_n^2 = \alpha_1 u_{n-1}^2 + \alpha_2 u_{n-2}^2 + \dots + \alpha_{n-1} u_1^2 = \sum_{i=1}^{n-1} \alpha_i u_{n-i}^2,$$

où les variables α_i correspondent au poids donné à la $n - i$ ème observation, de telle sorte que leur somme fasse 1. On obtient ainsi un modèle *ARCH(n)*. Un cas un peu plus général consiste à considérer un modèle avec constante, $\sigma_n^2 = \alpha_0 V + \alpha_1 u_{n-1}^2 + \alpha_2 u_{n-2}^2 + \dots + \alpha_{n-1} u_1^2$ où V correspond à une moyenne, à long terme de la volatilité.

Exemple 19 EWMA (Exponentially Weighted Moving Average) - Un cas particulier du cas précédent consiste à prendre des poids décroissantes de façon exponentielle, i.e. $\alpha_i = \lambda \alpha_{i-1}$ où $0 < \lambda < 1$, de telle sorte que l'équation ci-dessus se réécrit

$$\sigma_n^2 = \lambda \sigma_{n-1}^2 + (1 - \lambda) u_{n-1}^2 = (1 - \lambda) \sum_{i=1}^{n-1} \lambda^{i-1} u_{n-i}^2 + \lambda^n \sigma_0^2.$$

Exemple 20 GARCH - Dans le cas des modèles *GARCH(1,1)*, on considère des modèles de la forme

$$\sigma_n^2 = \kappa + \beta \sigma_{n-1}^2 + \alpha u_{n-1}^2 \text{ où } \kappa = \gamma V.$$

Dans le cas particulier où $\kappa = 0$, on retrouve le modèle *EWMA*.

⁹Il est possible, connaissant le prix de marché d'un call de déduire une valeur unique pour la volatilité, dans le cas du modèle de Black et Scholes. Cette valeur de volatilité, appelée volatilité implicite, peut s'écarte notablement de la volatilité historique (*écart type des rendements du sous-jacent*) car elle est censé refléter la volatilité future anticipée par le marché. En particulier, en calculant, à partir des prix de calls de même maturité, et de prix d'exercice différents, les valeurs de cette volatilité implicite. On obtient une courbe convexe assez caractéristique que l'on appelle le smile de volatilité.

2.6 Autres types de modèles non-linéaires

2.6.1 Les modèles bilinéaires - notés $BL(p, q, P, Q)$

Les modèles bilinéaires ont été introduits en 1978 par Granger et Andersen, de façon à généraliser les modèles $ARMA$,

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{i=1}^P \sum_{j=1}^Q \lambda_{ij} Y_{t-i} \varepsilon_{t-j},$$

où (ε_t) est un bruit blanc gaussien. Ce modèle sera noté $BL(p, q, P, Q)$. Une sous classe possible (*diagonal*) est définie par

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{i=1}^P \lambda_i Y_{t-i} \varepsilon_{t-i}.$$

Exemple 21 Le modèle $BL(0, 0, 2, 1)$ est donné par

$$Y_t = \varepsilon_t + \lambda Y_{t-2} \varepsilon_{t-1}.$$

Ce processus est centré, puisque le bruit est indépendant du passé (et donc $\text{cov}(\varepsilon_{t-1}, Y_{t-2})$),

$$\mathbb{E}(Y_t) = \mathbb{E}(\varepsilon_t) + \lambda \mathbb{E}(\varepsilon_{t-1} Y_{t-2}).$$

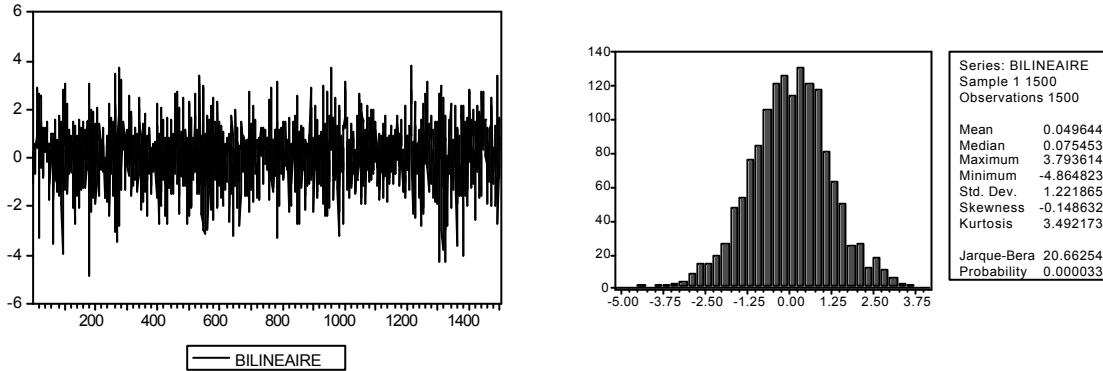
Le processus d'autocovariance est donné par

$$\mathbb{E}(Y_t Y_{t-h}) = \begin{cases} 0 & \text{pour } h \geq 1 \\ \sigma^2 + \lambda^2 \sigma^2 \mathbb{E}(Y_{t-2}^2) & \end{cases}$$

La variance marginale est alors $V(Y_t) = \sigma^2 / [1 - \lambda^2 \sigma^2]$. Il est alors possible de montrer que Y_t s'écrit aussi

$$Y_t = \varepsilon_t + \sum_{j=1}^{\infty} \left[\lambda^j \varepsilon_{t-2j} \cdot \prod_{k=1}^j \varepsilon_{t-2k+1} \right].$$

La série ci-dessous correspond à une simulation de la série $Y_t = \varepsilon_t + 0.6 Y_{t-2} \varepsilon_{t-1}$



2.6.2 Les modèles autorégressifs exponentiels - EXPAR

Ces modèles sont une généralisation des processus AR , permettant de prendre en compte les explosions en variance de la série :

$$X_t = \mu + \sum_{i=1}^p [a_i + \pi_i \cdot \exp(-\gamma X_{t-1}^2)] X_{t-i} + \varepsilon_t.$$

2.6.3 Les modèles autorégressifs à seuils - *TAR*, *STAR* ou *SETAR*

Les modèles autorégressifs à seuils ont été introduits par Tong en 1978, sous le nom de *threshold autoregressive*, ou *TAR*. On suppose, dans le cas d'un seuil unique, qu'il existe deux régimes différents,

$$Y_t = \begin{cases} \phi_1^{(1)} X_{t-1} + \phi_2^{(1)} X_{t-2} + \dots + \phi_{p_1}^{(1)} X_{t-p_1} + \varepsilon_t & \text{si } X_t \leq s \\ \phi_1^{(2)} X_{t-1} + \phi_2^{(2)} X_{t-2} + \dots + \phi_{p_2}^{(2)} X_{t-p_2} + \eta_t & \text{si } X_t > s, \end{cases}$$

où (X_t) est soit une variable exogène, soit une variable (Y_t) retardée (Y_{t-d}) . Dans ce dernier cas, on parlera éventuellement de modèle *SETAR*, *self excited threshold AR*. Il est d'ailleurs possible de généraliser davantage en considérant des sous-modèles *ARMA* au lieu de modèles *AR*. Il est à noter que les bruit (ε_t) et (η_t) sont indépendants, et peuvent être de variance différente.

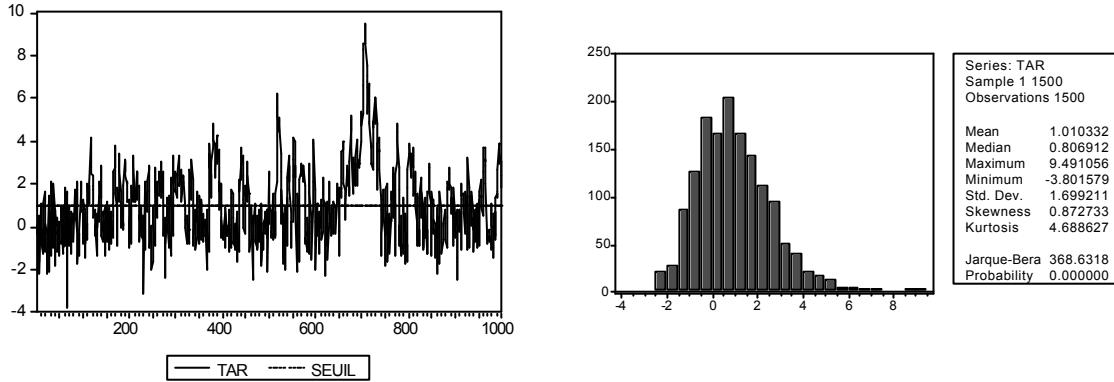
Exemple 22 Considérons le cas de modèles *AR*(1) avec un seuil unique

$$X_t = \begin{cases} \phi^1 X_{t-1} + \varepsilon_t & \text{si } X_{t-1} \leq s \\ \phi^2 X_{t-1} + \varepsilon_t & \text{si } X_{t-1} > s, \end{cases}$$

avec le même bruit. Une condition nécessaire et suffisante d'existence d'une solution stationnaire et $\phi^1 < 1$, $\phi^2 < 1$ et $\phi^1 \phi^2 < 1$. La série ci-dessous correspond à une simulation de la série

$$X_t = \begin{cases} -0.2 X_{t-1} + \varepsilon_t & \text{si } X_{t-1} \leq 1 \\ 0.9 X_{t-1} + \varepsilon_t & \text{si } X_{t-1} > 1, \end{cases}$$

avec un bruit blanc gaussien, centré réduit,



Ce type de processus, là aussi, permet d'avoir des queues de distribution plus épaisses (en l'occurrence ici pour les fortes valeurs de Y_t - queue à droite).

Une écriture équivalente du modèle à seuil à deux régimes, avec un seul retard, ou une seule variable exogène (X_t ou Y_{t-1}), est la suivante

$$Y_t = \begin{cases} \alpha_1 + \phi_1 X_t + \varepsilon_t & \text{si } X_t \leq s \\ \alpha_2 + \phi_2 X_t + \eta_t & \text{si } X_t > s, \end{cases} \quad \Downarrow \quad Y_t = (\alpha_1 + \phi_1 X_t) \mathbb{I}_{X_t \leq s} + (\alpha_2 + \phi_2 X_t) \mathbb{I}_{X_t > s} + u_t, \quad (20)$$

où (u_t) est une séquence de bruits indépendants, dont la variance est de la forme $V(u_t) = \sigma_\varepsilon^2 \mathbb{I}_{X_t \leq s} + \sigma_\eta^2 \mathbb{I}_{X_t > s}$. Il sera possible de se reporter à l'article de Ben Salem et Perraudin (2001) 'Tests de linéarité, spécification et estimation des modèles à seuil : une analyse comparée des méthodes de Tsay et Hansen' pour des compléments d'information sur le sujet.

Test de linéarité Afin d'être sûr qu'il est nécessaire d'utiliser un modèle *TAR*, il est possible d'utiliser des tests de linéarité, dont l'idée est de tester $H_0 : \alpha_1 = \alpha_2$ et $\phi_1 = \phi_2$ contre l'hypothèse alternative fournie par le modèle à seuil. Plus généralement, on teste $H_0 : \alpha_1 = \alpha_2$ et $\phi_i^{(1)} = \phi_i^{(2)}$ pour $i = 1, \dots, p$.

Tsay (1989) propose la statistique suivante, où m est le nombre d'états, et $Y_t = \phi_0^{(i)} + \phi_1^{(i)} Y_{t-1} + \phi_2^{(i)} Y_{t-2} + \dots + \phi_p^{(i)} Y_{t-p_1} + \varepsilon_t^{(i)}$ si $X_t \in$ état i ,

$$Q(m) = \left(\frac{1}{m+1} \sum \hat{\varepsilon}_t^2 - \sum \hat{a}_t^2 \right) \left(\frac{1}{n-b-2m-1} \sum \hat{a}_t^2 \right)^{-1},$$

où b désigne le nombre d'observation utilisées pour initialiser les estimations récursives, i.e. $b = n/10 + p$ (comme le conseille Tsay), $\hat{\varepsilon}_t$ correspondent aux résidus ε_t standardisés, et \hat{a}_t les résidus de la régression de $\hat{\varepsilon}_{(i)}$ sur $(1, X'_{(i)})$. Sous l'hypothèse nulle de linéarité, $Q(m)$ suit une loi de Fisher à $(m+1)$ et $(n-b-2m-1)$ degrés de liberté.

Estimation des paramètres Pour un nombre de retard p donné, et pour un seuil s fixé, les estimateurs par moindres carrés ordinaires des coefficients autorégressifs sont asymptotiquement normaux comme l'ont montré Chan et Tong en 1986. Pour chacun des régimes, on minimise la somme des carrés des résidus, i.e. dans le cas du modèle (20) par exemple

$$(\alpha_i, \phi_i) = \arg \min SCR_i \text{ où } SCR_i = \sum_{t, X_t \in \text{état } i} (Y_t - \alpha_i - \phi_i X_t)^2,$$

et l'estimateur de la matrice de variance covariance des coefficients estimés est alors donnée par $\widehat{V}_i = \widehat{\sigma}_i^2 (X'X)^{-1}$ où $X = (1, X'_t)_{t, X_t \in \text{état } i}$ et

$$\widehat{\sigma}_i^2 = \frac{SCR_i}{n_i - (m+1)} \text{ où } m \text{ est le nombre d'état (ici 2)},$$

et n_i est le cardinal de $\{t, X_t \in \text{état } i\}$.

Détection et choix du seuil Hansen a proposé, en 1996 d'utiliser des moindres carrés séquentiels pour estimer le seuil. Nous allons nous placer ici dans le cas d'un seuil unique s (*modèle à deux états*). Pour une valeur du seuil donnée, on estime par moindres carrés, pour chacun des régimes, les coefficients $\phi^{(i)}$, pour $i = 1, 2$. On évalue alors la variance résiduelle du modèle à seuil, conditionnellement à cette valeur du seuil : $\widehat{\sigma}_T^2(s) = \frac{1}{T} \sum_{t=1}^T (\hat{\varepsilon}_t(s))^2$ où

$$\hat{\varepsilon}_t(s) = Y_t - [(\alpha_1 + \phi_1 X_t) \mathbb{I}_{X_t \leq s} + (\alpha_2 + \phi_2 X_t) \mathbb{I}_{X_t > s}].$$

Ces estimations sont alors faites pour différents seuils s , et on choisit alors le seuil qui minimise la variance résiduelle,

$$\hat{s} = \arg \min_s \widehat{\sigma}_T^2(s).$$

Un intervalle de confiance peut également être obtenu pour s en utilisant la distribution asymptotique de la statistique du rapport de vraisemblance, et fait intervenir un coefficient tabulé par Hansen.

2.6.4 Les généralisations des modèles ARCH

Considérons un modèle linéaire usuel, de la forme

$$Y_t = X'_t \beta + U_t.$$

L'hypothèse faite en économétrie est que (U_t) est un bruit blanc, c'est à dire un processus *i.i.d.* Toutefois, il est possible que le bruit (U_t) soit autocorrélé, c'est à dire que (U_t) puisse s'écrire $U_t = \rho U_{t-1} + \varepsilon_t$ où (ε_t) est ici *i.i.d.*. Comme nous l'avons vu dans ce chapitre, il est possible que le bruit (U_t) soit conditionnellement hétéroscléastique, par exemple, modélisable par un processus *ARCH* (1) défini sous la forme

$$U_t = \eta_t \sqrt{h_t} \text{ où } h_t = \mu + \alpha U_{t-1}^2.$$

Ce processus peut se généraliser sous la forme *GARCH* (1, 1) suivante

$$U_t = \eta_t \sqrt{h_t} \text{ où } h_t = \mu + \alpha U_{t-1}^2 + \delta h_{t-1}.$$

Parmi les autres modèles possibles, on notera les modèles *ARCH-M* (*ARCH en moyenne*, ou *ARCH in Mean*), c'est à dire que le modèle linéaire s'écrira ici

$$Y_t = X'_t \beta + \gamma g(h_t) + U_t \text{ où } U_t = \eta_t \sqrt{h_t} \text{ avec } h_t = \mu + \alpha U_{t-1}^2 + \delta h_{t-1}.$$

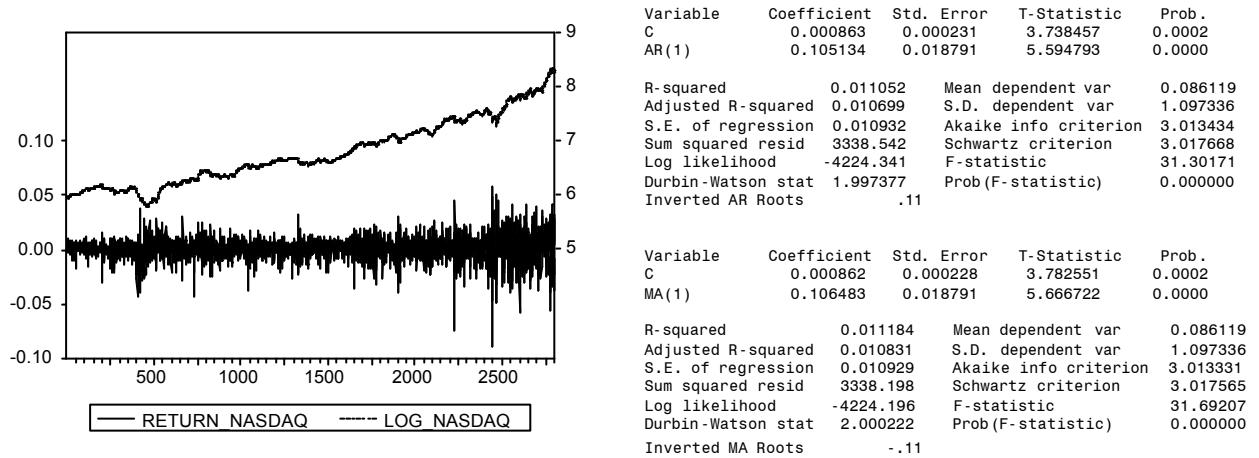
Ce type de modèle a été introduit en 1987 par Engle, Lilien et Robbins. Parmi les modèles asymétriques, on notera les modèles *TARCH* ou *EGARCH*. Les modèles *TARCH* (*ARCH à seuil*, ou *Threshold ARCH*) ont été introduits par Zakoian en 1994. Par exemple, le modèle *TARCH* (1, 1) s'écrit

$$U_t = \eta_t \sqrt{h_t} \text{ où } h_t = \mu + \alpha U_{t-1}^2 + \phi U_{t-1}^2 \mathbb{I}_{u_{t-1} < 0} + \delta h_{t-1}.$$

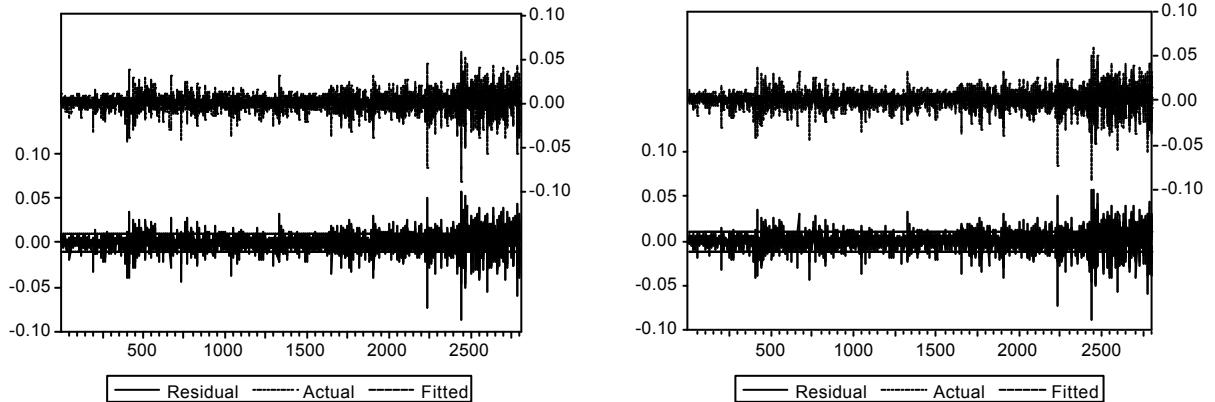
C'est à dire que l'impact de U_{t-1}^2 sera de α si $U_t > 0$, alors qu'il sera de $\alpha + \phi$ sinon. Le modèle *EGARCH* (*GARCH exponentiel*, ou *Exponential GARCH*), introduit par Nelson en 1991, peut s'écrire, dans le cas du modèle *EGARCH* (1, 1)

$$U_t = \frac{\eta_t}{\sqrt{h_t}} \text{ où } \log h_t = \mu + \alpha |U_{t-1}| + \delta \log h_{t-1} + \phi U_{t-1}.$$

Exemple 23 Considérons ici le rendement de l'indice NASDAQ, de janvier 1989 à février 2000. Le graphique ci-dessous représente le logarithme de l'évolution de l'indice NASDAQ ainsi que le rendement journalier



avec, à droite, les résultats d'un ajustement respectivement par un modèle *AR* (1) et *MA* (1) de la série des rendements. Une étude des résidus (avec ci-dessous l'analyse des résidus des modèles autorégressif *AR* (1) à gauche et moyenne mobile *MA* (1) à droite)

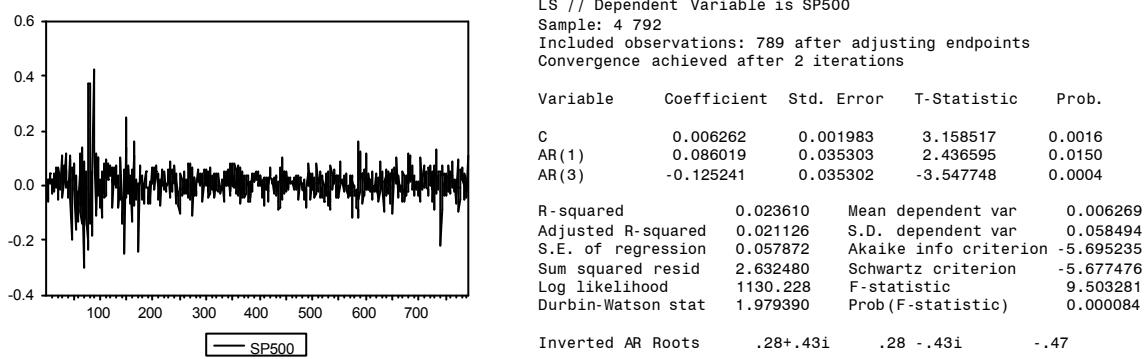


montre que l'hypothèse de bruit blanc est rejetée, mais surtout, les graphiques ci-dessus montrent des périodes de forte volatilité de la série des résidus : il semble nécessaire de modéliser l'erreur par un modèle non linéaire. Les sorties présentées ci-dessous montrent les ajustements de modèles *MA* (1) – *GARCH* (1, 1) (c'est à dire un modèle *MA* (1)

dont les erreurs suivent un modèle $GARCH(1,1)$ et $ARCH - M$, respectivement à gauche et à droite,

Variable	Coefficient	Std. Error	T-Statistic	Prob.	Variable	Coefficient	Std. Error	T-Statistic	Prob.
C	0.084907	0.017700	4.797124	0.0000	SQR(GARCH)	0.119204	0.022944	5.195479	0.0000
MA(1)	0.171620	0.020952	8.190983	0.0000	MA(1)	0.174104	0.020771	8.382098	0.0000
VARIANCE EQUATION									
C	0.027892	0.009213	4.797124	0.0000	C	0.031291	0.009545	3.278211	0.0000
ARCH(1)	0.121770	0.020448	5.955103	0.0000	ARCH(1)	0.133713	0.021011	6.363810	0.0000
GARCH(1)	0.857095	0.021526	39.81666	0.0000	GARCH(1)	0.843131	0.021785	38.70279	0.0000
R-squared	0.007104	Mean dependent var		0.086119	R-squared	0.010861	Mean dependent var		0.086119
Adjusted R-squared	0.005685	S.D. dependent var	1.097336		Adjusted R-squared	0.009448	S.D. dependent var	1.097336	
S.E. of regression	1.094213	Akaike info criterion	2.695856		S.E. of regression	1.092141	Akaike info criterion	2.694578	
Sum squared resid	3352.444	Schwartz criterion	2.706443		Sum squared resid	3339.759	Schwartz criterion	2.705165	
Log likelihood	-3775.938	F-statistic	5.008069		Log likelihood	-3339.759			
Durbin-Watson stat	2.129878	Prob(F-statistic)	0.000000		Durbin-Watson stat	2.132844			
Inverted AR Roots					Inverted AR Roots				
									- .17

Exemple 24 Considérons ici la série du rendement de l'indice SP500, représenté ci-dessous. La sortie présentée à droite est l'ajustement d'un modèle $AR(3)$,



Les sorties ci-après correspondent à l'ajustement de modèles $AR(3) - GARCH(1,1)$, $GARCH(1,1)$, $IGARCH(1,1)$, $GARCH(1,1) - M$,

2.7 Les tests de linéarité

Avant de construire un modèle non-linéaire, il est recommandé de vérifier qu'un modèle linéaire ne suffit à modéliser correctement la série. Il peut arriver (*surtout si les séries temporelles sont courtes*) que l'on estime - avec succès - un modèle non-linéaire alors que la vraie relation sous-jacente est linéaire. Le danger est alors de compliquer inutilement la construction du modèle. Tester la linéarité peut alors aider à ne pas compliquer à outrance, mais cela peut, en outre, aider pour la spécification du type de non-linéarité. Deux types de tests sont alors utilisés :

- les tests contre un modèle non-linéaire spécifique (*comme les tests du multiplicateur de Lagrange et le test du Cusum*)
- les tests sans alternative spécifique (*tels que le BDS ou le Reset*)

2.7.1 Les tests du multiplicateur de Lagrange

Puisque l'estimation de modèles non-linéaires est en général plus difficile que celle des modèles linéaires, il est naturel de considérer des tests qui, bien qu'avec des alternatives non-linéaires spécifiques, ne requièrent pas l'estimation de ces alternatives.

C'est à cette catégorie qu'appartiennent les tests du multiplicateur de Lagrange. Nous disposons ainsi de quatre tests de ce type, chacun testant un type de non-linéarité: $ARCH$, BL , $EXPAR$ et $SETAR$.

$AR(p)$ contre $AR(p)$ à erreurs $ARCH(q)$ Dans ce type de tests, on considère un processus (ε_t) bruit blanc gaussien, c'est à dire *i.i.d.* dont la loi est $\mathcal{N}(0, \sigma^2)$. Les hypothèses sont alors

$$(H_0) : X_t = \mu + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t \text{ contre } (H_1) : X_t = \mu + \sum_{i=1}^p \phi_i X_{t-i} + \sqrt{\alpha_0 + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}^2}.$$

On estime le modèle (H_0) par moindres carrés, et on calcule les résidus $\hat{\varepsilon}_t$ obtenus. On estime alors par les moindres carrés la régression

$$\hat{\varepsilon}_t^2 = \alpha_0 + \sum_{j=1}^q \alpha_j \hat{\varepsilon}_{t-j}^2 + \eta_t \text{ avec } R^2 = 1 - \frac{\sum \eta_t^2}{\sum (\hat{\varepsilon}_t - \bar{\varepsilon})^2}.$$

La statistique du multiplicateur de Lagrange est alors asymptotiquement équivalente à TR^2 . Si l'on pose $LM_0 = TR^2$ alors, sous (H_0) , LM_0 est asymptotiquement distribuée comme un chi-deux à q degrés de libertés.

AR(p) contre BL($p, 0, P, Q$) On considère un processus (ε_t) bruit blanc gaussien, c'est à dire *i.i.d.* dont la loi est $\mathcal{N}(0, \sigma^2)$. Les hypothèses sont alors

$$(H_0) : X_t = \mu + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t \text{ contre } (H_1) : X_t = \mu + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t + \sum_{i=1}^P \sum_{j=1}^Q \theta_{i,j} X_{t-i} \varepsilon_{t-j}.$$

Là encore, on estime le modèle (H_0) par moindre carré, ainsi que $\hat{\sigma}^2$, estimateur de σ^2 . La statistique du test du multiplicateur de Lagrange est

$$LM_1 = \hat{\sigma}^{-2} \left[\sum \hat{z}_{1,t} \hat{\varepsilon}_t \right]' \left[\widehat{M}_{11} - \widehat{M}_{10} \widehat{M}_{00}^{-1} \widehat{M}_{01} \right] \left[\sum \hat{z}_{1,t} \hat{\varepsilon}_t \right],$$

où

$$\begin{cases} \hat{z}_{0,t} = (-1, X_{t-1}, \dots, X_{t-p})' \\ \hat{z}_{1,t} = -(\hat{\varepsilon}_{t-1} X_{t-1}, \dots, \hat{\varepsilon}_{t-P} X_{t-P}, \hat{\varepsilon}_{t-2} X_{t-1}, \dots, \hat{\varepsilon}_{t-Q} X_{t-1}, \dots, \hat{\varepsilon}_{t-Q} X_{t-P}) \end{cases},$$

et

$$\begin{aligned} \widehat{M}_{ii} &= \sum \hat{z}_{i,t} \hat{z}'_{i,t} \text{ pour } i = 0, 1 \\ \widehat{M}_{01} &= \widehat{M}_{10} = \sum \hat{z}_{1,t} \hat{z}'_{0,t}. \end{aligned}$$

Il est possible de montrer que sous (H_0) , LM_0 est asymptotiquement distribuée comme un chi-deux à PQ degrés de libertés.

AR(p) contre EXPAR(p) On considère un processus (ε_t) bruit blanc gaussien, c'est à dire *i.i.d.* dont la loi est $\mathcal{N}(0, \sigma^2)$. Les hypothèses sont alors

$$(H_0) : X_t = \mu + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t \text{ contre } (H_1) : X_t = \mu + \sum_{i=1}^p [\phi_i + \pi_i \exp(-\gamma X_{t-1}^2)] X_{t-i} + \varepsilon_t.$$

De façon équivalente l'hypothèse alternative peut s'écrire

$$(H_1^*) : X_t = \mu + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=1}^p \theta_i X_{t-i} X_{t-1}^2 + \varepsilon_t.$$

Là encore, la statistique du test du multiplicateur de Lagrange est

$$LM_2 = \hat{\sigma}^{-2} \left[\sum \hat{z}_{1,t} \hat{\varepsilon}_t \right]' \left[\widehat{M}_{11} - \widehat{M}_{10} \widehat{M}_{00}^{-1} \widehat{M}_{01} \right] \left[\sum \hat{z}_{1,t} \hat{\varepsilon}_t \right],$$

où

$$\begin{cases} \hat{z}_{0,t} = (-1, X_{t-1}, \dots, X_{t-p})' \\ \hat{z}_{1,t} = -(X_{t-1}^3, X_{t-2} X_{t-1}^2, \dots, X_{t-p} X_{t-1}^2) \end{cases} \text{ et } \begin{aligned} \widehat{M}_{ii} &= \sum \hat{z}_{i,t} \hat{z}'_{i,t} \text{ pour } i = 0, 1 \\ \widehat{M}_{01} &= \widehat{M}_{10} = \sum \hat{z}_{1,t} \hat{z}'_{0,t}. \end{aligned}$$

Il est possible de montrer que sous (H_0) , LM_0 est asymptotiquement distribuée comme un chi-deux à p degrés de libertés.

2.7.2 Le test du CUSUM

Le test du CUSUM permet de détecter l'existence de régimes ou de seuils. Il consiste à reclasser les observations suivant les variables de transition. On estime alors les paramètres de manière récursive et on considère les sommes cumulées des résidus obtenus.

Par exemple, dans le cas d'un modèle *SETAR*, on suppose que l'on connaît le délai d , et que l'on dispose des observations $X_{-d+1}, \dots, X_0, \dots, X_T$. On ordonne les observations X_1, \dots, X_n selon les valeurs de X_{1-d}, \dots, X_{t-d} . Soient $X_{(1)}^T, \dots, X_{(T)}^T$ les observations ainsi réordonnées. Pour tout j , on considère le modèle linéaire suivant

$$X_{(i)}^T = \mu_j + \sum_{k=1}^p \phi_k^j X_{(i)-k} + \varepsilon_{(j)}^T \text{ pour tout } i \leq j.$$

L'hypothèse (H_0) est alors $\mu_j = \mu$, $\phi_k^j = \phi_k$ pour tout k , et l'hypothèse alternative (H_1) est que (H_0) n'est pas vérifiée.

2.7.3 Le test BDS

Ce test permet de détecter un bruit blanc indépendant, c'est-à-dire si les observations sont (*ou non*) indépendantes, à partir de l'étude de la "dimension de corrélation" (*notion issue de la théorie du chaos*).

Remarque 13 Cette notion, introduite par Grassberger et Procaccia en 1983 est très liée à la théorie des fractales et à la dimension de capacité. L'idée de base est la suivante : plus un ensemble est de dimension élevée, plus rapidement le nombre de voisins d'un point donné de cet ensemble augmentera avec la distance à ce point. Encore faut-il pouvoir évaluer le nombre de voisins. Dans ce but, nous introduisons l'intégrale de corrélation, définie comme la probabilité de trouver une paire de points distants de moins d'un certain rayon r (donné à priori) dans un espace d'immersion m -dimensionnel.

A partir d'une série temporelle, on commence donc par former des m -uples. Si l'on dispose de N observations, l'intégrale de corrélation sera donnée par

$$C_{n,N}(r) = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=1, j \neq i}^N \theta(r - d(i,j)),$$

où $d(i,,j)$ est la distance entre les points i et j et $\theta(k)$ est la fonction de Heaviside (ou fonction signe) qui, par définition, vaut 1 si $k \geq 0$ et 0 dans le cas contraire. Ceci revient donc à compter, dans un espace d'immersion, le nombre de voisins d'un point quelconque distants de moins de r . En principe, $d(i,,j)$ est la distance euclidienne usuelle, mais on prend en général la norme-sup (ou norme de Takens) pour distance. Cette dernière est définie comme la plus grande des différences (en valeur absolue) entre deux coordonnées équivalentes de 2 vecteurs. Quand $N \rightarrow \infty$, on pose alors $C_{m,N}(r) \rightarrow C_m(r) = \alpha r^{\nu_m}$. L'exposant ν_m ainsi défini sera la dimension de corrélation dans l'immersion m , dont l'estimateur naturel est

$$\hat{\nu}_m = \lim_{r \rightarrow 0} \frac{d \log(C_m(r))}{d \log r}.$$

A partir de cette notion, il est aisément de mettre en place le test BDS de non-linéarité.

Ce test peut être interprété comme un test de non linéarité : il suffit en effet d'estimer le modèle linéaire retenu et de calculer les résidus obtenus. On applique ensuite le BDS à ces résidus; si ce test révèle que les résidus sont indépendants et identiquement distribués, alors l'hypothèse de linéarité est acceptée. Dans le cas contraire, la série est soit stochastique non-linéaire, soit chaotique.

Soit le vecteur m -historique $X_t^m = (X_t, X_{t+1}, \dots, X_{t+m-1})$, et $T_m = n - m + 1$ le nombre de vecteurs m -historiques pour la série (X_t) avec n observations. On désigne par $I_\varepsilon(\cdot)$ la fonction indicatrice de $[0, \varepsilon]$, où ε est une constante positive préalablement choisie. On pose alors

$$C_{n,m}(\varepsilon) = \frac{2}{T_m(T_m-1)} \sum_{t < s} I_\varepsilon(\|X_t^m - X_s^m\|_\infty),$$

et on a, sous l'hypothèse où les (X_t) sont indépendants et identiquement distribués

$$\lim_{n \rightarrow \infty} C_{n,m}(\varepsilon) = \left[\lim_{n \rightarrow \infty} C_{n,1}(\varepsilon) \right]^m.$$

Remarque 14 Cette propriété peut être illustrée simplement dans le cas où $m = 2$: $C_{n,2}(\varepsilon)$ correspond à la fréquence des observations "plus proches voisins" dans le plan (X_{t-1}, X_t) tandis que $C_{1,n}(\varepsilon)$ mesure toujours la proportion de points de la série distants de moins de ε de moins de ε de leur valeur à la date précédente. Dans le cas d'une série indépendante, on voit que l'on considère une surface élémentaire, et que l'on double la longueur des côtés ($C_{1,n}$ est alors multiplié par 2), le nombre de points contenu dans la surface est alors multiplié par 4 (à cause de la répartition purement aléatoire des points dans le plan). En revanche, avec une série non-indépendante, le nombre de points n'est plus multiplié par 4 lorsque la surface élémentaire voit ses côtés doublés : $C_{n,1}(\varepsilon)$ est bien multiplié par 2 mais $C_{n,2}(\varepsilon)$ n'est pas multiplié par 4.

Compte tenu de cette remarque, on définit la statistique BDS par

$$BDS = \sqrt{n} \frac{C_{n,m}(\varepsilon) - C_{n,1}(\varepsilon)}{\sigma_m(\varepsilon)},$$

avec

$$\sigma_m(\varepsilon) = 2 \sqrt{K^m + 2 \sum_{j=1}^{m-1} K^{m-j} C^{2j} + (m-1)^2 C^{2m} - m^2 K C^{2m-2}},$$

et

$$C = C_{n,1}(\varepsilon) \text{ et } K = \frac{1}{n(n-1)(n-2)} \sum_{(t,r,s)} I_\varepsilon(|X_t - X_s|) I_\varepsilon(|X_r - X_s|),$$

où la somme s'effectue sur les (r, s, t) distincts deux à deux. Sous (H_0) : (X_t) est *i.i.d.* la statistique BDS est asymptotiquement normale, centrée et réduite.

2.7.4 Le test RESET (*Regression error specification test*)

On estime ici les paramètres du modèle linéaire

$$(H_0) : X_t = \mu + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t,$$

et on calcule les résidus obtenus $\hat{\varepsilon}_t$, les valeurs ajustées $\hat{X}_t = \hat{\mu} + \sum_{i=1}^p \hat{\phi}_i X_{t-i} = X_t - \hat{\varepsilon}_t$ et la somme des carrés des résidus $SCR_0 = \sum \hat{\varepsilon}_t^2$. On estime alors, par la méthode des moindres carrés, les paramètres de

$$\hat{\varepsilon}_t = b_0 + \sum_{i=1}^p b_i X_{t-i} + \sum_{j=1}^h \theta_j \hat{X}_t^j + \eta_t,$$

et on calcule $SCR_1 = \sum \hat{\eta}_t^2$. La statistique de test est

$$RESET = \frac{[SCR_0 - SCR_1] / (h-1)}{SCR_1 / (n-p-h)},$$

qui suit sous (H_0) (*hypothèse de modèle AR(p)*) une loi de Fisher $F(h-1, n-p-h)$. En fait, le test du RESET peut être interprété comme un test du multiplicateur de Lagrange.

2.7.5 La procédure AUTOREG sous SAS/ETS

La syntaxe de la procédure AUTOREG est la suivante

```
PROC AUTOREG options;
  BY variables;
  MODEL dependent = regressors / options;
  HETERO variables/options;
  RESTRICT equation,...,equation;
  TESTequation,...,equation / option;
  OUTPUT OUT=SAS data set options;
```

Estimation de modèles GARCH(p, q) Le modèle que nous souhaitons tester est de la forme suivante :

$$Y_t | Y_{t-1}, Y_{t-2}, \dots \sim \mathcal{N}(0, h_t) \text{ où } h_t = \omega + \sum_{i=1}^q \alpha_i Y_{t-i}^2 + \sum_{j=1}^p \gamma_j h_{t-j},$$

où h_t est alors appelée "variance conditionnelle". Le modèle ARCH(q) est obtenu quand $p = 0$. L'estimation sous SAS se fait par maximum de vraisemblance, sous l'hypothèse où le processus (η_t) est normalement distribué (*centré réduit*)

La déclaration MODEL

- Pour les modèles GARCH :

- GARCH = (option-list) : spécifie la forme du modèle *GARCH*, par exemple, pour un modèle *GARCH* (1,1) l'instruction sera

```
MODEL y=x1 x2 / GARCH=(q=1,p=1);
```

Toute la classe des modèles généralisant les modèles *GARCH* sont d'ailleurs obtenus à l'aide de l'instruction TYPE= : modèles *EGARCH* (*GARCH exponentiel*), *IGARCH* (*GARCH intégré*), sans contraintes de non-négativité, contrainte de stationnarité...

- DIST=T ou DIST=NORMAL : spécifie la distribution du terme d'erreur : soit une loi normale, soit une loi de Student,

- Pour les tests statistiques

- ARCHTEST : cette option permet d'obtenir les statistiques *Q* et *LM* permettant de tester l'absence d'effet *ARCH*,

- CHOW =(obs₁, ..., obs_n) : met en place les tests de Chow afin de tester la stabilité des coefficients de régression,

- COVEST =OP|HESSIAN|QML : cette option spécifie le type de matrice de covariance pour le modèle *GARCH*. OP signifie que la matrice produit est utilisée la matrice, HESSIAN signifie que la matrice Hessienne est utilisée pour calculer la matrice de covariance, et QML signifie que la méthode retenue est basée sur du quasi maximum de vraisemblance,

- DW =n : met en place le test de Durbin Watson à l'ordre *n* (*par défaut égal à 1*). L'instruction DWPROB donne la p-value du test.

- GODFREY =r : met en place le test du multiplicateur de Lagrange, avec comme hypothèse alternative des erreurs *ARMA*,

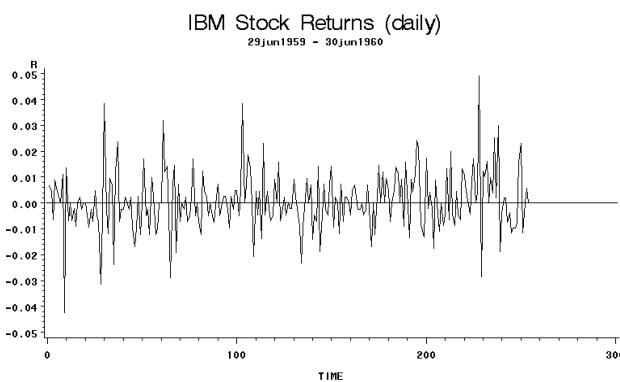
- RESET : met en place le test *RESET*, avec *p* valant 2, 3 ou 4,

- STATIONNARITY =(PHILLIPS) : met en place le test de Phillips-Perron, test de stationnarité du modèle

La déclaration OUPUT Un certain nombre d'output peuvent être imprimés, à l'aide des instructions CEV= variable ou - HT= variable. L'instruction CEV= option permet d'obtenir, en sortie, la table de la variance des erreurs du modèle hétéroscléastique (*spécifié dans l'instruction HETERO ou la valeur de la variance de l'erreur conditionnelle ht de la déclaration GARCH= option dans l'instruction MODEL*)

Parmi les tests possibles, on notera les instructions CUSUM et CUSUMSQ, dont les intervalles de confiance (haut et bas) peuvent être obtenus à l'aide des instructions CUSUMUB= variable, CUSUMUB= option, CUSUMLB= variable, CUSUMLB= option, CUSUMSQUB= variable, CUSUMSQUB= option ou encore CUSUMSQLB= option.

Estimation d'un modèle ARCH (p) sous SAS L'exemple ci-dessous (*tiré de Box et Jenkins (1976) page 527*) correspond au cours de l'action IBM, entre le 29 juin 1959 et le 30 juin 1960. Le rendement (*journalier*) de ce titre est la variable *r*, représenté ci-dessous,



Le modèle que nous souhaitons tester est de la forme suivante :

$$ARCH(2) : Y_t | Y_{t-1}, Y_{t-2}, \dots \sim \mathcal{N}(0, h_t) \text{ où } h_t = \omega + \alpha_1 Y_{t-1}^2 + \alpha_2 Y_{t-2}^2.$$

Afin de tester ce modèle, les instructions sont les suivantes

```
proc AUTOREG data=IBM;
  model r=/noint GARCH=(Q=2);
  output out=a cev=v;
```

et la sortie *SAS* (*estimation des paramètres et tests statistiques*) est alors

```

Autoreg Procedure
Dependent Variable = R

Ordinary Least Squares Estimates
      SSE      0.032143   DFE      254
      MSE      0.000127   Root MSE  0.011249
      SBC     -1558.8    AIC     -1558.8
      Reg Rsq   0.0000   Total Rsq  0.0000
      Durbin-Watson 2.1377

NOTE: No intercept term is used. R-squares are redefined.

GARCH Estimates
      SSE      0.032143   OBS      254
      MSE      0.000127   UVAR     0.000126
      Log L    781.0174   Total Rsq  0.0000
      SBC     -1545.42   AIC     -1556.03
      Normality Test 105.8557 Prob>Chi-Sq  0.0001

      Variable   DF      B Value   Std Error   t Ratio Approx Prob
      ARCH0      1      0.000112   7.561E-6   14.851      0.0001
      ARCH1      1      0.041348   0.0511      0.810      0.4181
      ARCH2      1      0.069749   0.0432      1.616      0.1062

```

c'est à dire que $\hat{\omega}$, $\hat{\alpha}_1$ et $\hat{\alpha}_2$ valent respectivement 0.00011, 0.04136 et 0.06976. Les coefficients donnés sont, pour l'estimation par moindre carrés :

- *SSE* : (Sum of Square Errors)
- *MSE* : (Mean Square Error)
- *SBC* : critère de Schwarz défini par $SBC = -2 \ln \mathcal{L} + K \ln N$, où \mathcal{L} est la vraisemblance, N le nombre d'observations et K le nombre de paramètres estimés.
- Regress *R-square* : donné par

$$R_{REG}^2 = 1 - \frac{TSSE}{TSST},$$

où *TSST* est la somme des carrés ...

- *DW* : statistique de Durbin-Watson
- *DFE* : (Error Degrees of Freedom) the number of observations minus the number of parameters
- *Root MSE* : (Root Mean Square Error)
- *AIC* : critère d'Akaike défini par $AIC = -2 \ln \mathcal{L} + 2K$, où \mathcal{L} est la vraisemblance et K le nombre de paramètres estimés.
- Total *R-square* : donné par

$$R_{TOT}^2 = 1 - \frac{SSE}{SST},$$

Et pour l'estimation des paramètres *GARCH*

- Log Likelihood : valeur de la log-vraisemblance (*sous hypothèse de normalité des résidus*)
- Normal Test : test de Bera et Jarque (1982) basé sur l'utilisation du skewness et de la kurtosis (*qui doivent être proches respectivement de 0 et de 3 dans le cas normal*)

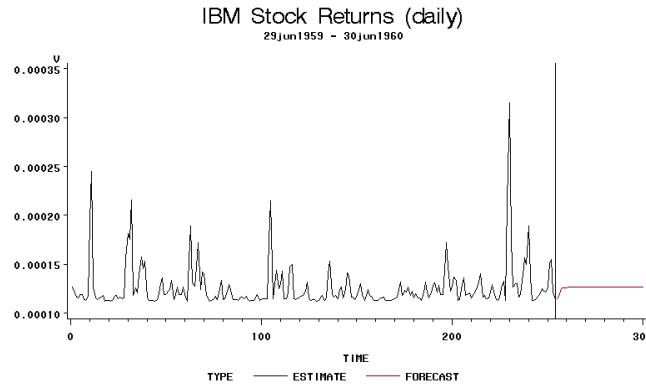
- Observations : nombre d'observations utilisées

- Uncond Var : variance non-conditionnelle

- Total *R-square* :

- Pr>ChiSq : *p*-value associée au test de normalité

SAS renvoie également la sortie graphique suivante, correspondant à la variance conditionnelle

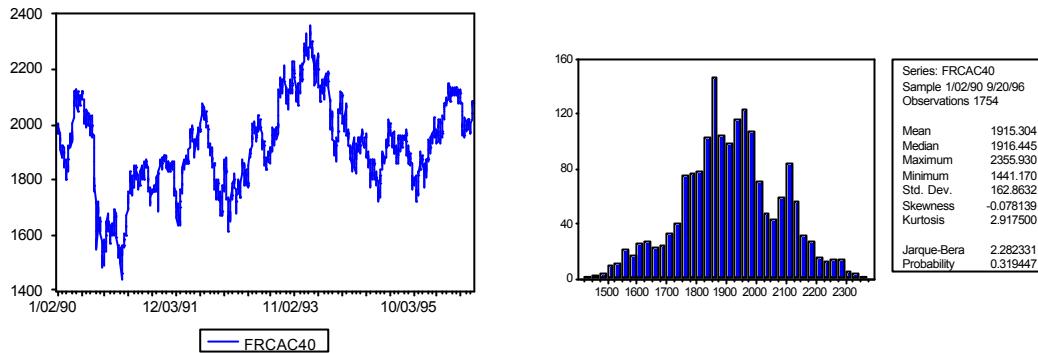


ainsi que la prévision.

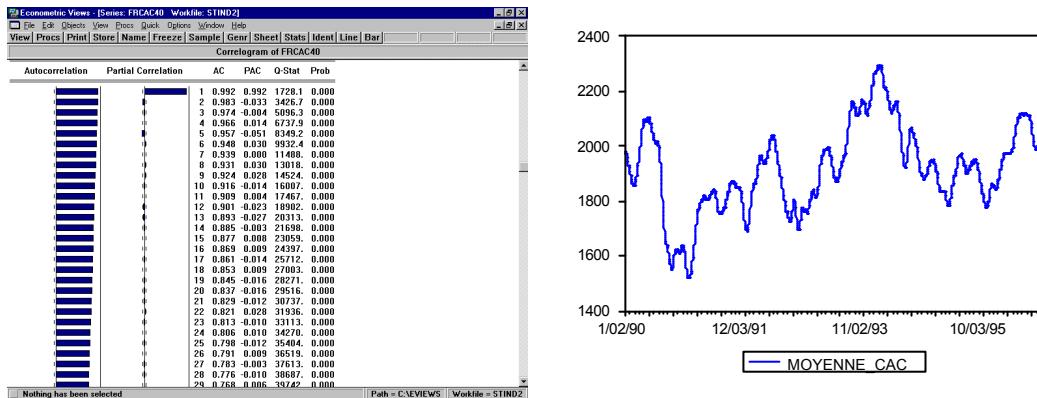
2.8 Application sur des données réelles

2.8.1 Modélisation GARCH du CAC40

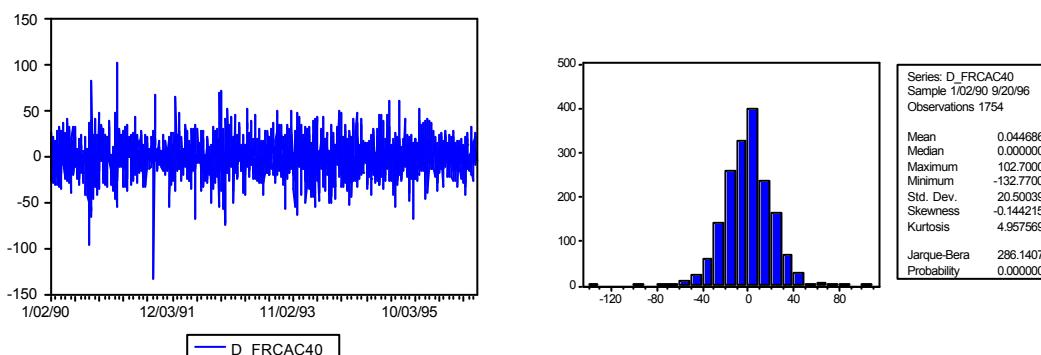
Présentation de la série et analyse préliminaire Considérons la série du CAC40, observée du 2 janvier 1990 au 20 septembre 1996 (soit un peu plus de 1750 observations)



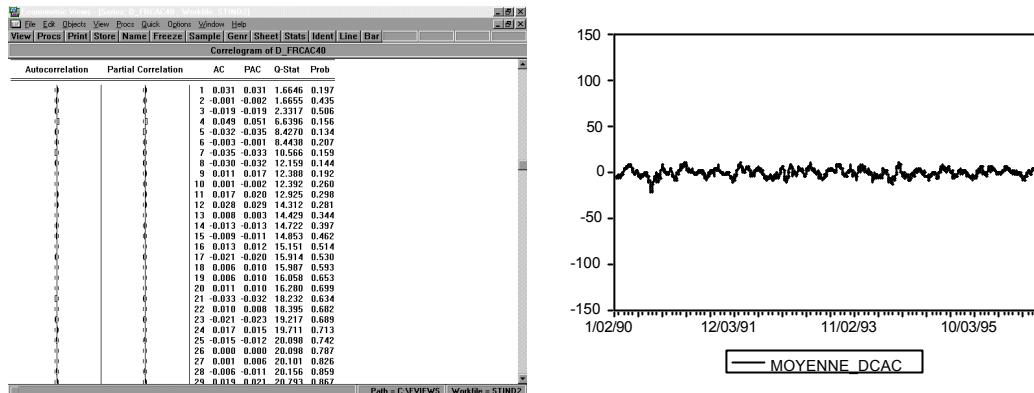
Comme le suggère l'autocorrélogramme ci-dessous, ainsi que le processus moyenne glissante (avec équiperdération, ± 10 jours (soit 21 jours)), cette série n'est pas stationnaire



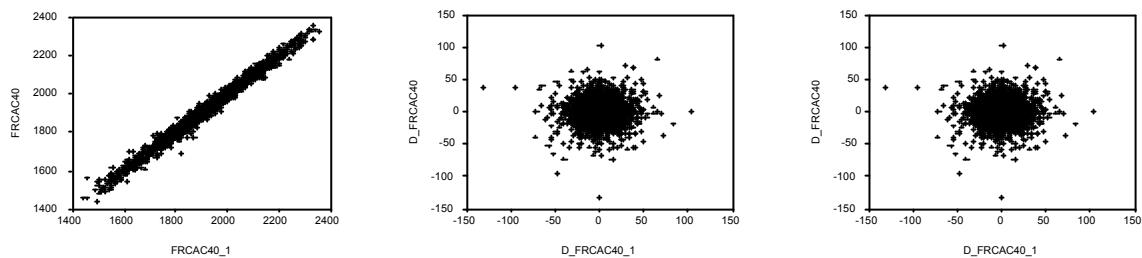
Pour stationnariser la série, considérons la série différenciée, correspondant à la variation de l'indice CAC40 sur une journée (*en points*)



Cette transformation semble stationnariser la série au second ordre, comme le laissent à penser les graphiques ci-dessous,



Si la série initiale du CAC40 présentait - a priori - une racine unité, comme cela peut se voir sur le scatterplot ci-dessous à gauche, on peu noter que la série différenciée une fois est elle stationnaire (*la corrélation entre deux dates - 1 jour et 5 jours ci-dessous - est d'ailleurs très faible*),

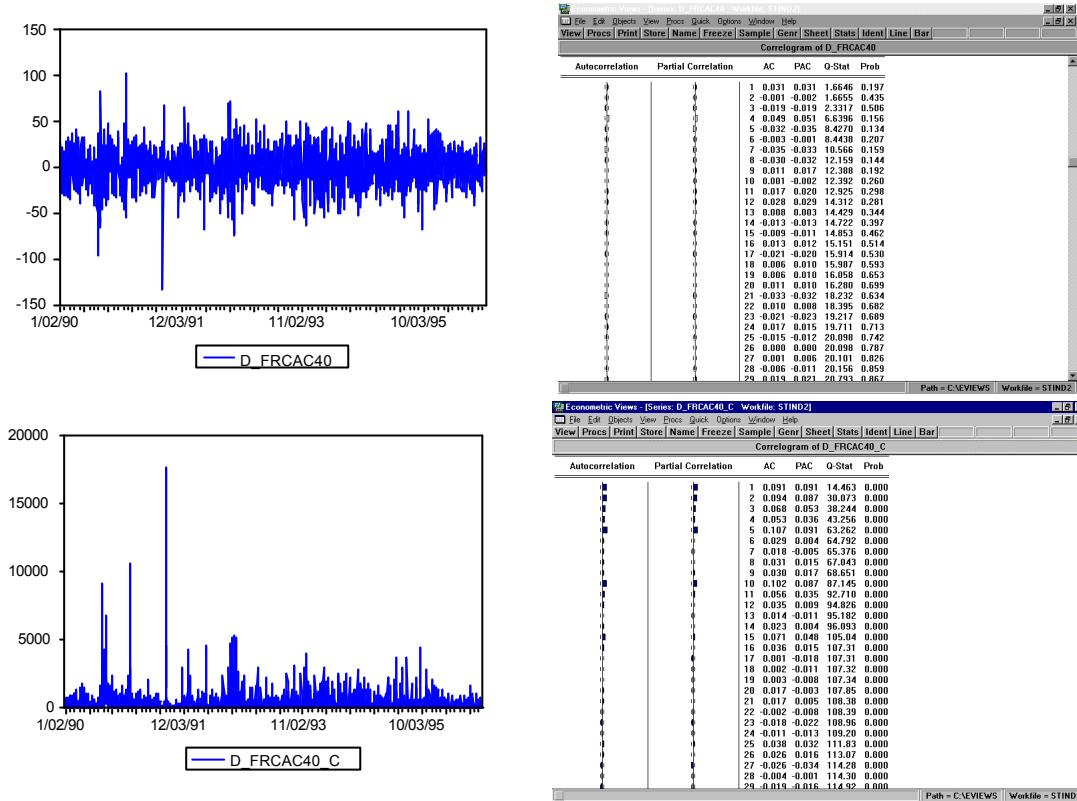


Tests de non-linéarité Afin de déterminer le type de modélisation à adopter, un certain nombre de tests simples de non-linéarité peuvent être utilisés : multiplicateur de Lagrange et *RESET*. Quel que soit le test, on peut noter une forte non-linéarité de la série (*cela peut toutefois être dû à la nature particulière des tests effectués, postulant une forme AR (1) comme hypothèse alternative*) :

Test	Valeur de la statistique	p-value	Linéaire
LM (AR contre ARCH)	206.6051	0.00000	non
LM (AR contre BL)	9.236277	0.00237	non
LM (AR contre EXPAR)	39.06230	0.00000	non
RESET	8.876685	0.00293	non

L'étude des autocorréogrammes de la série des variations et de la série au carré semblent confirmer ce phénomène : si les autocorréogrammes de la série n'indiquent pas de corrélations importantes (*laissant à penser que la série pourrait être un bruit blanc*), des autocorrelations non-nulles apparaissent en considérant la série définie comme le carré de la

variation du *CAC40*,



Ces résultats incitent à adapter une modélisation de type *ARCH/GARCH* et non pas *ARIMA*.

Modélisation *ARCH – GARCH* de la variation du *CAC40* Le critère d'Akaike incite à retenir un modèle *ARCH(1)*. Il pourrait être également intéressant de considérer des modèles *GARCH(2,1)* ou *ARMA* (à titre de comparaison). Nous noterons dans la suite X_t la série correspondant à la variation de l'indice *CAC40*.

- **Modèle *ARCH(1)* - avec erreurs conditionnellement normales**

Le modèle est de la forme

$$X_t | X_{t-1} \sim N(0, \sigma_t^2) \text{ où } \sigma_t^2 = 398.6559 + 0.1488X_{t-2}^2.$$

Considérons la série des résidus standardisés (*la série des "résidus" coïncidant avec la série des variations elle-même*).

- Modèle *GARCH(2,1)*

[A INSERER]

- Modèle *ARMA*

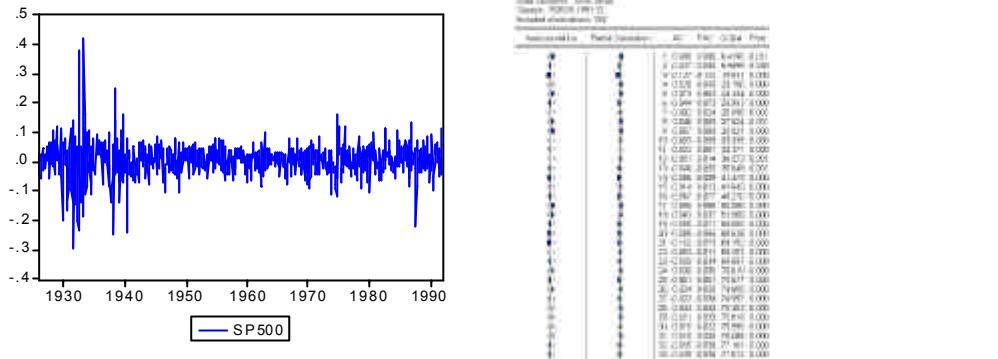
[A INSERER]

2.8.2 Modélisation des rendements du cours Intel : modèle *ARCH*

[A INSERER]

2.8.3 Modélisation des rendements de l'indice S&P : modèle GARCH

Considérons la série du rendement mensuel de l'indice SP500, depuis 1926 (soit de l'ordre de 800 observations),



Au vu de l'autocorrélogramme présenté ci-dessus, il est possible de modéliser ce processus soit par un $AR(3)$, soit par un $MA(3)$ (voire un $ARMA(3,3)$)

Dependent Variable: SP500				
Method: Least Squares				
Date: 05/28/03 Time: 10:04				
Sample: 1926-01 1991:12				
Included observations: 792				
Convergence achieved after 5 iterations				
Backcast: 1925:10 1925:12				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.006161	0.001955	3.151074	0.0017
MA(1)	0.094167	0.035150	2.678963	0.0075
MA(3)	-0.141599	0.035173	-4.025809	0.0001
R-squared	0.026877	Mean dependent var	0.006143	
Adjusted R-squared	0.024410	S.D. dependent var	0.058459	
S.E. of regression	0.057741	Akaike info criterion	2.861921	
Sum squared resid	2.630535	Schwarz criterion	-2.844215	
Log likelihood	1136.321	F-statistic	10.89573	
Durbin-Watson stat	1.995371	Prob(F-statistic)	0.000021	

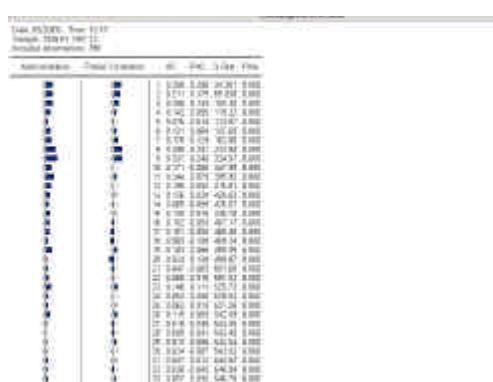
Dependent Variable: SP500				
Method: Least Squares				
Date: 05/28/03 Time: 09:59				
Sample(adjusted): 1926-04 1991:12				
Included observations: 789 after adjusting endpoints				
Convergence achieved after 3 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.006262	0.001983	3.158519	0.0016
AR(1)	0.086030	0.035303	2.436595	0.0150
AR(3)	-0.125241	0.035302	-3.547748	0.0004
R-squared	0.023610	Mean dependent var	0.006269	
Adjusted R-squared	0.021126	S.D. dependent var	0.058494	
S.E. of regression	0.057872	Akaike info criterion	2.867358	
Sum squared resid	2.632480	Schwarz criterion	-2.839599	
Log likelihood	1130.228	F-statistic	9.503281	
Durbin-Watson stat	1.979390	Prob(F-statistic)	0.000084	
Inverted AR Roots	.28 -.43i	.28+.43i	-.47	

i.e.

$$\begin{cases} X_t = 0.006 + \varepsilon_t + 0.094\varepsilon_{t-1} - 0.141\varepsilon_{t-3} \\ X_t = 0.006 + 0.088X_{t-1} - 0.123X_{t-2} + \eta_t. \end{cases}$$

Toutefois, si l'on regarde par exemple l'autocorrélogramme de la série (η_t^2) , ci-dessous à droite, on peut constater que (η_t) n'est pas un bruit blanc,

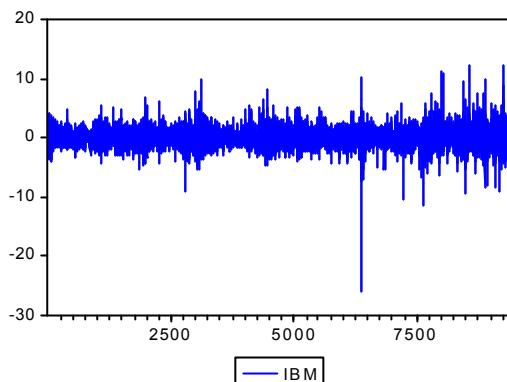
Dependent Variable: RESID_AR3				
Method: ML - ARCH (Marquardt)				
Date: 05/28/03 Time: 10:17				
Sample(adjusted): 1926-04 1991:12				
Included observations: 789 after adjusting endpoints				
Convergence achieved after 14 iterations				
Variance backcast: ON				
Variable	Coefficient	Std. Error	z-Statistic	Prob.
Variance Equation				
C	7.97E-05	2.42E-05	3.289844	0.0010
ARCH(1)	0.113723	0.020449	5.561203	0.0000
GARCH(1)	0.862719	0.020530	42.02175	0.0000
R-squared	0.000000	Mean dependent var	7.05E-14	
Adjusted R-squared	-0.002545	S.D. dependent var	0.057799	
S.E. of regression	0.057872	Akaike info criterion	-3.189878	
Sum squared resid	2.632480	Schwarz criterion	-3.172118	
Log likelihood	1261.407	Durbin-Watson stat	1.979390	



Il est alors possible de tester un modèle $AR(3) - GARCH(1,1)$
[A INSERER]

2.8.4 Modélisation des rendements de l'action IBM (avec dividendes) : modèle TAR

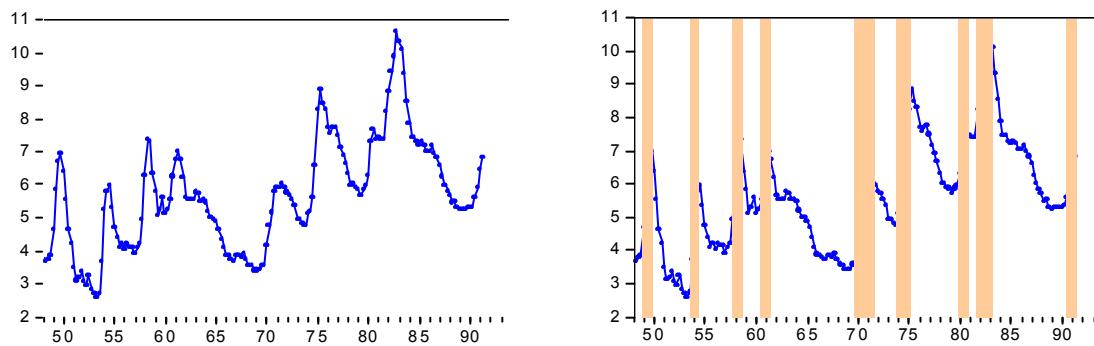
Considérons la série des rendements journaliers de l'action *IBM*, de juillet 62 à décembre 99



[A INSERER]

2.8.5 Modélisation du taux de chômage aux Etats Unis : modèles TAR et ARCH

Considérons la série du taux de chômage désaisonnalisé en données trimestrielles¹⁰ aux Etats-Unis, de 1948 à 1993,



Comme le montre le graphique de droite, cette série présente des phases 'cycliques', avec de lentes variations, suivies de très fortes variations, laissant à penser que les comportements à la hausse et à la baisse sont sensiblement différents. Ce type de constatation empirique pousse à tester un modèle *TAR*. L'utilisation de modèles *TAR* sur des données macroéconomiques a été confirmée, par exemple pour modéliser le *PIB* américain, dans 'A Nonlinear Approach to US GNP' par Potter en 1995.

¹⁰Cette série est étudiée dans Montgomery, Zarnowitz, Tsay et Tiao (1998), 'Forecasting the U.S. Unemployment Rate', dans Journal of the American Statistical Association.

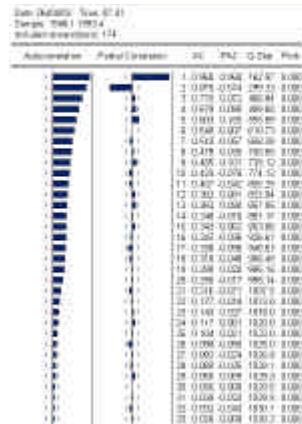
Modèle linéaire ARIMA L'autocorrélogramme (ou uniquement une analyse graphique de la série), tendent à rejeter l'hypothèse de stationnarité de la série, ainsi que le test de Phillips Perron

Null Hypothesis: X has a unit root Exogenous: Constant, Linear Trend Bandwidth: 2 (Newey-West using Bartlett kernel)		
	Adj. t-Stat	Prob.*
Phillips-Perron test statistic	-2.980680	0.1407
Test critical values:		
1% level	-4.011977	
5% level	-3.436009	
10% level	-3.142085	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(X) has a unit root Exogenous: Constant, Linear Trend Bandwidth: 17 (Newey-West using Bartlett kernel)		
	Adj. t-Stat	Prob.*
Phillips-Perron test statistic	-4.778448	0.0007
Test critical values:		
1% level	-4.012296	
5% level	-3.436163	
10% level	-3.142175	

*MacKinnon (1996) one-sided p-values.



Cette présence de racine unité correspond au phénomène économique appelé "présence d'effets d'hystérosis du taux de chômage". Le test de Dickey Fuller (avec constante et tendance) rejette en revanche l'hypothèse de présence de racine unité.

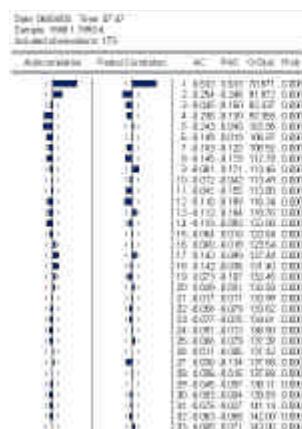
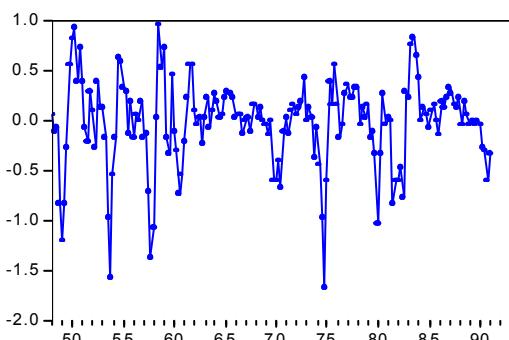
Null Hypothesis: X has a unit root Exogenous: Constant, Linear Trend Lag Length: 1 (Automatic based on SIC, MAXLAG=13)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-4.745048	0.0008
Test critical values:		
1% level	-4.012296	
5% level	-3.436163	
10% level	-3.142175	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: DX has a unit root Exogenous: Constant Lag Length: 1 (Automatic based on SIC, MAXLAG=13)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-7.087081	0.0000
Test critical values:		
1% level	-3.468749	
5% level	-2.878311	
10% level	-2.575791	

*MacKinnon (1996) one-sided p-values.

La série différenciée une fois semblant stationnaire, il est alors possible de tester un modèle ARMA, en particulier des modèles $AR(p)$,



Les sorties suivantes correspondent à l'estimation de modèles AR(2) - les deux premières autocorrélations partielles étant clairement significatives - et AR(4) - la série étant trimestrielle, il convient de tester une '*saisonnalité*'

Dependent Variable: DX				
Method: Least Squares				
Date: 06/04/03 Time: 07:53				
Sample(adjusted): 1948:3 1991:1				
Included observations: 171 after adjusting endpoints				
<u>Convergence achieved after 3 iterations</u>				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.790459	0.074567	10.60065	0.00000
AR(2)	-0.245083	0.074975	-3.268889	0.00130
R-squared	0.437927	Mean dependent var		-0.017930
Adjusted R-squared	0.434601	S.D. dependent var		0.439577
S.E. of regression	0.330531	Akaike info criterion		0.635393
Sum squared resid	18.46335	Schwarz criterion		0.672138
Log likelihood	-52.32612	Durbin-Watson stat		2.078329
Inverted AR Roots	.40+.30i	.40 -.30i		

Assumption	Final Estimate	AE	PAC	BLR	Test
1. $\alpha = 0.05$	0.000	0.000	0.000	0.000	0.000
2. $\alpha = 0.01$	0.000	0.000	0.000	0.000	0.000
3. $\alpha = 0.001$	0.000	0.000	0.000	0.000	0.000
4. $\beta = 0.05$	-0.161	0.205	0.000	0.000	0.000
5. $\beta = 0.01$	-0.161	0.205	0.000	0.000	0.000
6. $\beta = 0.001$	-0.161	0.205	0.000	0.000	0.000
7. $\delta = 0.05$	-0.161	0.205	0.000	0.000	0.000
8. $\delta = 0.01$	-0.161	0.205	0.000	0.000	0.000
9. $\delta = 0.001$	-0.161	0.205	0.000	0.000	0.000
10. $\sigma^2 = 0.05$	-0.161	0.205	0.000	0.000	0.000
11. $\sigma^2 = 0.01$	-0.161	0.205	0.000	0.000	0.000
12. $\sigma^2 = 0.001$	-0.161	0.205	0.000	0.000	0.000
13. $\sigma^2 = 0.0001$	-0.161	0.205	0.000	0.000	0.000
14. $\alpha = 0.10$	0.000	0.024	0.000	0.000	0.000
15. $\alpha = 0.02$	0.000	0.024	0.000	0.000	0.000
16. $\alpha = 0.002$	0.000	0.024	0.000	0.000	0.000
17. $\alpha = 0.0002$	0.000	0.024	0.000	0.000	0.000
18. $\alpha = 0.00002$	0.000	0.024	0.000	0.000	0.000
19. $\alpha = 0.000002$	0.000	0.024	0.000	0.000	0.000
20. $\alpha = 0.0000002$	0.000	0.024	0.000	0.000	0.000
21. $\alpha = 0.00000002$	0.000	0.024	0.000	0.000	0.000
22. $\alpha = 0.000000002$	0.000	0.024	0.000	0.000	0.000
23. $\alpha = 0.0000000002$	0.000	0.024	0.000	0.000	0.000
24. $\alpha = 0.00000000002$	0.000	0.024	0.000	0.000	0.000
25. $\alpha = 0.000000000002$	0.000	0.024	0.000	0.000	0.000
26. $\alpha = 0.0000000000002$	0.000	0.024	0.000	0.000	0.000
27. $\alpha = 0.00000000000002$	0.000	0.024	0.000	0.000	0.000
28. $\alpha = 0.000000000000002$	0.000	0.024	0.000	0.000	0.000
29. $\alpha = 0.0000000000000002$	0.000	0.024	0.000	0.000	0.000
30. $\alpha = 0.00000000000000002$	0.000	0.024	0.000	0.000	0.000
31. $\alpha = 0.000000000000000002$	0.000	0.024	0.000	0.000	0.000

Dependent Variable: DX				
Method: Least Squares				
Date: 06/04/03 Time: 07:51				
Sample(adjusted): 1949:1 1991:1				
Included observations: 169 after adjusting endpoints				
Convergence achieved after 3 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.735183	0.075316	9.761291	0.000000
AR(2)	-0.171892	0.078142	-2.199747	0.029237
AR(4)	-0.161908	0.060665	-2.668864	0.008446
R-squared	0.469192	Mean dependent var	-0.012817	
Adjusted R-squared	0.462797	S.D. dependent var	0.437634	
S.E. of regression	0.320760	Akaike info criterion	0.581346	
Sum squared resid	17.07926	Schwarz criterion	0.636907	
Log likelihood	46.12376	Durbin-Watson stat	0.23372	
Inverted AR Roots	.64 -.45i	.64+.45i	-.27+.43i	-.27-.43i

On peut noter que, dans les deux cas, l'hypothèse de bruit blanc des résidus doit être rejetée. Un modélisation à l'aide d'un modèle *ARMA*(4, 4) sur la série différenciée donne

Dependent Variable: DX					
Method: Least Squares					
Date: 06/04/03 Time: 07:58					
Sample(adjusted): 1949:1 1991:1					
Included observations: 169 after adjusting endpoints					
Convergence achieved after 36 iterations					
Backcast: 1948: 1948:4					
Variable	Coefficient	Std. Error	t-Statistic	Prob.	
AR(1)	1.297687	0.163808	7.922013	0.0000	
AR(2)	-0.993817	0.256960	-3.867598	0.0002	
AR(3)	0.760099	0.226467	3.356336	0.0010	
AR(4)	-0.229478	0.130645	-1.755605	0.0805	
MA(1)	-0.600397	0.160055	-3.751166	0.0002	
MA(2)	0.560191	0.175028	3.200575	0.0017	
MA(3)	-0.500735	0.158612	-3.156981	0.0019	
MA(4)	-0.377559	0.141700	-2.664495	0.0895	
R-squared	0.544289	Mean dependent var		-0.012817	
Adjusted R-squared	0.524475	S.D. dependent var		0.437634	
S.E. of regression	0.301785	Akaike info criterion		0.487797	
Sum squared resid	14.66296	Schwarz criterion		0.636136	
Log likelihood	-33.23409	Durbin-Watson stat		2.031076	
Inverted AR Roots	.79	.49	.01	-.771	.01+.771
Inverted MA Roots	.97	.02	-.971	.02+.971	

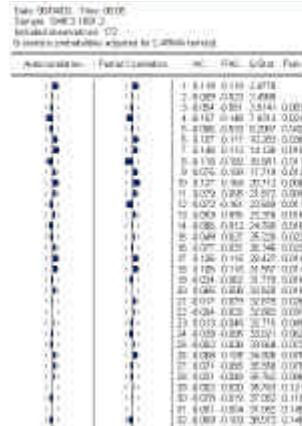
Automotive	Partial Concentration	AC	EC	SCHE	Pre
1	7.43E-009	1.172			
2	4.02E-008	1.138			
3	2.01E-007	1.104			
4	9.94E-007	4.938			
5	5.07E-006	5.849			
6	2.53E-005	3.870			
7	1.27E-004	1.937			
8	6.30E-004	4.091	4.437		
9	3.17E-003	1.651	6.248		
10	1.58E-002	3.011	4.512	8.185	
11	7.50E-002	1.500	1.600	1.723	
12	3.75E-001	0.989	1.088	1.241	
13	1.875E+000	0.692	0.768	0.847	
14	9.375E+000	0.154	0.3932	0.326	
15	4.687E+001	0.002	0.3591	0.314	
16	2.344E+002	0.0002	0.3591	0.314	
17	1.172E+003	0.00002	0.3591	0.314	
18	5.860E+003	0.000002	0.3591	0.314	
19	2.930E+004	0.0000002	0.3591	0.314	
20	1.465E+005	0.00000002	0.3591	0.314	
21	7.325E+005	0.000000002	0.3591	0.314	
22	3.662E+006	0.0000000002	0.3591	0.314	
23	1.831E+007	0.00000000002	0.3591	0.314	
24	9.155E+007	0.000000000002	0.3591	0.314	
25	4.577E+008	0.0000000000002	0.3591	0.314	
26	2.288E+009	0.00000000000002	0.3591	0.314	
27	1.144E+010	0.000000000000002	0.3591	0.314	
28	5.722E+010	0.0000000000000002	0.3591	0.314	
29	2.861E+011	0.00000000000000002	0.3591	0.314	
30	1.430E+012	0.000000000000000002	0.3591	0.314	
31	7.150E+012	0.0000000000000000002	0.3591	0.314	

L'hypothèse de bruit blanc des résidus semble ici validée.

⁷ Une modélisation *ARMA* (ici *AR* (2) à la vue du corrélogramme) de la série brute (X_t) - et non plus $(1 - L)X_t$ -

donne

Dependent Variable: X
Method: Least Squares
Date: 06/04/03 Time: 08:05
Sample(adjusted): 1948:3 1991:2
Included observations: 172 after adjusting endpoints
Convergence achieved after 3 iterations
Variable Coefficient Std. Error t-Statistic Prob.
C 5.801028 0.420081 13.80931 0.0000
AR(1) 1.602622 0.057652 27.79812 0.0000
AR(2) -0.662140 0.057465 -11.52241 0.0000
R-squared 0.960964 Mean dependent var 5.689140
Adjusted R-squared 0.960502 S.D. dependent var 1.642964
S.E. of regression 0.326525 Akaike info criterion 0.616667
Sum squared resid 18.01853 Schwarz criterion 0.671565
Log likelihood -50.03334 F-statistic 2080.160
Durbin-Watson stat 1.761995 Prob(F-statistic) 0.00000000
Inverted AR Roots 80.-14i 80+.14i



où là encore, l'hypothèse de bruit blanc des résidus est encore rejetée.

Modèle à deux niveaux : modèle TAR Comme nous l'avions noté empiriquement, il existe un comportement différent, suivant que le taux de chômage monte ou baisse, i.e. si la variation du taux est positive ou négative. Il conviendrait donc d'essayer un modèle *TAR* sur $Y_t = (1 - L) X_t$, avec comme seuil 0. Nous allons tester ici des modèles à deux retards, de la forme

$$Y_t = \begin{cases} \alpha + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} & \text{si } Y_{t-2} \leq 0 \\ \beta + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} & \text{si } Y_{t-2} > 0 \end{cases}$$

En testant le modèle *AR*(2) sous la contrainte $Y_{t-2} \leq 0$, on obtient que

[A INSERER]

Modèle ARMA avec erreurs ARCH Dans les modèles *ARMA* testés ci-dessus, on peut noter qu'un test de présence d'effet *ARCH* sur les résidus est validée à chaque fois (*AR*(4) et *ARMA*(4,4) respectivement)

ARCH Test:				
F-statistic	10.09410	Probability	0.001774	
Obs*R-squared	9.630129	Probability	0.001914	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 06/04/03 Time: 09:20				
Sample(adjusted): 1949:2 1991:1				
Included observations: 168 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.075393	0.015049	5.009880	0.0000
RESID^2(-1)	0.238223	0.074981	3.177122	0.0018
R-squared	0.057322	Mean dependent var	0.099608	
Adjusted R-squared	0.051643	S.D. dependent var	0.172709	
S.E. of regression	0.168190	Akaike info criterion	-0.715610	
Sum squared resid	4.695797	Schwarz criterion	-0.678419	
Log likelihood	62.11120	F-statistic	10.09410	
Durbin-Watson stat	2.010038	Prob(F-statistic)	0.001774	

ARCH Test:				
F-statistic	14.80236	Probability	0.000170	
Obs*R-squared	13.75422	Probability	0.000208	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 06/04/03 Time: 09:21				
Sample(adjusted): 1949:2 1991:1				
Included observations: 168 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.061079	0.012965	4.711009	0.0000
RESID^2(-1)	0.285836	0.074294	3.847383	0.0002
R-squared	0.081870	Mean dependent var	0.086026	
Adjusted R-squared	0.076339	S.D. dependent var	0.151415	
S.E. of regression	0.145521	Akaike info criterion	-1.005156	
Sum squared resid	3.515287	Schwarz criterion	-0.967966	
Log likelihood	86.43312	F-statistic	14.80236	
Durbin-Watson stat	2.079100	Prob(F-statistic)	0.000170	

[A INSERER]

2.9 Conseils bibliographiques

- DROESBEKE,J.J., FICHET,B. & TASSI,P. (1994). Modélisation *ARCH* : théorie statistique et applications dans le domaine de la finance. *Association pour la Statistique et ses Utilisations*. Ellipses
- ENGLE,R. (1995). *ARCH* : selected readings. Oxford University Press.
- GOURIEROUX,C. (1997). Modèles hétéroscédastiques. *dans "Encyclopédie des Marchés Financiers"*, Economica
- GOURIEROUX,C, SCAILLET,O. & SZAFARZ,A.. (1998). *Econométrie de la finance : analyses historiques*. Economica
- JOHNSTON,J. (1988). Méthodes Econométriques. *Econometrica*.

2.10 Compléments sur les séries temporelles en finance

On pourra consulter sur ce sujet l'article de Walter (2003) '*1900-2000 : un siècle de descriptions statistiques des fluctuations boursières ou les aléas du modèle de marché aléatoire en finance*', ainsi que les deux ouvrages de Berstein (19) et (19), 'Against the gods : ' et '.

2.10.1 Introduction historique

2.10.2 Complément sur les données disponibles

Comme nous l'avons vu en introduction, il est possible de travailler par cotation, c'est à dire en utilisant le prix des titres à chaque fois qu'ils sont échangés sur le marché. Soit N_t le nombre de fois où le titre a été échangé entre la date 0 et la date t . On observe alors $\Sigma_1, \Sigma_2, \dots, \Sigma_{N_t}$ la suite des cotations, et $\nu_1, \nu_2, \dots, \nu_{N_t}$ la suite correspondant au volume échangé à chaque cotation. Entre la date 0 et la date t il est possible d'écrire $S_t = \Sigma_0 + (\Sigma_1 - \Sigma_0) + (\Sigma_2 - \Sigma_1) + \dots + (\Sigma_{N_t} - \Sigma_{N_t-1})$, alors que le volume final échangé est $V_t = \nu_1 + \nu_2 + \dots + \nu_{N_t}$. On suit alors (S_t) le processus du cours de l'actif, ou processus du prix d'échange, (N_t) le processus de comptage correspondant au nombre de transaction et (ν_t) le processus du nombre de titres échangés.

Dans ce cours, on s'intéresse à la série (S_t) observée à des dates régulièrement espacées (*jour, mois...*), au processus de gain (Y_t) où $Y_t = S_t - S_0$, ou encore au taux de rentabilité (r_t) où $r_t = \log S_t - \log S_0$. De façon plus générale, il est possible de considérer des rendements sur des périodes τ , et on étudie alors $(r_{t,\tau})$ où $r_{t,\tau} = \log S_t - \log S_{t-\tau}$.

- Le choix du temps

Le niveau du Nil, ses crues et ses étiages sont connus depuis l'année 622. Il s'agit là de données exceptionnelles, fascinantes, tant pour les hydrologues que pour les statisticiens. Quel est le bon modèle pour prédire les crues importantes ou l'étiage des années de sécheresse ? Est-ce que les années de sécheresse sont distribuées de manière uniforme, « au hasard » ? Supposons¹¹ vouloir prédire les crues entre 1107 et 1206 à partir des crues observées entre 1007 et 1106. Si l'on cherche à construire un modèle à mémoire courte comme en économie, la prévision devient vite simplement la moyenne du siècle passé, du XI^e siècle. Elle ne rend pas bien compte du phénomène réellement observé au XII^e siècle. Il faut donc introduire une mémoire longue, même très longue, mais à la condition d'être capable de la faire avec un nombre restreint de paramètres qu'il faudra estimer. Il existe plusieurs sortes de modèles de ce type donnant d'assez bons résultats.

La Bible (Genèse 41) dit: "Sept années d'abondance vont venir sur la terre d'Egypte, mais elles seront suivies de sept années de famine". Nous ne possédons pas de données sur le niveau du Nil à l'époque biblique mais sur les données connues, on observe de longues périodes d'humidité et de longues périodes de sécheresse et non de brèves alternances. Un des promoteurs des modèles à longue mémoire, B. Mandelbrot, a donc appelé ce phénomène l' "effet Joseph". Son nom scientifique est l'effet Hurst. Ce spécialiste d'hydrologie a pu montrer en construisant un modèle de réservoir qui peut représenter le bassin du Nil, que de simples considérations sur les qualités du réservoir, indépendamment de toutes variations du climat, pouvaient expliquer cet "effet Joseph". Si l'on prend en compte les données du Nil sur une période plus longue, par exemple, sur les cinq siècles allant du VII^e au XIII^e, on peut estimer avec précision la forme de la mémoire, si l'on a une idée du type de modèle qui décrit le régime du fleuve. Pour trouver ce modèle on utilise la constatation suivante : si l'on fait un changement d'échelle convenable sur les données de 50 années consécutives, le résultat a la même allure que la suite des données sur 500 ans. Cette permanence par changement d'échelle s'appelle l'autosimilarité. Elle guide le choix du modèle à longue mémoire¹².

Mais le Nil pose une autre question; il semble qu'au cours du X^e et du XVII^e siècle son régime ait changé, qu'il y ait eu une rupture du niveau moyen des crues puis de la forme de la mémoire. Et les tests faits à partir de modèles à longue mémoire ont tendance à conforter cette hypothèse dont le motif physique est mal connu. Le problème des crues du Nil concerne la recherche d'un changement de régime qui a eu lieu il y a plusieurs siècles. Pour cette recherche, si l'on se place à l'instant de cette rupture hypothétique, on dispose d'un passé mais aussi d'un futur comportant un nombre significatif de données. Il n'en est pas ainsi en économie puisqu'il n'y a pas de futur à l'instant même de la rupture hypothétique. En 1992, les pouvoirs publics ont tablé sur 2,4 % de croissance l'an, certains instituts de prévision, quant à eux, sur 3,4 %. En fait de croissance, on a constaté plus de 1% de récession. L'I.N.S.E.E. n'est pas en cause. Il utilise des modèles d'analyse économétrique dont les paramètres sont estimés à partir de l'observation d'un passé récent. Les prévisions fournies n'ont de validité qu'en faisant l'hypothèse de stationnarité. Ces analyses ne sont pas faites pour intégrer l'imprévu lorsqu'il est structurel, c'est-à-dire lorsqu'il invalide les modèles économétriques admis jusqu'alors, lorsqu'il faut envisager un changement très probable des paramètres du modèle. Après quelques mois de flottement, une fois accumulées les nouvelles données, les modèles économétriques peuvent recommencer à donner des prédictions correctes, mais les valeurs de leurs paramètres ont changé. On a eu « rupture de modèle ». De telles ruptures peuvent parfois être discernées assez rapidement, mais jamais instantanément. Quelques mois sont nécessaires en économie pour admettre, avec une forte probabilité, qu'il y a eu un changement de modèle; une fois cela admis, il ne reste plus qu'à rebâtir le modèle de prévision. Dans le cas de l'I.N.S.E.E. en 1992, la rapidité n'a évidemment pas été suffisante pour que les erreurs de prévision n'aient pas été récupérées par le débat politique.

(Source : Dacunha-Castelle 'Chemins de l'aléatoire, le hasard et le risque dans la société moderne')

¹¹Voir à ce sujet: J. Beran, *Statistics for long-memory processes*, Chapman et Hall, 1994.

¹²Ces modèles autosimilaires découverts par Paul Ikey ont un caractère fractal. Ils ont été popularisés par Benoit Mandelbrot.

3 Introduction à la notion de "mémoire longue"

Les processus $ARMA$ sont souvent appelé "*processus à mémoire courte*" compte tenu de la décroissance rapide de l'autocorrélation entre X_t et X_{t-h} quand $h \rightarrow \infty$ (*décroissance exponentielle de $\rho(h)$*). En fait, il est possible de montrer que, pour un processus $ARMA$ stationnaire, sa fonction d'autocorrélation est bornée,

$$|\rho(h)| \leq Cr^h \text{ pour } h = 1, 2, \dots$$

pour une certaine constante $C > 0$ et $0 < r < 1$.

Définition 10 *Un processus stationnaire sera dit à **mémoire longue** si les autocorrélations $\rho(h)$ satisfont*

$$\rho(h) \sim Ch^{2d-1} \text{ quand } h \rightarrow \infty \quad (21)$$

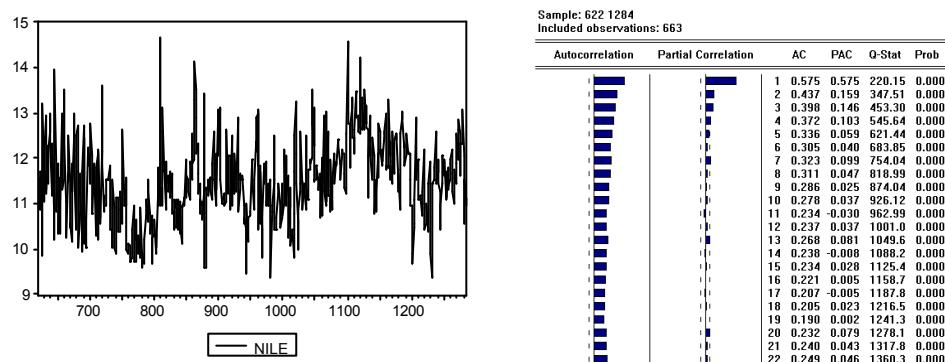
où C est une constante non nulle et $d < 1/2$.

Il est possible de faire une distinction suivant la valeur de d :

- si $d < 0$: mémoire intermédiaire, la série $\rho(h)$ est absolument convergente $\sum |\rho(h)| < \infty$.

- si $0 < d < 1/2$: mémoire longue, la série $\rho(h)$ n'est plus absolument convergente $\sum |\rho(h)| = \infty$

Ce type de comportement a été mis en évidence dès 1951 sur des données hydrologiques, puis économiques. Le graphique ci-dessous à gauche correspond au minimum annuel atteint par le Nil, et à droite, son autocorrélogramme,



Remarque 15 L'expression (21) dit uniquement la forme asymptotique des autocorrélations, sous forme d'un équivalent, mais ne dit en aucun cas que certains retards spécifiques doivent être particulièrement importants.

Remarque 16 Cette condition (21) se réécrit également

$$\lim_{h \rightarrow \infty} \frac{\rho(h)}{Ch^{2d-1}} = 1.$$

De façon heuristique, une série à mémoire longue est caractérisée par une fonction d'autocorrélation qui décroît lentement lors que le retard h augmente. En terme de fréquence et de densité spectrale, le spectre de telles séries ont un pôle à la fréquence 0. Cette propriété est très fréquente sur les données économiques, à tel point que Granger avait parlé en 1966 de "*forme spectrale telle d'une variable économique*", comme le note Valérie Mignon dans "*Méthodes d'estimation de l'exposant de Hurst*". Historiquement, le premier processus à mémoire longue est le **mouvement brownien fractionnaire**, développé par Mandelbrot et Van Ness en 1968. Ce processus est caractérisé par un paramètre appelé exposant de Hurst, noté H , permettant de classer les séries temporelles en fonction de leur structure de dépendance : mémoire courte ou nulle, mémoire longue et positive (appelé *phénomène de persistance*), et anti-persistante. Les séries empiriques étant des processus en temps discret, Mandelbrot et Wallis ont défini, en 1969, l'analogue du mouvement brownien fractionnaire en temps discret, appelé **bruit gaussien fractionnaire**.

Une seconde classe de modèles, liés au bruit gaussien fractionnaire, est constitué des processus *ARFIMA* (*AutoRegressive Fractionnaly Integrated Moving Average*) parfois également appelés *FARIMA*. Ceux-ci constituent une généralisation des processus *ARIMA* (p, d, q) standards, dans lesquels le paramètre de différenciation d était un entier. Le paramètre d explique alors le comportement de long terme de la série, le comportement de court terme étant expliqué par les composantes autorégressives et moyenne-mobiles.

Si ces processus sont très présents en économie (comme le notait Granger), ils sont également présents en finance. En 1991, Lo a ainsi étudié les implications des processus à mémoire longue dans la théorie financière, et Mignon

a étudié en 1995 l'impact de ce comportement sur l'efficience des marchés. En effet, on dit que les marchés sont efficients (*efficient d'un point de vue informationnel*) si les cours des actifs côtés reflètent toute l'information disponible : il est alors impossible de prévoir les rentabilités futures à partir des rentabilités passées puisque toute l'information (*anticipée*) est incorporée dans les prix. Mais si les rentabilités ont une structure de dépendance à long terme, alors les observations présentent des autocorrelations significatives. La connaissance du passé fournit alors de l'information pour prévoir les valeurs futures. En fait, Samuelson a montré en 1965 que si un marché est efficient, le prix observé sur le marché correspond à la valeur fondamentale (*c'est à dire la somme actualisé des dividendes futurs anticipés par les agents*), et la mémoire longue impliquerait un écart durable entre le cours et la valeur fondamentale.

Remarque 17 Si la mémoire longue a été définie à partir de la fonction d'autocorrélation, elle peut aussi être définie à l'aide de la densité spectrale. Les processus ARMA, exemples classiques de processus à mémoire courte, la densité spectrale est positive et finie quand la fréquence tend vers 0. On dira qu'un processus (X_t) est à mémoire longue s'il existe une constante β telle que $0 < \beta < 1$ et une constante $C > 0$ telles que

$$\lim_{\omega \rightarrow 0} \frac{f(\omega)}{C\omega^{-\beta}} = 1,$$

où f est la densité spectrale du processus (X_t) . Ainsi, la densité spectrale présente un pôle à la fréquence 0.

3.1 Processus self-similaires

Définition 11 On dit qu'il processus est **self-similaire**, de paramètre de self-similarité H si, pour tout c , le processus (Y_{ct}) à la même distribution que le processus $(c^H Y_t)$, pour tout $t \geq 0$

Cette classe de processus a été introduite dès 1968 par Mandelbrot, même si, d'un point de vue probabiliste, cette propriété avait été étudiée dès 1940 par Kolmogorov.

Exemple 25 En temps continu, le mouvement brownien est un processus self-similaire de paramètre $H = 1/2$. Aussi, dans les modèles financiers de type Black et Scholes, où le rendement est modélisé par une équation de la forme $dR_t = \mu dt + \sigma dW_t$, où (W_t) est un mouvement brownien standard, on obtient, en notant $R(\tau)$ le rendement sur une période τ , que la distribution de $R(\tau)/\sqrt{\tau}$ est indépendante du temps τ : en terme financier, l'ordre de grandeur de la rentabilité d'un titre financier pendant une période τ est proportionnelle à la racine carrée de cette période. Une des applications a été que la volatilité sur 12 mois est égale à la volatilité sur 1 mois, multiplié par $\sqrt{12}$. La plupart de règles prudentielles imposées pour le contrôle des risques reposent sur ce genre d'approche permettant d'extrapoler le risque à long terme à partir du risque à court terme. Toutefois, il est apparu en pratique, dans les années 70 que le coefficient H était plutôt de l'ordre de 0.3. Ces observations ont mis en évidence des phénomènes dits de "retour à la moyenne" des rentabilités, introduisant la notion d'horizon à long terme sur les marchés.

3.1.1 Accroissements stationnaires de processus self-similaires

Soit (Y_t) un processus H -self-similaire, à accroissements stationnaires. Posons alors

$$X_t = Y_t - Y_{t-1} \text{ pour } t = 1, 2, \dots$$

Soit $\sigma^2 = V(X_t)$. Alors la fonction d'autocovariance est donnée par

$$\gamma(h) = \frac{\sigma^2}{2} \left[|h+1|^{2H} - 2|h|^{2H} + |h-1|^{2H} \right]. \quad (22)$$

3.2 Processus FARIMA - ARIMA Fractionnaires

Dans la partie sur les processus ARIMA, nous avions introduit un aspect tendanciel aux processus ARMA en retenant un polynôme autorégressif admettant 1 comme racine (éventuellement multiple). Ceci a conduit à considérer les processus écrits sous la forme

$$(1 - L)^d \Phi(L) X_t = \Theta(L) \varepsilon_t,$$

où d est un entier positif. Le fait que le paramètre d soit entier pose un certain nombre de problèmes, des problèmes d'identification par exemple, et c'est pour cela que d est estimé séparément des autres paramètres. Les processus FARIMA sont une généralisation où l'on considère d a priori quelconque (réel).

Définition 12 Un processus fractionnaire est un processus (X_t) , défini pour $t \geq 0$, satisfaisant une équation de la forme

$$(1 - L)^d \Phi(L) X_t = \Theta(L) \varepsilon_t \quad \text{pour } t \geq 0,$$

avec comme condition initiale $X_t = 0$ pour $t < 0$ et ε_t un bruit blanc pour $t \geq 0$, nul sinon, de variance finie σ^2 , et où $\Phi(L)$ et $\Theta(L)$ sont deux polynômes ayant leurs racines à l'extérieur du disque unité, et où d est réel, avec la convention

$$(1 - L)^d = 1 + \sum_{j=1}^{\infty} \frac{d(d-1)\dots(d-j+1)}{j!} (-1)^j L^j.$$

La définition de $(1 - L)^d$ est fondée sur le développement en série entière de la fonction puissance (qui peut se simplifier en introduisant la fonction Gamma) : soit

$$\Gamma(d) = \int_0^{+\infty} x^{d-1} \exp(-x) dx,$$

alors

$$(1 - L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(d+j)}{\Gamma(d).j!} L^j.$$

Propriété 7 Un processus fractionnaire admet une représentation moyenne mobile (infinie) de la forme

$$X_t = H(L) \varepsilon_t \quad \text{où } H(L) = \sum_{j=0}^{\infty} h_j L^j,$$

où les coefficients h_j vérifient asymptotiquement

$$h_j \sim \frac{\Theta(1)}{\Phi(1)\Gamma(d)} j^{d-1} \text{ quand } j \rightarrow \infty.$$

Définition 13 Cette série est

- asymptotiquement non-stationnaire si $\sum h_j^2 = +\infty$,
- asymptotiquement stationnaire si $\sum h_j^2 < +\infty$, et dans ce cas, on dira que
 - à mémoire courte si $\sum |h_j| < +\infty$,
 - à mémoire longue si $\sum |h_j| = \infty$,

Propriété 8 Les processus ARMA sont des processus asymptotiquement stationnaires à mémoire courte.

Compte tenu du comportement asymptotique des h_j , on a la propriété suivante,

Propriété 9 Le processus fractionnaire est asymptotiquement stationnaire si et seulement si $d < 1/2$. Il est à mémoire courte si $d \leq 0$ et $0 < d < 1/2$.

Remarque 18 Ces processus sont parfois appelés "processus intégrés d'ordre d ", et noté $I(d)$.

3.2.1 Processus fractionnaire sans composante AR et MA : processus FARIMA $(0, d, 0)$

Considérons un processus de la forme $(1 - L)^d X_t = \varepsilon_t$ avec $d < 1/2$. (ε_t) est ici un bruit blanc, et nous supposerons, dans un premier temps, qu'il est de variance finie σ^2 .

Exemple 26 Comme nous l'avons vu auparavant, les processus ARIMA $(0, 1, 0)$ - c'est à dire la marche aléatoire $X_t = X_{t-1} + \varepsilon_t$ - dans le cas où (ε_t) est un bruit blanc gaussien de variance σ^2 , peuvent être vus comme des discrétisations du mouvement brownien. De façon analogue, dans le cas où le bruit (ε_t) est gaussien, le processus (X_t) ARFIMA $(0, d, 0)$ est à rapprocher du **mouvement brownien fractionnaire** (noté FBM - fractional Brownian motion) : un processus (en temps continu) $B_H(t)$ est appelé FBM s'il est à trajectoires continues, telles que $B_H(t)$ soit gaussien pour tout t , $B_H(0) = 0$, $\mathbb{E}(B_H(t) - B_H(s)) = 0$ pour tout s, t , et

$$\text{cov}(B_H(t), B_H(s)) = \frac{\sigma^2}{2} \left[|t|^{2H} - |t-s|^{2H} + |s|^{2H} \right], \quad (23)$$

pour tout $0 < H < 1$. On peut noter que le cas $H = 1/2$ correspond au mouvement brownien. La version discrète d'un tel processus est alors appelé bruit blanc gaussien fractionnaire (FGN - fractional Gaussian noise). La relation (23) est d'ailleurs à rapprocher de (22).

La fonction d'autocovariance $\gamma(t, h) = \text{cov}(X_t, X_{t+h})$ converge vers

$$\gamma(h) = \sigma^2 \frac{\Gamma(h+d)}{\Gamma(h+1-d)} \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \text{ quand } t \rightarrow \infty,$$

où σ^2 est la variance du bruit blanc.

Remarque 19 La fonction d'autocovariance $\gamma(t, h)$ dépend de t : le processus n'est pas stationnaire (mais il est asymptotiquement stationnaire).

La vitesse de convergence de l'autocorrélogramme vers 0 est donnée, pour h grand, par

$$\gamma(h) \sim \sigma^2 h^{2d-1} \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \text{ ou } \rho(h) \sim \gamma_d \cdot h^{2d-1} \text{ quand } h \rightarrow \infty.$$

A titre de comparaison, pour les processus *ARMA*, nous avions vu que la vitesse de décroissance était de la forme exponentielle, c'est à dire beaucoup plus rapide.

Par défaut, nous considérerons les processus *FARIMA* Gaussien, c'est à dire que le bruit (ε_t) est supposé être un processus gaussien. Toutefois, toute une théorie existe pour des processus plus généraux, non-gaussiens, mais de variance finie, ou alors de variance infinie. Ce dernier cas est d'ailleurs relativement utilisé dans la pratique. Parmi ces familles, on peut considérer les distributions de Pareto (c'est à dire que $\mathbb{P}(\varepsilon > x) \sim Cx^{-\alpha}$) de paramètre $\alpha < 2$, ou les distributions α -stables (c'est à dire que leur fonction caractéristique est de la forme $\phi(t) = \mathbb{E}(e^{it\varepsilon}) = \exp(-\sigma^\alpha |x|^\alpha)$ où σ est un paramètre d'échelle) de paramètre $1 < \alpha \leq 2$ (le cas $\alpha = 2$ correspondant au cas gaussien)¹³.

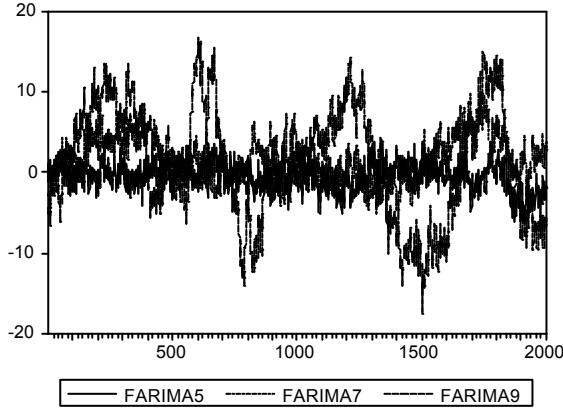
Les processus *ARFIMA* $(0, d, 0)$ peuvent s'écrire sous la forme

$$X_t = \sum_{k=0}^{+\infty} \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)} (-1)^k \varepsilon_{t-k} \text{ pour } d < 1/2,$$

dont la densité spectrale est donnée par

$$f_X(\omega) = \frac{\sigma^2}{|2\sin(\pi\omega)|^{2d}} \approx \frac{\sigma^2}{|2\pi\omega|^{2d}}.$$

Le graphique ci-dessous présente les trajectoires de 3 processus *FARIMA*, avec $d = 0.9, 0.7$ et 0.5 ,



Remarque 20 La densité spectrale de ces processus *ARIMA* $(0, d, 0)$ vérifie la relation suivante

$$f(\omega) = \left(2 \sin \frac{\omega}{2}\right)^{-2d} \text{ pour } 0 < \omega \pi, \quad (24)$$

de telle sorte que le comportement en 0 est donné par

$$\text{également } f(\omega) = \omega^{-2d} \text{ quand } \omega \rightarrow 0. \quad (25)$$

¹³De plus amples informations peuvent être trouvées dans Samorodnitsky et Taqqu (1994) 'Stable non-gaussian processes : stochastic models with infinite variance'

3.2.2 Autocorrélations des processus ARFIMA

L'écriture des autocorrélations se fait à l'aide de la forme MA du processus, si elle existe. Une autre possibilité, en utilisant la partie précédente est d'utiliser le paragraphe précédent et le résultat suivant, dû à Hosking (1981),

Propriété 10 Soit X_t un processus ARFIMA (p, d, q) , vérifiant $(1 - L)^d \Phi(L) X_t = \Theta(L) \varepsilon_t$. Les autocorrélations γ_k^X du processus X_t peuvent s'écrire en fonction des autocorrélations des deux processus

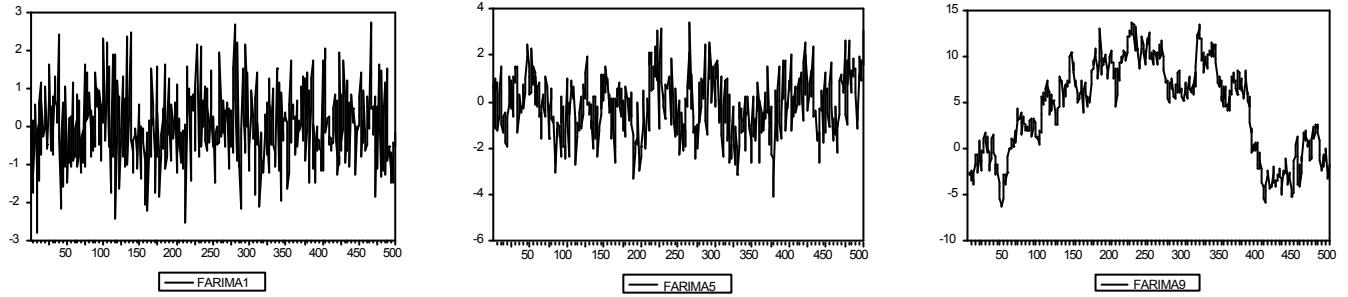
$$\begin{cases} Y_t = (1 - L)^d X_t & : \text{processus ARMA } (p, q) \\ Z_t = \Phi(L) \Theta(L)^{-1} X_t & : \text{processus ARFIMA } (0, d, 0). \end{cases}$$

Alors

$$\gamma_k^X = \sum_{i=-\infty}^{+\infty} \gamma_i^Y \gamma_{k-i}^Z.$$

Chacune des processus d'autocorrélations γ_i^Y et γ_{k-i}^Z pouvant se calculer aisément par des méthodes itératives.

Exemple 27 Considérons les simulations du processus ARFIMA $(0, d, 0)$ X_t défini par $(1 - L)^d X_t = \varepsilon_t$ où ε_t est un bruit blanc gaussien. Les graphiques ci-dessous représentent X_t dans le cas où $d = 0.1$ à gauche $d = 0.5$ au centre et $d = 0.9$ à droite



avec les autocorrélogrammes ($d = 0.3, d = 0.5, d = 0.9$) ci-dessous,



Les cas limites $d \rightarrow 0$ et $d \rightarrow 1$ correspondant respectivement au bruit blanc gaussien et à la marche aléatoire gaussienne (c'est à dire au processus intégré d'ordre 1 ARIMA $(0, 1, 0)$).

3.3 Estimation du paramètre d

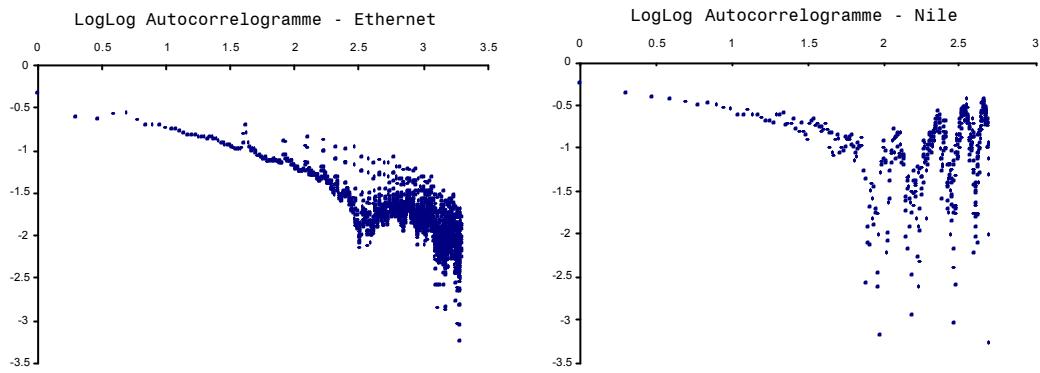
Parmi les méthodes pour estimer le paramètre d , nous en retenons 3.

3.3.1 Méthode du log-autocorrélogramme

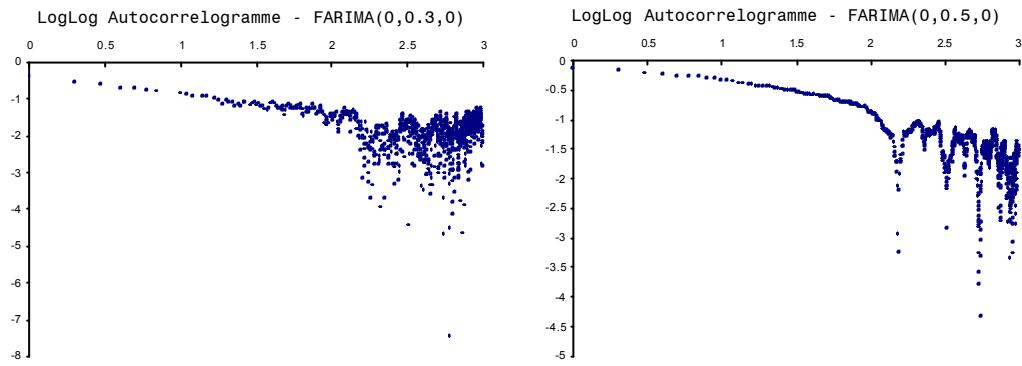
La méthode la plus simple pour détecter la présence de mémoire longue est l'utilisation du log-autocorrélogramme. Comme nous l'avons vu, asymptotiquement, les autocorrélations sont de la forme $\rho(h) \sim \gamma_d h^{2d-1}$. En prenant le logarithme, $\log \rho(h) \sim \kappa_d + (2d-1) \log h$, on peut noter une relation linéaire entre le logarithme des retards $\log h$ et le logarithme de l'autocorrélation $\log \rho(h)$. Le coefficient d peut alors être estimé en considérant la pente de la droite : soit $\hat{\alpha}$ et $\hat{\beta}$ tels que

$$\log |\hat{\rho}(h)| \sim \hat{\alpha} + \hat{\beta} \log h \text{ alors } \hat{d} = \frac{1}{2} (\hat{\beta} + 1).$$

Exemple 28 Les graphiques ci-dessous correspondent des log-log autocorrélogrammes de données internet à gauche (nombre de bytes échangés par seconde) et de la crue du Nil chaque année à droite



Si l'aspect linéaire apparaît difficilement sur des données réelles, il apparaît davantage sur des données simulées,



avec à gauche un processus ARFIMA $(0,0.3,0)$ et un processus ARFIMA $(0,0.5,0)$ à droite. On peut toutefois noter que les autocorrélations ne sont pas nulles au bout d'un temps relativement faible (<200) comme cela devrait être le cas pour les processus à mémoire courte.

3.3.2 Méthode du log-périodogramme

Soit (X_t) un processus stationnaire, dont la densité spectrale est de la forme $f_X(\lambda) = |2 \sin(\lambda/2)|^{-2d} f^*(\lambda)$ où $-0.5 < d < 0.5$ est le paramètre de mémoire et où f^* est une fonction continue, bornée sur $[-\pi, +\pi]$. Le périodogramme associé à un échantillon de taille T est défini par

$$I_T(\lambda) = \frac{1}{2\pi T} \left| \sum_{t=1}^T X_t e^{-i\lambda t} \right|^2.$$

On définit les fréquences de Fourier par $\lambda_j = 2\pi j/T$ pour $j = 1, \dots, (T-1)/2$. A partir de la densité spectrale et du périodogramme pris aux fréquences de Fourier, on obtient la régression suivante : $\log I_T(\lambda_j) = \alpha + \beta Z_j + \varepsilon_j$ où

$$\begin{cases} \alpha = \log f^*(0) - \xi \\ \beta = d \\ Z_j = -2 \log |2 \sin(\lambda_j/2)| \\ \varepsilon_j = \log(I_T(\lambda_j)) - \log f^*(\lambda_j) + \xi \end{cases} \quad \text{où } \xi \text{ est la constante d'Euler.}$$

Pour $0 \leq s < t < T$, on utilise l'estimateur

$$\hat{d}_R(s, t) = \frac{\sum_{j=s+1}^t (Z_j - \bar{Z}) \log(I_T(\lambda_j))}{\sum_{j=s+1}^t (Z_j - \bar{Z})^2} \quad \text{où } \bar{Z} \text{ est la moyenne des } Z_j.$$

3.3.3 Méthode de Whittle

Soit (X_t) un processus $FARIMA(0, d, 0)$ défini par $(1 - L)^d X_t = \varepsilon_t$ où (ε_t) est un bruit blanc de variance finie σ^2 . La méthode de Whittle revient à estimer le vecteur $\theta = (\sigma^2, d)'$ de paramètres. La fonction que l'on cherche à minimiser est appelée vraisemblance de Whittle, et est donnée par

$$\mathcal{L}_W(X, \theta) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \left(\log f_X(\lambda, \theta) + \frac{I_T(\lambda)}{f_X(\lambda, \theta)} \right) d\lambda,$$

où f_X et I_T sont respectivement la densité spectrale et le périodogramme du processus. En notant $\theta^* = (1, d)'$ et en choisissant σ^2 (*paramètre d'échelle*) tel que $f_X(\lambda, \theta) = \sigma^2 f_X(\lambda, \theta^*)$, et $\int_{-\pi}^{+\pi} \log f_X(\lambda, \theta^*) d\lambda = 0$, alors la vraisemblance de Whittle, $\mathcal{L}_W(X, \theta)$ s'écrit simplement sous la forme

$$\mathcal{L}_W^*(X, \theta) = \mathcal{L}_W(X, \theta^*) = \log \sigma^2 + \frac{1}{2\pi\sigma^2} \int_{-\pi}^{+\pi} \frac{I_T(\lambda)}{f_X(\lambda, \theta^*)} d\lambda.$$

L'estimateur de Whittle $\hat{\theta} = (\hat{\sigma}^2, \hat{d})'$ est alors obtenu en minimisant la fonction $\mathcal{L}_W^*(X, \theta)$ par rapport à θ^* . Puisque l'échantillon est de taille finie, cela revient à minimiser par rapport à θ^* la fonction $Q_T(\theta^*)$ définie par

$$Q_T(\theta^*) = \sum_{i=1}^{(T-1)/2} \frac{I_T(\lambda_i)}{f_X(\lambda_i, \theta^*)} \text{ et } \hat{\sigma}^2 \text{ vaut alors } \frac{4\pi}{T} Q_T(\theta^*).$$

3.4 Exemples d'applications des processus à mémoire longue

Le développement rapide des séries longues a été principalement dû à des questions posées en hydrologie. La résolution à l'aide de modèle statistique "à mémoire longue" a commencé des 1951, suit à un article de Hurst sur le stockage d'eau par les barrages.

Ce phénomène de mémoire longue se manifeste par la présence de cycles de toutes périodicités, qui se rencontrent dans un grand nombre de phénomènes physiques. Ainsi, en 1975, Mandelbrot a appliqué cette théorie aux communication, Halett et Rafterey (en 1989) sur de la météorologie, Ogata et Abe (en 1991) en sismologie...etc. Des tentatives ont également été faites en économie, en particulier pour faire de la prévision à moyen ou long terme.

3.4.1 Applications en finance : rendements d'indices boursier

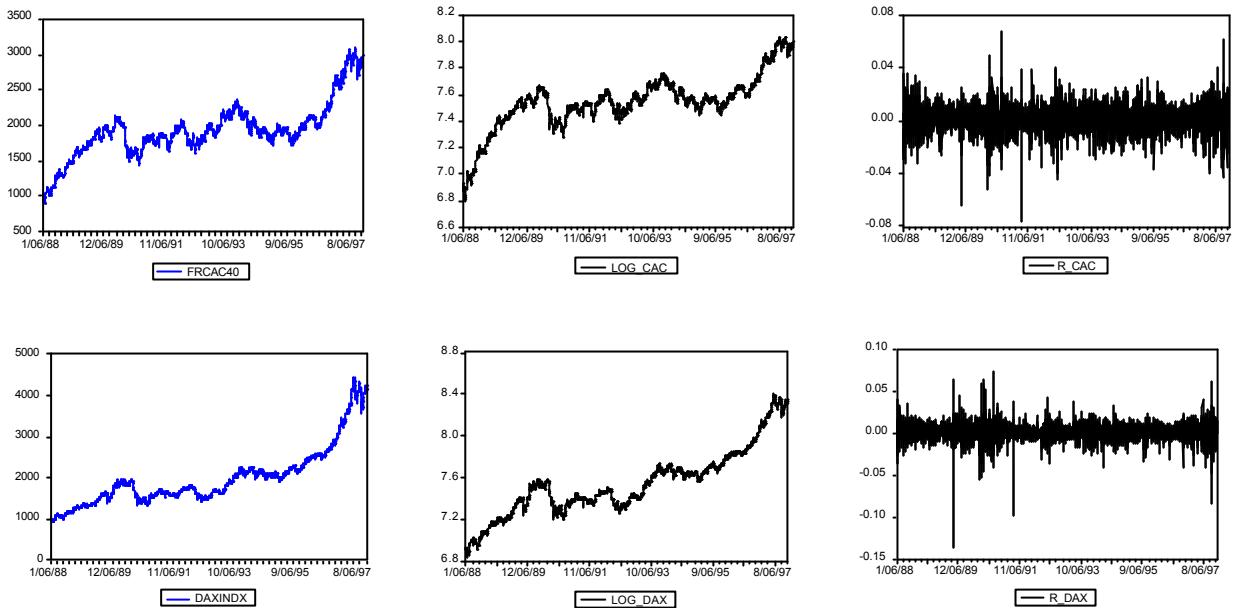
Les prix d'actifs financiers (*taux de change, cours boursiers... etc*) sont des séries qui présentent des comportement de type hétérosédastique, avec très souvent de la persistance. En outre, la plupart du temps, les distributions des prix ne sont pas normales. A partir des années 80 (*et de l'article de Engle en 1982*), une large classe de modèles à temps discret a été proposée : *ARCH, GARCH, EGARCH, TARCH, IGARCH... etc.*

Une des caractéristiques des modèles est la prise en compte explicite de l'hypothèse d'hétérosédasticité dans les séries observées. Néanmoins, leur fonction d'autocorrélation est semblable à celle d'un bruit blanc ou d'un processus à mémoire courte : la fonction d'autocorrélation d'un modèle *ARCH* décroît exponentiellement vers 0. Pour h grand, l'autocorrélation entre X_t et X_{t-h} est alors supposée très faible.

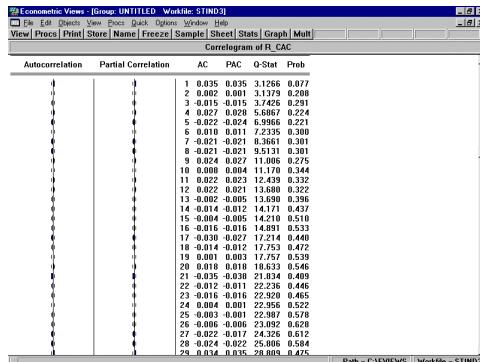
A partir des années 90, de nouveaux modèles, permettant la prise en compte des hypothèses d'hétérosédasticité et de mémoire longue ont été proposés. En 1986, Taylor remarquait le premier que les valeurs absolues des rendements des indices boursiers ont une fonction d'autocorrélation qui décroît très lentement. En 1993, Ding, Granger et Engle observent le même genre de comportement en prenant une puissance quelconque (*en prenant le carré par exemple*) du rendement. Plusieurs méthodes ont été proposées pour estimer le coefficient d sur des données financières.

La méthode du log-périodogramme a été proposée par Geweke et Porter-Hudak en 1983, afin d'estimer l'ordre d , en régressant le logarithme du périodogramme sur les basses fréquences. Cet estimateur, noté d_{GPH} , semble toutefois fortement biaisé et non-asymptotiquement normal pour $-0.5 < d < 0.5$. Cet estimateur sera amélioré par Robinson en 1995. La méthode de Whittle proposée en 1951 repose sur l'approximation de la matrice de variance-covariance du processus. Cet estimateur, noté d_W , est non-biaisé et sa distribution asymptotique est normale (*comme Fox et Taqqu l'ont montré en 1986 dans le cas des processus gaussiens, puis généralisé par Giraitis et Surgalis en 1990*).

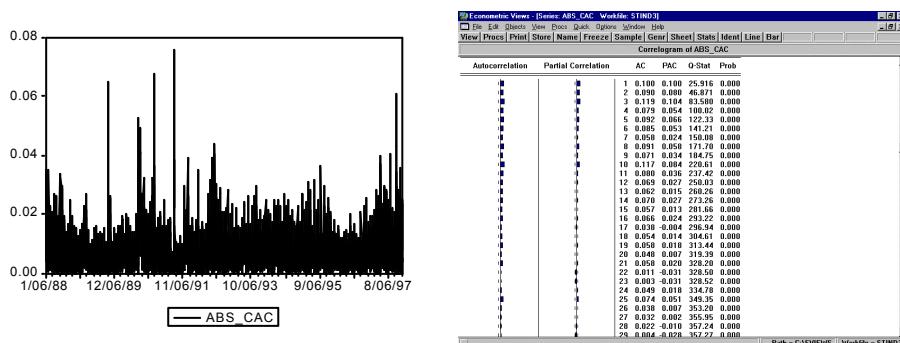
Rendements d'indices boursiers R_t Considérons le processus (X_t) , indice boursier, et posons (R_t) le rendement associé, défini par $R_t = \log X_t - \log X_{t-1}$. Une analyse rapide montre que, s'il y a des phases de fortes volatilités, les séries de rendement sont généralement stationnaires. Les graphiques ci-dessous représentent les indices CAC40 (en haut) et DAX30 (en bas), avec, de gauche à droite, la série de l'indice (X_t) , le logarithme de l'indice $(\log X_t)$, et enfin, le rendement (R_t) ,



La sortie ci-dessous montre l'autocorrélogramme dans le cas de l'indice CAC40



Etude de la valeur absolue du rendement $|R_t|$ La série représentée ci-dessous correspond à la valeur absolue du rendement pour l'indice CAC40



On observe effectivement une décroissance relativement lente de l'autocorrélogramme¹⁴. Un modèle de type FARIMA pourrait modéliser le comportement de cette série. Sur la période allant de Janvier 1992 à Décembre 2000, les résultats

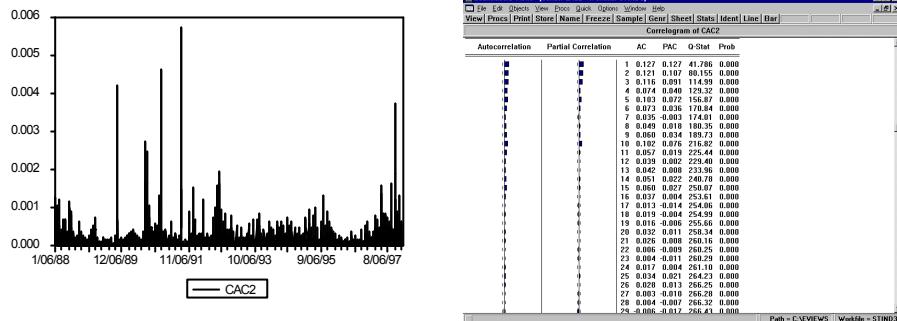
¹⁴ $\gamma(25) = 0.074, \gamma(38) = 0.059, \gamma(69) = 0.053, \gamma(86) = 0.062 \dots$ etc. Un grand nombre d'autocorrelations sont encore significatives ($\gamma(179), \gamma(183), \gamma(185)$ par exemple sont encore significativement non nulles).

suivants sont obtenus, par la méthode de Whittle,

	<i>SP500</i>	<i>CAC40</i>	<i>DAX30</i>
\hat{d}_W	0.149	0.126	0.178
\mathcal{L}_W	-14.13	-13.71	-13.59

où \mathcal{L}_W désigne la valeur de la vraisemblance de Whittle, définie dans la partie précédante. Les coefficients estimés sont significativement positifs, ce qui confirme l'hypothèse de mémoire lonue persistante.

Etude du carré du rendement R_t^2 La série représentée ci-dessous correspond au carré du rendement pour l'indice *CAC40*



De façon analogue à la valeur absolue du rendement, on obtient

	<i>SP500</i>	<i>CAC40</i>	<i>DAX30</i>
\hat{d}_W	0.170	0.154	0.162
\mathcal{L}_W	-20.95	-20.21	-19.76

Etude par sous-périodes En étudiant par sous-périodes, on peut noter, par exemple pour l'indice *CAC40*, que le comportement "mémoire longue" dépend de la période considérée : avec ci-dessous le comportement de $|R_t|$

	87/89	90/94	95/00	87/00
\hat{d}_W	0.202	0.118	0.126	0.150
\mathcal{L}_W	-13.27	-13.84	-13.71	-13.59

suivi du comportement de R_t^2

	87/89	90/94	95/00	87/00
\hat{d}_W	0.182	0.134	0.166	0.170
\mathcal{L}_W	-18.59	-20.31	-20.31	-19.66

Remarque 21 Comme nous l'avions rappelé au début de cette partie, la notion d'efficience de marché est très liée à l'absence de mémoire : l'hypothèse d'efficience est alors associée au modèle de marche aléatoire. Les cours boursiers suivent une marche aléatoire, et les processus de rentabilité sont alors des bruits blancs : le prix observé sur le marché fluctue de façon aléatoire autour de sa valeur fondamentale. Cette absence de mémoire étant empiriquement une hypothèse trop restrictive, Samuelson a proposé en 1965 de remplacer le modèle de marche aléatoire par un modèle de martingale : ce modèle est alors moins restrictif puisqu'aucune condition n'est alors imposée sur les autocorrelations des résidus.

En fait, Fama a montré en 1970 puis en 1991 que la mémoire courte ne remet pas en cause l'hypothèse d'efficience puisque le fait que quelques autocorrelations soient significatives à court terme ne peut pas être utilisé pour spéculer. En revanche, la présence d'une mémoire longue pose d'avantage de problèmes. En 1986, Summers a étudié en détail le phénomène de retour à la moyenne (mean reversion) des prix : à la suite d'un choc, le prix sécarde de sa valeur fondamentale mais finit toujours par y revenir. Ce phénomène implique alors l'existence possible d'un écart entre le cours et la valeur fondamentale. Toutefois, si l'écart est durable, cela peut être traduit comme la présence d'une mémoire longue. Ce phénomène va alors à l'encontre de la propriété d'efficience de Samuelson : plus une série sera persistante, plus il sera possible d'établir une stratégie rémunératrice sur les marchés, ce qui remet fortement en cause l'efficience de ces marchés.

3.4.2 Applications en finance : taux de change

Cette étude a été faite dans la thèse de Christelle Lecourt (*'Les variations de taux de change au jour le jour : une approche économétrique à partir des processus à mémoire longue'*).

De la même façon que pour les rendements d'indice boursier, la représentation *ARIMA* a été la première modélisation retenue. Toutefois, compte tenu des caractéristiques de la plupart des séries financières (*en particulier la présence d'une dépendance temporelle dans la variance conditionnelle*) a conduit à modéliser les seconds moments (*et moments plus élevés*). La classe la plus utilisée avait trait aux modèles d'hétéroscléasticité conditionnelle autorégressive : les modèles *ARCH* (*puis les modèles GARCH*). Toutefois, ces modèles n'ont jamais permis de modéliser correctement les taux de change.

Dans les années 90, les modèles *FARIMA* ou *ARFIMA* introduit par Granger en 1980 ont été utilisés afin de mieux prendre en compte la dépendance temporelle dans les premiers moments.

[A COMPLÉTER]

3.4.3 Applications en hydrologie

Comme l'avait fait Hurst, considérons une suite d'observations (X_t) (*par exemple la quantité d'eau déchargée par une rivière dans un réservoir, le jour t*), et définissons la suite (X_t^*) correspondant au cumul de ces (X_t)

$$X_0^* = 0 \text{ et } X_t^* = \sum_{i=1}^t X_i$$

Soit d un entier, et considérons $R(d)$, étendue de (X_t) , ajustée sur l'intervalle $[0, d]$, c'est à dire

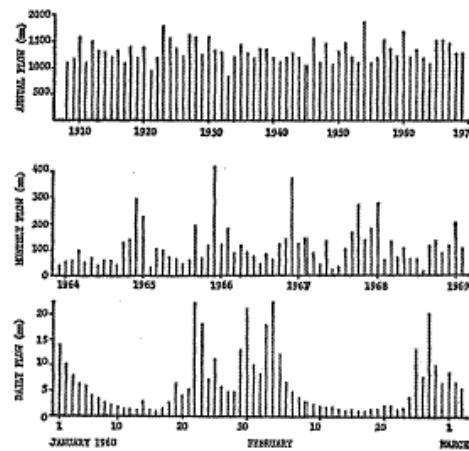
$$R(d) = \max_{0 \leq u \leq d} \left[X_u^* - \frac{u}{d} X_d^* \right] - \min_{0 \leq u \leq d} \left[X_u^* - \frac{u}{d} X_d^* \right]$$

Dans notre exemple (X_t) est la quantité d'eau déchargée par une rivière dans un réservoir, le jour t , et (X_t^*) la quantité totale d'eau déversée jusqu'au jour t . A un horizon d , la quantité moyenne d'eau arrivée par jour est X_d^*/d . Aussi, si la quantité (X_u^*) est la quantité d'eau arrivée jusqu'en u , la quantité uX_d^*/d est la quantité d'eau qui aurait du arriver "en moyenne". Supposons qu'un barrage permette un écoulement constant, à un débit journalier de X_d^*/d . Pour être certain de ne pas avoir d'inondation, la quantité $R(d)$ est le volume du réservoir. Il est également possible de définir $S(t)$ de telle sorte que

$$S^2(t) = \frac{1}{d} \sum_{i=1}^d X_i^2 - \frac{1}{d^2} \left(\sum_{i=1}^d X_i \right)^2$$

La statistique dite R/S a été introduite par Hurst pour étudier les crues du Nil. Cette statistique est liée aux processus self-similaires puisque, si X_t^* est self-similaire alors la statistique $R(d)$ est égale, en distribution, à $d^H R(1)$.

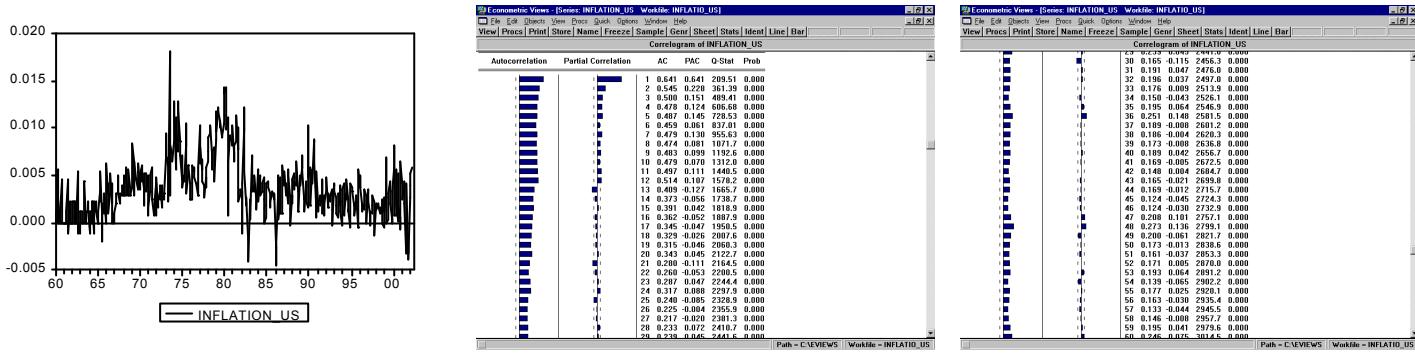
Le graphique ci-dessous est extrait de 'Stochastic modelling of riverflow time series', de Lawrence et Kottekoda (1977) :



On y voit un débit annuel (*en haut*), mensuel (*au milieu*) et journalier (*en bas*).

3.4.4 Applications en économie

Les processus à mémoire longue peuvent être utiles pour modéliser, par exemple, des données économiques, comme l'inflation. Dans cette partie, nous considérerons le taux d'inflation aux Etats Unis. Les courbes ci-dessous montrent le taux d'inflation (*variation de l'indice par rapport au mois précédent*) à gauche, et les fonctions d'autocorrélation à droite



Comme on peut le voir sur l'autocorrélogramme, les fonctions d'autocorrélations sont toutes significativement non nulles même après 60 retards (5 ans puisque les données sont mensuelles). Le modèle testé ici est de la forme suivante

$$(1 - \phi L^{12}) (1 - L)^d (X_t - [c + \delta_{1980-07}]) = (1 - \theta L^{12}) \varepsilon_t,$$

où δ_t est une fonction *dummy* permettant de modérer une chute brutale de l'inflation en Juillet 1980. Différentes méthodes peuvent être utilisées pour estimer les paramètres : *EML* (), *MPL* (), ou *NLS* (), donnant les résultats suivants sur la période 1959 – 01 / 1995 – 12,

	<i>EML</i>	<i>MPL</i>	<i>NLS</i>
\hat{d}	0.400 (0.033)	0.411 (0.034)	0.409 (0.036)
$\hat{\phi}$	0.769 (0.101)	0.783 (0.091)	0.620 (0.110)
$\hat{\theta}$	-0.605 (0.123)	-0.621 (0.113)	-0.447 (0.121)
\hat{c}	0.003 (0.002)	0.003 (0.003)	0.04 (0.002)
$\hat{\delta}$	-0.011 (0.002)	-0.011 (0.002)	-0.011 (0.002)
$\hat{\sigma}^2$	3.626	3.625	3.548

On peut noter que pour chacune des méthodes $d < 0.5$.

3.5 Conseils bibliographiques

Sur ce sujet, et pour une application des modèles *ARFIMA* (*et des processus GARMA*) sur des données réelles, je renvoie au mémoire de Bouette, JC., Chassagneux, J.F, Sibai, D. et Terron, R. (2003) sur la modélisation du vent en ireland¹⁵.

- AVOUYI-DOVIS., GUEGAN,D. & LADOUCETTE,S. (2000) Application des processus de longue mémoire à l'analyse des indices boursiers
- BERAN,J. (1994). Statistics for long-memory processes. Chapman & Hall.
- JASIAK,J. (1999). Long memory in economics : a comment. Working Paper - York University

¹⁵Ce mémoire est téléchargeable à <http://www.crest.fr/pageperso/lfa/charpent/statapp2003.pdf.pdf>

4 Compléments : exercices

4.1 Exercices avec correction

Exercice 1 Soit (ε_t) un bruit blanc. Les processus suivants sont-ils stationnaires ? (éventuellement avec certaines restrictions sur les paramètres)

- (i) $X_t = a + b\varepsilon_t + c\varepsilon_{t-1}$ où $a, b, c \in \mathbb{R}$
- (ii) $X_t = \varepsilon_t \cos(ct) + \varepsilon_{t-1} \sin(ct)$ où $c \in \mathbb{R}$
- (iii) X_t tel que $X_t - X_{t-1} = \varepsilon_t$

(i) Le processus X_t vérifie $\mathbb{E}(X_t) = a = \text{constante}$. Sa variance est donnée par $V(X_t) = [b^2 + c^2]\sigma^2$ où σ^2 est la variance du bruit, puisque le bruit est non-autocorrélé : $\text{cov}(\varepsilon_t, \varepsilon_{t-1}) = 0$.

De plus, $\text{cov}(X_t, X_{t-1}) = bc\sigma^2$ qui ne dépend pas de t et $\text{cov}(X_t, X_{t-h}) = 0$ pour tout $h > 1$. Aussi, le processus X_t est stationnaire.

(ii) Le processus X_t vérifie $\mathbb{E}(X_t) = 0$ (= constante). De plus, sa variance est donnée par $V(X_t) = [\cos^2(ct) + \sin^2(ct)]\sigma^2$ où σ^2 est la variance du bruit, puisque le bruit est non-autocorrélé : $\text{cov}(\varepsilon_t, \varepsilon_{t-1}) = 0$.

Enfin, $\text{cov}(X_t, X_{t-1}) = \sigma^2[\sin ct \cos c(t-1)]$ qui dépend de t pour $c \notin \pi\mathbb{Z}$. Aussi, le processus X_t n'est pas stationnaire

(iii) Le processus X_t peut se réécrire

$$X_t = \sum_{k=1}^t \varepsilon_k + X_0$$

qui correspond à une marche aléatoire. Aussi, le processus X_t n'est pas stationnaire. On peut montrer que le processus explose en \sqrt{t} .

Exercice 2 Soit (ε_t) un bruit blanc et (X_t) un processus vérifiant

$$X_t - \frac{7}{2}X_{t-1} + \frac{3}{2}X_{t-2} = \varepsilon_t \text{ pour tout } t \in \mathbb{Z} \quad (26)$$

Montrer qu'il existe un processus stationnaire, solution de (26) sous la forme

$$\tilde{X}_t = \sum_{k \in \mathbb{Z}} a_k \varepsilon_{t-k}$$

En déduire que ε_t n'est pas l'innovation du processus \tilde{X}_t .

Le polynôme caractéristique de la forme AR est $\Phi(Z) = 1 - 7Z/2 + 3Z^2/2$, qui peut s'écrire

$$\Phi(Z) = 1 - \frac{7}{2}Z + \frac{3}{2}Z^2 = (1 - 3Z)\left(1 - \frac{1}{2}Z\right)$$

Les racines sont alors $1/3$ et 2 . La racine $1/3$ n'est pas à l'extérieur du disque unité : la représentation n'est pas canonique. Toutefois, il est possible d'inverser le polynôme retard Φ . Soit \tilde{X}_t est défini par

$$\begin{aligned} \tilde{X}_t &= \Phi^{-1}(L)\varepsilon_t = \frac{1}{(1-3L)(1-\frac{1}{2}L)}\varepsilon_t = \left[\frac{6/5}{1-3L} - \frac{1/5}{1-L/2} \right] \varepsilon_t \\ &= -\frac{2}{5} \underbrace{\sum_{k=1}^{\infty} \left(\frac{1}{3}\right)^{k-1} \varepsilon_{t+k}}_{\substack{\text{dépend du futur} \\ \text{de } \varepsilon_t}} - \frac{1}{5} \underbrace{\sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k \varepsilon_{t-k}}_{\substack{\text{dépend du passé} \\ \text{de } \varepsilon_t}} \text{ soit } \sum_{i=-\infty}^{+\infty} \alpha_i \varepsilon_{t-i} \end{aligned}$$

Le processus ε_t est alors corrélé avec $\tilde{X}_{t-1}, \tilde{X}_{t-2}, \dots$ (puisque \tilde{X}_t s'exprime en fonction du passé et du futur de ε_t) : le processus ε_t n'est pas l'innovation.

En revanche, en considérant un bruit blanc de la forme $\eta_t \sim BB(1, \sigma^2/9)$ (le coefficient $1/9$ venant de $(1/3)^2$, $1/3$ étant la racine appartenant au disque unité), on peut mettre \tilde{X}_t sous forme AR,

$$\left(1 - \frac{1}{2}L\right) \left(1 - \frac{1}{3}L\right) \tilde{X}_t = \eta_t \text{ soit } \tilde{X}_t - \frac{5}{6}\tilde{X}_{t-1} + \frac{1}{6}\tilde{X}_{t-2} = \eta_t \text{ où } \eta_t \sim BB(1, \sigma^2/9)$$

Cette écriture correspondra à l'écriture canonique du processus \tilde{X}_t .

Exercice 3 On considère le processus aléatoire AR (2) suivant

$$X_t = 40 + 0.4X_{t-1} - 0.2X_{t-2} + \varepsilon_t \text{ où } \varepsilon_t \sim BB(0, \sigma^2 = 12.8)$$

- (i) vérifier que le processus est stationnaire
- (ii) calculer l'espérance de X_t
- (iii) donner les équations de Yule-Walker du processus, calculer la variance, ainsi que les 5 premières valeurs des autocorrelations
- (iv) calculer les 3 premières autocorrelations partielles

(i) Le processus peut s'écrire $(1 - 0.4L + 0.2L^2) X_t = 40 + \varepsilon_t$. Le processus est stationnaire si les racines du polynôme sont à l'extérieur du disque unité :

$$1 - 0.4Z + 0.2Z^2 = 0 \text{ssi } Z = \frac{0.4 \pm i\sqrt{0.64}}{0.4} \text{ c'est à dire } Z = \begin{cases} Z = 1 + 2i \\ Z = 1 - 2i \end{cases}$$

dont le module est $\sqrt{1+2^2} = \sqrt{5} > 1$: le processus est stationnaire.¹⁶

(ii) Le processus est un AR (2) avec constante : il n'est pas centré.

$$\mathbb{E}(X_t) = 0.4\mathbb{E}(X_{t-1}) - 0.2\mathbb{E}(X_{t-2}) + 40 + 0 \text{ soit } \mu = \mathbb{E}(X_t) = 40/0.8 = 50$$

(iii) Soit $Y_t = X_t - 50$. Alors Y_t satisfait $Y_t = 0.4Y_{t-1} - 0.2Y_{t-2} + \varepsilon_t$ avec $\mathbb{E}(Y_t) = 0$. La fonction d'autocovariance de Y_t (qui est la même que celle de X_t) est obtenue de la façon suivante

$$\mathbb{E}(Y_t Y_{t-h}) = 0.4\mathbb{E}(Y_{t-1} Y_{t-h}) - 0.2\mathbb{E}(Y_{t-2} Y_{t-h}) + \mathbb{E}(\varepsilon_t Y_{t-h})$$

ce qui donne l'équation de Yule Walker $\gamma_h = 0.4\gamma_{h-1} - 0.2\gamma_{h-2}$. Les initialisations sont obtenue par les relations

$$\begin{cases} \gamma_1 = 0.4\gamma_0 - 0.2\gamma_1 \\ \gamma_2 = 0.4\gamma_1 - 0.2\gamma_0 \end{cases}$$

La première équation permet d'écrire $\gamma_1 = 0.4\gamma_0/1.2 = \gamma_0/3$, et la seconde $\gamma_2 = -0.2\gamma_0/3$. Or γ_0 vérifie

$$\gamma_0 = \mathbb{E}(Y_t Y_t) = 0.4\mathbb{E}(Y_{t-1} Y_t) - 0.2\mathbb{E}(Y_{t-2} Y_t) + \mathbb{E}(\varepsilon_t Y_t) = 0.4\gamma_1 - 0.2\gamma_2 + \mathbb{E}(\varepsilon_t Y_t)$$

Et comme $\varepsilon_t Y_t = 0.4\varepsilon_t Y_{t-1} - 0.2\varepsilon_t Y_{t-2} + \varepsilon_t^2$, on en déduit que $\mathbb{E}(\varepsilon_t Y_t) = \sigma^2$. Et donc $\gamma_0 = 0.4\gamma_1 - 0.2\gamma_2 + \sigma^2$. D'où finallement, par substitution

$$\gamma_0 = 0.4 \times \gamma_0/3 - 0.2 \times (-0.2\gamma_0/3) + \sigma^2 \text{ et donc } \gamma_0 = 3\sigma^2/2.56 = 15$$

D'où finalement

$$\begin{cases} \gamma_1 = \gamma_0/3 = 5 \\ \gamma_2 = -0.2\gamma_0/3 = -1 \\ \gamma_3 = 0.4 \times -1 - 0.2 \times 5 = -1.4 \\ \gamma_4 = 0.4 \times -1.4 - 0.2 \times -1 = 0.36 \\ \gamma_5 = 0.4 \times 0.36 - 0.2 \times -1.4 = 0.136 \end{cases} \text{ et donc } \begin{cases} \rho_1 = 0.333 \\ \rho_2 = -0.068 \\ \rho_3 = -0.093 \\ \rho_4 = -0.024 \\ \rho_5 = 0.009 \end{cases}$$

(iv) Les autocorrelations partielles se calculent à l'aide des déterminants des matrices d'autocorrelations :

$$\alpha(1) = \frac{|\rho_1|}{|1|} = 0.33$$

$$\alpha(1) = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = -0.199$$

¹⁶Ceci pouvait également se noter directement en notant que les conditions de stationnarité dans le cas d'un AR (2) s'écrivent

$$\begin{cases} \phi_1 + \phi_2 < 1 \\ -\phi_1 + \phi_2 < 1 \\ \phi_2 > -1 \end{cases}$$

$$\alpha(1) = \frac{1}{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \\ \hline 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} = -0.00069$$

On peut noter que $\alpha(3)$ est très faible, ce qui confirme la nature $AR(2)$ du modèle (*en théorie, les autocorrélations partielles d'un processus $AR(p)$ sont nulles au delà du rang p*).

Exercice 4 Considérons deux processus (X_t) et (Y_t) liés par les relations suivantes

$$\begin{cases} Y_t = \phi Y_{t-1} + \alpha X_t + \varepsilon_t \\ X_t = \theta X_{t-1} + \eta_t \end{cases}$$

où ε_t et η_t sont des bruits blancs non-corrélés, de variance respective σ_ε^2 et σ_η^2 . Pour les applications pratiques, on prendra

$$\alpha = 1.5, \phi = 0.4, \theta = 0.6, \sigma_\varepsilon^2 = 0.016 \text{ et } \sigma_\eta^2 = 0.036$$

(i) on pose $\omega_t = (1 - \phi L)(1 - \theta L)Y_t$. Calculer les moments de ω_t et en déduire qu'il s'agit d'un processus moyenne mobile (à préciser). Donner la représentation canonique de ω_t

(ii) en déduire que Y_t est un processus ARMA dont on précisera les ordres. Donner la représentation canonique de Y_t

(i) Le processus ω_t est défini par $\omega_t = (1 - \phi L)(1 - \theta L)Y_t$. X_t et Y_t étant centrés, on peut déjà noter que $\mathbb{E}(\omega_t) = 0$.

$$\begin{aligned} (1 - \theta L)Y_t &= Y_t - \theta Y_{t-2} = \phi Y_{t-1} + \alpha X_t + \varepsilon_t - \theta [\phi Y_{t-2} + \alpha X_{t-1} + \varepsilon_{t-1}] \\ &= \phi [Y_{t-1} - \theta Y_{t-2}] + \alpha [X_t - \theta X_{t-1}] + \varepsilon_t - \theta \varepsilon_{t-1} \\ &= \phi [Y_{t-1} - \theta Y_{t-2}] + \alpha \eta_t + \varepsilon_t - \theta \varepsilon_{t-1} = \phi [1 - \theta L]Y_{t-1} + \alpha \eta_t + \varepsilon_t - \theta \varepsilon_{t-1} \end{aligned}$$

Ainsi $(1 - \theta L)Y_t = \phi [1 - \theta L]LY_t + \alpha \eta_t + \varepsilon_t - \theta \varepsilon_{t-1}$, soit $(1 - \theta L)(1 - \phi L)Y_t = \omega_t = \alpha \eta_t + \varepsilon_t - \theta \varepsilon_{t-1}$. Le processus ω_t vérifie alors,

$$\begin{cases} \gamma(0) = \mathbb{E}(\omega_t^2) = \alpha^2 \sigma_\eta^2 + (1 + \theta^2) \sigma_\varepsilon^2 = 0.10276 \\ \gamma(1) = \mathbb{E}(\omega_t \omega_{t-1}) = -\theta \sigma_\varepsilon^2 = -0.0096 \\ \gamma(h) = 0 \text{ pour } h \geq 2 \end{cases}$$

Ce comportement de l'autocorrélogramme est caractéristique des processus $MA(1)$ (*nullité des autocorrélations au delà du rang 1*) : ω_t est un processus moyenne-mobile.

Cherchons à écrire ω_t sous forme $MA(1)$: $\omega_t = u_t - \beta u_{t-1}$ où u_t est un bruit blanc de variance σ^2 . Les paramètres β et σ^2 doivent alors vérifier

$$\begin{cases} \gamma(0) = (1 + \beta^2) \sigma^2 \\ \gamma(1) = -\beta \sigma^2 \end{cases}$$

c'est à dire que β doit vérifier

$$-\frac{\beta}{1 + \beta^2} = \frac{\gamma(1)}{\gamma(0)} = \frac{-\theta \sigma_\varepsilon^2}{\alpha^2 \sigma_\eta^2 + (1 + \theta^2) \sigma_\varepsilon^2} = 0.0934$$

β doit alors vérifier $0.0934\beta^2 - \beta + 0.0934 = 0$ soit $\beta = 0.095$ ou $\beta = 10.61$. On choisit la racine de module inférieur à 1 pour la représentation canonique (β étant l'inverse de la racine du polynôme caractéristique). ω_t satisfait $\omega_t = u_t - \beta u_{t-1}$ où u_t est un bruit blanc de variance $\sigma^2 = 0.101$ et où $\beta = 0.095$.

(ii) D'après ce que nous venons de montrer $\omega_t = (1 - \theta L)(1 - \phi L)Y_t = (1 - \beta L)u_t$ où u_t est un bruit blanc, c'est à dire que Y_t suite un processus $ARMA(2,1)$.

La forme développée de la dynamique de Y_t est

$$Y_t = (\phi + \theta)Y_{t-1} - \phi\theta Y_{t-2} + u_t - \beta u_{t-1} \quad (27)$$

Il est également possible de calculer les autocorrélations $\gamma(h)$ du processus : pour cela, on va multiplier cette expression par X_{t-h} , puis prendre l'espérance (*le processus Y_t étant centré*¹⁷) :

$$\begin{aligned} Y_t^2 &= (\phi + \theta)Y_{t-1}Y_t - \phi\theta Y_{t-2}Y_t + u_t Y_t - \beta u_{t-1} Y_t \\ &= (\phi + \theta)Y_{t-1}Y_t - \phi\theta Y_{t-2}Y_t + u_t [(\phi + \theta)Y_{t-1} - \phi\theta Y_{t-2} + u_t - \beta u_{t-1}] \\ &\quad - \beta u_{t-1} [(\phi + \theta)Y_{t-1} - \phi\theta Y_{t-2} + u_t - \beta u_{t-1}] \end{aligned}$$

¹⁷Ce qui implique que $\text{cov}(X_t, X_{t-h}) = \mathbb{E}(X_t X_{t-h}) - \mathbb{E}(X_t)\mathbb{E}(X_{t-h}) = \mathbb{E}(X_t X_{t-h})$ pour tout h . En particulier $V(X_t) = \mathbb{E}(X_t^2)$.

(on remplace Y_t par (27) dans $u_t Y_t$ et $u_{t-1} Y_t$). En prenant l'espérance, on obtient

$$\gamma(0) = (\phi + \theta)\gamma(1) - \phi\theta\gamma(2) + \sigma^2 [1 - \beta(\phi + \theta) + \beta^2]$$

De plus, en multipliant (27) par Y_{t-1} on peut écrire

$$\begin{aligned} Y_t Y_{t-1} &= (\phi + \theta) Y_{t-1} Y_{t-1} - \phi\theta Y_{t-2} Y_{t-1} + u_t Y_{t-1} - \beta u_{t-1} Y_{t-1} \\ &= (\phi + \theta) Y_{t-1} Y_{t-1} - \phi\theta Y_{t-2} Y_{t-1} + u_t Y_{t-1} - \beta u_{t-1} [(\phi + \theta) Y_{t-2} - \phi\theta Y_{t-3} + u_{t-1} - \beta u_{t-2}] \end{aligned}$$

(u_t est indépendant du passé de Y_t , en particulier indépendant de Y_{t-1}) d'où, en prenant l'espérance

$$\gamma(1) = (\phi + \theta)\gamma(0) - \phi\theta\gamma(1) - \beta\sigma^2$$

En multipliant (27) par Y_{t-2} on a

$$\gamma(2) = (\phi + \theta)\gamma(1) - \phi\theta\gamma(0)$$

et de façon plus générale

$$\gamma(h) = (\phi + \theta)\gamma(h-1) - \phi\theta\gamma(h-2)$$

Exercice 5 Décomposition de Beveridge-Nelson d'un processus intégré

On s'intéresse dans cet exercice à la décomposition d'un processus ARIMA, intégré d'ordre 1, en la somme d'une marche aléatoire, et d'un processus stationnaire.

(1) Soit Y_t admettant la décomposition $(1 - L)Y_t = \Theta(L)\varepsilon_t$ avec

$$\Theta(L) = \sum_{k=0}^{+\infty} \theta_k L^k \text{ où } \theta_0 = 1 \text{ et } \sum_{k=0}^{+\infty} |\theta_k| < +\infty$$

et ε_t bruit blanc de variance σ^2 . Posons alors $\theta = \Theta(1)$.

(1 - i) Soit Θ^* tel que $\Theta(L) = \theta + (1 - L)\Theta^*(L)$. Donner l'expression des coefficients θ_k^* du polynôme Θ^* .

(1 - ii) Montrer que si $\sum k|\theta_k| < +\infty$ alors $\sum |\theta_k^*| < +\infty$.

(1 - iii) Montrer que l'on peut décomposer Y_t sous la forme $Y_t = T_t + C_t$ où $(1 - L)T_t = \theta\varepsilon_t$ et où C_t est un processus stationnaire.

(1 - iv) Quelle est la limite de $V(Y_t)/t$ quand $t \rightarrow \infty$?

(1 - v) Considérons une décomposition quelconque du processus Y_t , $Y_t = \bar{T}_t + \bar{C}_t$ en une somme d'une marche aléatoire T_t et d'un processus stationnaire C_t . Montrer qu'alors, on a nécessairement

$$V((1 - L)\bar{T}_t) = \theta^2\sigma^2$$

(2) On supposera désormais que le processus Y_t vérifie une équation de la forme

$$(1 - L)(1 - \phi L)Y_t = \varepsilon_t \text{ où } 0 < \phi < 1 \quad (28)$$

(2 - i) Ecrire la relation précédente sous la forme $(1 - L)Y_t = \Theta(L)\varepsilon_t$ (Forme de Wold)

(2 - ii) Décomposer le processus Y_t en somme d'une marche aléatoire et d'un processus stationnaire. Les processus C_t et T_t sont-ils indépendants ?

(3) On veut montrer ici qu'il est impossible d'écrire le processus Y_t sous la forme $Y_t = T_t + C_t$ où $(1 - L)T_t = u_t$ et $C_t = \Theta(L)v_t$ où les processus u_t et v_t sont des bruits blancs indépendants, et où $\Theta(L)$ est un polynôme en L dont les racines sont de module strictement supérieur à 1.

(3 - i) En utilisant (28), calculer la densité spectrale de $X_t = (1 - L)Y_t$

(3 - ii) En supposant que $Y_t = T_t + C_t$ où $(1 - L)T_t = u_t$ et $C_t = \Theta(L)v_t$, où les processus u_t et v_t sont des bruits blancs indépendants, calculer la densité spectrale de $X_t = (1 - L)Y_t$

(3 - iii) Etudier les comportements pour $\omega = 0$ des deux densités spectrales trouvées en (3 - i) et en (3 - ii).

(3 - iv) Conclure.

(1 - i) Par hypothèse, $(1 - L)Y_t = \Theta(L)\varepsilon_t$. Soit alors $X_t = (1 - L)Y_t$. Le polynôme Θ vérifie

$$\Theta(L) = \sum_{k=0}^{+\infty} \theta_k (L^k - 1) + \sum_{k=0}^{+\infty} \theta_k = \sum_{k=0}^{+\infty} \theta_k (L^k - 1) + \Theta(1)$$

Or la différence $L^k - 1$ peut se réécrire

$$L^k - 1 = (L - 1)[1 + L + \dots + L^{k-1}]$$

et donc

$$\begin{aligned}\Theta(L) &= \Theta(1) + (1-L) \sum_{k=0}^{+\infty} \theta_k [-(1+L+\dots+L^{k-1})] \\ &= \Theta(1) - (1-L) \sum_{k=0}^{+\infty} \left(\sum_{j=k+1}^{+\infty} \theta_j \right) L^k\end{aligned}$$

D'où finallement l'écriture $\Theta(L) = \theta + (1-L)\Theta^*(L)$ où

$$\Theta^*(L) = \sum_{k=0}^{+\infty} \theta_k^* L^k \quad \text{avec } \theta_k^* = - \sum_{j=k+1}^{+\infty} \theta_j$$

(1 - ii) En utilisant cette expression,

$$\sum_{k=0}^{+\infty} \theta_k^* = \sum_{k=0}^{+\infty} \sum_{j=k+1}^{+\infty} \theta_j = [\theta_1 + \theta_2 + \theta_3 + \dots] + [\theta_2 + \theta_3 + \dots] + [\theta_3 + \dots] + \dots = \sum_{k=0}^{+\infty} j \theta_j$$

et

$$\sum_{k=0}^{+\infty} |\theta_k^*| \leq \sum_{k=0}^{+\infty} k |\theta_k|$$

Donc finallement si $\sum k |\theta_k| < +\infty$ alors $\sum |\theta_k^*| < +\infty$.

(1 - iii) D'après la question précédente, on peut écrire $X_t = \Theta(L) \varepsilon_t = \theta \varepsilon_t + (1-L)\Theta^*(L) \varepsilon_t$
Si $Y_t = T_t + C_t$, alors $X_t = (1-L)Y_t = (1-L)T_t + (1-L)C_t$. Aussi, on devrait avoir

$$\begin{cases} (1-L)T_t = \theta \varepsilon_t \\ C_t = \Theta^*(L) \varepsilon_t \end{cases}$$

- Montrons que le processus C_t est stationnaire :

$$C_t = \Theta^*(L) \varepsilon_t = \sum_{k=0}^{+\infty} \theta_k^* \varepsilon_{t-k}$$

qui sera stationnaire si et seulement si $\sum |\theta_k^*| < +\infty$, qui sera vérifié dès lors que $\sum k |\theta_k| < +\infty$.

- Le processus T_t définie par $(1-L)T_t = \theta \varepsilon_t$ est une marche stationnaire :

$$T_t = T_{t-1} + \theta \varepsilon_t = \theta \cdot \sum_{k=0}^t \varepsilon_{t-k} + T_{-1}$$

(1 - iv) La variance de Y_t peut se décomposer en

$$V(Y_t) = V(C_t) + 2\text{cov}(C_t, T_t) + V(T_t)$$

avec $V(C_t) = \text{constante}$ et $V(T_t)$ est équivalent à $\theta^2 \sigma_\varepsilon^2 t$, car $T_t = T_0 + \theta [\varepsilon_1 + \dots + \varepsilon_t]$. De plus,

$$|\text{cov}(T_t, C_t)| \leq \sqrt{V(T_t)} \sqrt{V(C_t)} = O(\sqrt{t})$$

et finalement, on a

$$\frac{V(Y_t)}{t} \rightarrow \theta^2 \sigma_\varepsilon^2 \text{ quand } t \rightarrow \infty$$

En effectuant exactement le même genre de calcul sur $Y_t = \bar{T}_t + \bar{C}_t$, p, a

$$V(Y_t) = V(\bar{C}_t) + 2\text{cov}(\bar{C}_t, \bar{T}_t) + V(\bar{T}_t)$$

et on obtient alors

$$\lim_{t \rightarrow \infty} \frac{V(Y_t)}{t} = \lim_{t \rightarrow \infty} \frac{V(\bar{T}_t)}{t} = \theta^2 \sigma_\varepsilon^2$$

D'où finallement de résultat souhaité, $V(\Delta \bar{T}_t) = \theta^2 \sigma_\varepsilon^2$.

PARTIE 2

(3 - i et ii) Le processus Y_t se décompose en $Y_t = T_t + C_t$, avec $(1 - L) Y_t = u_t$ et $C_t = B(L) v_t$. Aussi, en différenciant Y_t , on a

$$(1 - L) Y_t = (1 - L) T_t + (1 - L) C_t = u_t + (1 - L) B(L) v_t = \Theta(L) \varepsilon_t$$

où $\Theta(L) \varepsilon_t = u_t + (1 - L) B(L) v_t$. Le processus $X_t = (1 - L) Y_t = \Theta(L) \varepsilon_t$ admet pour densité spectrale

$$f(\omega) = \frac{1}{2\pi} |\Theta(e^{i\omega})|^2 \sigma_\varepsilon^2 = \frac{\sigma_u^2}{2\pi} + \frac{1}{2\pi} |(1 - e^{i\omega}) B(e^{i\omega})|^2 \sigma_v^2$$

en utilisant l'écriture $\Theta(L) \varepsilon_t = u_t + (1 - L) B(L) v_t$

(3 - iii) En comparant les deux écritures pour $\omega = 0$ on constate alors que

$$\frac{\sigma_u^2}{2\pi} = \frac{\sigma_\varepsilon^2}{2\pi} |\Theta(1)|^2 \text{ soit } \frac{\sigma_u^2}{\sigma_\varepsilon^2} = |\Theta(1)|^2$$

Et comme $|(1 - e^{i\omega}) B(e^{i\omega})|^2 \sigma_v^2 \geq 0$, alors $f(\omega) \geq f(0) = \sigma_u^2/2\pi$: f admet un minimum global en $\omega = 0$.

(3 - iv) Soit $(1 - L)(1 - \phi L) Y_t = \varepsilon_t$ pour $0 < \phi < 1$. Alors $(1 - \phi L) X_t = \varepsilon_t$ en posant $X_t = (1 - L) Y_t$, on peut écrire

$$f(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{1}{|1 - \phi e^{i\omega}|^2} \text{ et } f(0) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{1}{(1 - \phi)^2}$$

Montrons qu'alors 0 n'est pas un minimum global : si $\omega = \pi/2$ alors $|1 - \phi e^{i\omega}|^2 = 1 + \phi^2 > (1 - \phi)^2$ puisque $0 < \phi < 1$, et donc

$$\frac{1}{(1 - \phi)^2} > \frac{1}{1 + \phi^2} \text{ et donc } f(0) > f\left(\frac{\pi}{2}\right)$$

(Y_t) ne peut donc pas être décomposée en tendance - cycle, d'après le modèle de la question (2). Par contre il peut l'être d'après le modèle (1).

4.2 Examen de 2001/2002

Exercice 6 (i) Soit (ε_t) un bruit blanc et X_t défini par

$$X_t = \sum_{k=0}^t \lambda^k (\varepsilon_{t-k} - \varepsilon_{t-k-1}) \text{ où } \lambda \in \mathbb{R}$$

Le processus X_t est-il stationnaire ? (éventuellement avec certaines restrictions sur les paramètres)

(ii) Soit (ε_t) un bruit blanc, et X_t , Y_t et Z_t défini par $X_t = \varepsilon_t$, $Y_t = (-1)^t \varepsilon_t$ et $Z_t = X_t + Y_t$ pour tout $t \in \mathbb{Z}$. Les processus X_t , Y_t et Z_t sont-ils stationnaires ?

(iii) La somme de deux processus stationnaires est-elle stationnaire ?

(i) Le processus X_t est un processus centré, $\mathbb{E}(X_t) = 0$. Le processus X_t peut s'écrire

$$X_t = \lambda^0 [\varepsilon_t - \varepsilon_{t-1}] + \lambda^1 [\varepsilon_{t-1} - \varepsilon_{t-2}] + \lambda^2 [\varepsilon_{t-2} - \varepsilon_{t-3}] + \dots + \lambda^{t-1} [\varepsilon_1 - \varepsilon_0] + \lambda^t [\varepsilon_0 - 0]$$

avec la convention $\varepsilon_{-1} = 0$. Cette relation peut se réécrire, en mettant en facteur les ε_{t-i} ,

$$\begin{aligned} X_t &= \lambda^0 \varepsilon_t + [\lambda^1 - \lambda^0] \varepsilon_{t-1} + [\lambda^2 - \lambda^1] \varepsilon_{t-2} + [\lambda^3 - \lambda^2] \varepsilon_{t-3} + \dots + [\lambda^{t-1} - \lambda^{t-2}] \varepsilon_1 + [\lambda^t - \lambda^{t-1}] \varepsilon_0 \\ &= \lambda^0 \varepsilon_t + (\lambda - 1) [\varepsilon_{t-1} + \lambda \varepsilon_{t-2} + \lambda^2 \varepsilon_{t-3} + \lambda^3 \varepsilon_{t-4} + \dots + \lambda^{i-1} \varepsilon_{t-i} + \dots + \lambda^{t-2} \varepsilon_1 + \lambda^{t-1} \varepsilon_0] \end{aligned}$$

Aussi, pour $\lambda = 1$, alors $X_t = \varepsilon_t$ et le processus est stationnaire.

Pour $\lambda \neq 1$, on peut écrire

$$X_t = \varepsilon_t + (\lambda - 1) \sum_{k=0}^{t-1} \lambda^k \varepsilon_{t-k-1}$$

Aussi,

$$V(X_t) = \sigma^2 + (\lambda - 1)^2 \sum_{k=0}^{t-1} \lambda^{2k} \sigma^2 = \sigma^2 + (\lambda - 1)^2 \frac{1 - \lambda^{2t}}{1 - \lambda^2} \sigma^2$$

qui n'est pas indépendante de t : X_t n'est pas un processus stationnaire.

(ii) Le processus X_t est un bruit blanc donc X_t est stationnaire.

Le processus Y_t vérifie $\mathbb{E}(Y_t) = 0$ (=constante), et $\gamma(0) = \mathbb{E}(Y_t^2) = \mathbb{E}(\varepsilon_t^2) = \sigma^2$ (=constante). Enfin, pour tout $h \neq 0$, $\gamma(h)$ vérifie

$$\gamma(h) = \mathbb{E}(Y_t Y_{t-h}) = \mathbb{E}\left((-1)^{2t-h} \varepsilon_t \varepsilon_{t-h}\right) = \pm \mathbb{E}(\varepsilon_t \varepsilon_{t-h}) = 0$$

Le processus Y_t est stationnaire (et c'est même un bruit blanc).

Le processus Z_t est défini de la façon suivante

$$Z_t = \begin{cases} 2\varepsilon_t & \text{pour } t \text{ pair} \\ 0 & \text{pour } t \text{ impair} \end{cases}$$

Si $\mathbb{E}(Z_t) = 0$ (=constante), on peut noter que la variance, elle, ne sera pas constante : $\mathbb{E}(Z_t^2) = 4\sigma^2$ pour t pair, et $\mathbb{E}(Z_t^2) = 0$ pour t impair (*car Z_t^2 est alors une constante*). La variance de Z_t dépend de t : le processus Z_t n'est pas stationnaire.

(iii) Dans la question précédente, nous avions vu que X_t et Y_t étaient des processus stationnaire, et pourtant, leur somme Z_t n'est pas un processus stationnaire : **la somme de deux processus stationnaires n'est pas forcément stationnaire**.

Exercice 7 Soit ε_t un bruit blanc, centré de variance $5/18$, et considérons le processus Y_t défini par

$$Y_t = 2Y_{t-1} + \varepsilon_t$$

On suppose que Y_t n'est observable qu'avec une erreur d'observation : on ne peut observer que le processus $X_t = Y_t + \eta_t$ où η_t est un bruit blanc non corrélé avec ε_t , de variance $1/6$ (avec de plus $\text{cov}(\varepsilon_t, \eta_{t-h}) = 0$ pour tout $h \in \mathbb{Z}$)

(i) montrer que le processus $\omega_t = \varepsilon_t + \eta_t - 2\eta_{t-1}$ est un processus MA, et en déduire que X_t est un processus ARMA(1,1) que l'on précisera,

(ii) donner la représentation canonique de X_t ,

(iii) en déduire une représentation de X_t du type

$$X_t = -\sum_{i=1}^{\infty} \alpha_i X_{t-i} + u_t \quad \text{avec } \sum_{i=1}^{\infty} |\alpha_i| < \infty \tag{29}$$

et où u_t est un bruit blanc que l'on explicitera (donner sa volatilité), non corrélé avec X_t . A quoi sert une telle représentation ?

(i) Soit $\omega_t = \varepsilon_t + \eta_t - 2\eta_{t-1}$. Ce processus est centré : $\mathbb{E}(\omega_t) = 0 + 0 - 0 = 0$. Sa variance est donnée par

$$\begin{aligned} \gamma(0) &= \mathbb{E}(\omega_t^2) = \mathbb{E}([\varepsilon_t + \eta_t - 2\eta_{t-1}] [\varepsilon_t + \eta_t - 2\eta_{t-1}]) \\ &= \mathbb{E}(\varepsilon_t^2 + \eta_t^2 + 4\eta_{t-1}^2 + 2\varepsilon_t \eta_t - 4\eta_t \eta_{t-1} - 4\varepsilon_t \eta_{t-1}) \\ &= \sigma_\varepsilon^2 + 5\sigma_\eta^2 + 0 + 0 + 0 = \sigma_\varepsilon^2 + 5\sigma_\eta^2 \end{aligned}$$

car $\mathbb{E}(\varepsilon_t \eta_t) = \mathbb{E}(\varepsilon_t \eta_{t-1}) = 0$ (*les processus sont indépendants*) et $\mathbb{E}(\eta_t \eta_{t-1}) = 0$ (*car η_t est un bruit blanc*). Donc $\mathbb{E}(\omega_t^2) = \sigma_\varepsilon^2 + 5\sigma_\eta^2$ qui est une constante : $\gamma(0) = 5/18 + 5/6 = 20/18 = 10/9$. De plus

$$\begin{aligned} \gamma(1) &= \mathbb{E}(\omega_t \omega_{t-1}) = \mathbb{E}([\varepsilon_t + \eta_t - 2\eta_{t-1}] [\varepsilon_{t-1} + \eta_{t-1} - 2\eta_{t-2}]) \\ &= \mathbb{E}(\varepsilon_t \varepsilon_{t-1} + \varepsilon_t \eta_{t-1} - 2\varepsilon_t \eta_{t-2} + \eta_t \varepsilon_{t-1} + \eta_t \eta_{t-1} - 2\eta_t \eta_{t-2} - 2\eta_{t-1} \varepsilon_{t-1} - 2\eta_{t-1} \eta_{t-1} + 4\eta_{t-1} \eta_{t-2}) \\ &= 0 + 0 - 0 + 0 - 0 - 0 - 2\sigma_\eta^2 + 0 = -2\sigma_\eta^2 \end{aligned}$$

d'où $\gamma(1) = -2/6 = -1/3$. Enfin, on peut montrer que $\gamma(2) = 0$ et, plus généralement, $\gamma(h) = 0$ pour $h \geq 2$. Ce processus ω_t vérifie

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = -\frac{1}{3} \frac{9}{10} = -\frac{3}{10} \text{ et } \rho(h) = 0 \text{ pour } h \geq 2$$

Le processus ω_t est un processus MA(1). Aussi, il existe un bruit blanc ν_t et un paramètre θ tel que

$$\omega_t = \nu_t - \theta \nu_{t-1}$$

Le processus X_t vérifie

$$\begin{aligned} X_t &= Y_t + \eta_t = [2Y_{t-1} + \varepsilon_t] + \eta_t = 2[X_{t-1} - \eta_{t-1}] + \varepsilon_t + \eta_t \\ &= 2X_{t-1} + [\varepsilon_t + \eta_t - 2\eta_{t-1}] = 2X_{t-1} + \omega_t \end{aligned}$$

donc, d'après la question précédente, il existe bruit blanc ν_t et un paramètre θ tel que $X_t = 2X_{t-1} + \nu_t - \theta\nu_{t-1}$, c'est à dire que X_t est un processus *ARMA*(1, 1).

(ii) Pour expliciter l'écriture de ce processus *ARMA*, il convient de donner deux paramètres : θ (*composante moyenne-mobile*) et σ_ν^2 , la variance du nouveau bruit blanc. Pour identifier ces deux paramètres, on utilise les deux équations que nous avions obtenues auparavant. Pour le processus $\omega_t = \nu_t - \theta\nu_{t-1}$, on a

$$\begin{cases} \gamma(0) = \mathbb{E}(\omega_t^2) = [1 + \theta^2]\sigma_\nu^2 = 10/9 \\ \gamma(1) = \mathbb{E}(\omega_t\omega_{t-1}) = -\theta\sigma_\nu^2 = -1/3 \end{cases}$$

En divisant la première équation par la seconde, on obtient que

$$\frac{1 + \theta^2}{\theta} = \frac{10}{9} \frac{3}{1} = \frac{10}{3}$$

ce qui donne l'équation de degré 2 suivante, $3\theta^2 - 10\theta + 3 = 0$, dont les racines sont

$$\theta = \frac{10 \pm \sqrt{10^2 - 4 \times 3 \times 3}}{2 \times 3} = \frac{10 \pm \sqrt{64}}{6} = \frac{10 \pm 8}{6} = \begin{cases} 3 \\ 1/3 \end{cases}$$

Pour mettre le processus sous forme canonique **on choisit θ de telle sorte que la racine du polynôme moyenne-mobile ($\Theta(L) = 1 - \theta L$) est à l'extérieur du disque unité**, c'est à dire $\theta < 1$. On choisit $\theta = 1/3$. Cette valeur permet d'en déduire la variance du bruit blanc associé : $-\theta\sigma_\nu^2 = -1/3$ donc $\sigma_\nu^2 = 1$.

La forme canonique du processus est alors

$$X_t = 2X_{t-1} + \nu_t - \frac{1}{3}\nu_{t-1} \text{ où } \nu_t \sim BB(0, 1)$$

(iii) Nous avons écrit le processus sous la forme $\Phi(L)X_t = \Theta(L)\nu_t$. Afin d'obtenir une représentation de la forme (29), il convient d'inverser le polynôme retard $\Theta(L)$: $\Theta^{-1}(L)\Phi(L)X_t = \nu_t$. Puisque nous avons choisi la racine de Θ à l'extérieur du disque unité, le polynôme $\Theta(L)$ est inversible en fonction des opérateurs passés, et

$$\Theta^{-1}(L)\Phi(L) = \frac{1}{1 - \theta L} (1 + \phi L) = (1 + \phi L) \sum_{i=0}^{\infty} \theta^i L^i = 1 + \sum_{i=1}^{\infty} \underbrace{[\theta^{i-1}\phi + \theta^i]}_{\alpha_i} L^i$$

Avec cette écriture, on obtient finalement

$$X_t = - \sum_{i=1}^{\infty} \alpha_i X_{t-i} + \nu_t \text{ avec } \begin{cases} \alpha_i = \theta^{i-1}\phi + \theta^i = 2/3^{i-1} + 1/3^i \\ \nu_t \sim BB(0, 1) \end{cases}$$

Cette écriture est utile pour faire de la prévision : la meilleure prévision possible, faite à la date T pour un horizon h est

$$T\widehat{X}_{T+h}^* = - \sum_{k=1}^{h-1} \alpha_{k,T} \widehat{X}_{T+h-k}^* - \sum_{k=h}^{T+h} \alpha_k X_{T+h-k}$$

Remarque 22 L'exercice suivant utilisait les sorties informatiques disponibles sur <http://www.crest.fr/pageperso/lfa/cha>

Exercice 8 Considérons la série brute X_t , série trimestrielle observée de 1980 à 2000. Nous noterons dans toute la suite

- X_t la série brute
- Y_t la série différenciée une fois $Y_t = \Delta X_t = (X_t - X_{t-1})$
- Z_t la série différenciée deux fois $Z_t = \Delta^2 X_t = \Delta Y_t = X_t - 2X_{t-1} + X_{t-2}$
- $Y2_t$ la série $Y2_t = Y_t - Y_{t-2}$ et $Z2_t$ la série $Z2_t = Z_t - Z_{t-2}$
- $Y4_t$ la série $Y4_t = Y_t - Y_{t-4}$ et $Z4_t$ la série $Z4_t = Z_t - Z_{t-4}$

A partir de l'ensemble des séries informatiques suivantes (courbes, histogrammes, tests, estimation d'équations...etc), le but est ici de trouver un (ou plusieurs) modèle permettant de modéliser au mieux la série X_t .

- (i) quelle(s) série(s) peut-on essayer de modéliser à l'aide d'un processus ARMA ?
(ii) les modèles suivants sont-ils valides ?

$$(1 - L)(1 - \alpha_1 L - \alpha_2 L^2 - \alpha_3 L^3 - \alpha_4 L^4 - \alpha_5 L^5 - \alpha_6 L^6 - \alpha_7 L^7) X_t = \varepsilon_t \text{ où } \varepsilon_t \text{ est un bruit blanc} \quad (30)$$

$$(1 - L)(1 - \alpha L^2) X_t = \varepsilon_t \text{ où } \varepsilon_t \text{ est un bruit blanc} \quad (31)$$

$$(1 - L)(1 - L^2)(1 - \alpha L^2) X_t = \varepsilon_t \text{ où } \varepsilon_t \text{ est un bruit blanc} \quad (32)$$

$$(1 - L)(1 - L^4) X_t = \varepsilon_t - \theta \varepsilon_{t-1} \text{ où } \varepsilon_t \text{ est un bruit blanc} \quad (33)$$

- (iii) les méthodes SCAN et ESCF suggèrent-elles de 'bons' modèles ? si oui, lesquels ?

- (iv) est-il possible de modéliser la série X_t sans composante MA ?

- (v) la série X_t a en fait été simulée sur 80 valeurs, à l'aide d'un bruit blanc gaussien $N(0, 1)$, de telle sorte que

$$(1 - L)(1 - L^4)(1 - 0.8L) X_t = (1 - 0.6L^2) \varepsilon_t$$

Les simulations correspondent-elles effectivement à un tel modèle ? Existe-t-il de 'meilleurs' modèles, ou d'autres modèles permettant de modéliser la série X_t ?

(i) Si l'on regarde la série brute X_t , nous sommes en présence d'une série non-stationnaire : nous n'observons pas la décroissance exponentielle de l'autocorrélogramme, que l'on pourrait observer sur un processus AR, et les 20 premières autocorrelations sont significatives : X_t est non-stationnaire, avec présence d'une racine unité.

La série différenciée (*une fois*) Y_t semble stationnaire, avec toutefois présence d'une racine unité saisonnière : toutes les autocorrelations obtenues avec un retard pair sont significativement non nulle. Y_t n'est pas stationnaire, avec présence d'une racine unité saisonnière.

La série $(1 - L^2) Y_t$ présente le même genre de comportement que Y_t , c'est donc que 2 n'était pas la fréquence de la racine unité. Y_{2t} n'est pas stationnaire. En revanche la série Y_{4t} est stationnaire. Les autres séries proposées étant obtenues à partir d'une série (Z_t) différenciée davantage, nous n'allons pas les prendre en compte. De toutes façons, un raisonnement analogue aurait poussé à ne retenir que Z_{4t} , qui est la seule série stationnaire. Or $Y_{4t} = (1 - L)(1 - L^4) X_t$ et $Z_{4t} = (1 - L)^2(1 - L^4) X_t = (1 - L)Y_{4t}$: puisque Y_{4t} est déjà stationnaire, il est inutile de différencier davantage. Nous allons modéliser $Y_{4t} = (1 - L)(1 - L^4) X_t$.

Ceci est validé par les tests de Dickey-Fuller.

- (ii) L'estimation du modèle (30) - modèle AR(7) pour Y_t - défini par

$$(1 - L)(1 - \alpha_1 L - \alpha_2 L^2 - \alpha_3 L^3 - \alpha_4 L^4 - \alpha_5 L^5 - \alpha_6 L^6 - \alpha_7 L^7) X_t = \varepsilon_t \text{ où } \varepsilon_t \text{ est un bruit blanc}$$

est donnée dans le slide (50). Si tous les paramètres semblent significatifs, on peut noter que l'estimation aboutit à une racine unité (1.01 est racine du polynôme AR, et Eviews précise 'Estimated AR process is nonstationary') : ce modèle ne peut pas être retenu.

L'estimation du modèle (31) - modèle AR(2) pour Y_t - défini par

$$(1 - L)(1 - \alpha L^2) X_t = \varepsilon_t \text{ où } \varepsilon_t \text{ est un bruit blanc}$$

est donnée dans le slide (67). Le paramètre α est significativement non nul. Toutefois, si l'on considère l'autocorrélogramme des résidus obtenus après régression (slide (69)), on peut noter que l'on n'a pas un bruit blanc : toutes les autocorrelations paires sont non nulles. Ce modèle ne peut être retenu.

L'estimation du modèle (32) - modèle AR(2) pour Y_{2t} - défini par

$$(1 - L)(1 - L^2)(1 - \alpha L^2) X_t = \varepsilon_t \text{ où } \varepsilon_t \text{ est un bruit blanc}$$

est donnée dans le slide (79). Le paramètre α est significativement non nul. Toutefois, si l'on considère l'autocorrélogramme des résidus obtenus après régression (slide (81)), on peut noter que l'on n'a pas un bruit blanc : toutes les autocorrelations paires sont non nulles. Ce modèle ne peut pas être retenu.

L'estimation du modèle (33) - modèle MA(1) pour Y_{4t} - défini par

$$(1 - L)(1 - L^4) X_t = \varepsilon_t - \theta \varepsilon_{t-1} \text{ où } \varepsilon_t \text{ est un bruit blanc}$$

est donnée dans le slide (96). Le paramètre θ est significativement non nul. Si l'on considère l'autocorrélogramme des résidus obtenus après régression (slide (98)), on peut noter qu'aucune autocorrélation n'est significativement non nulle. Si les tests de Box-Pierce (Q) ne sont pas validés à 5%, on peut malgré tout noter que ce modèle reste relativement bon Ce modèle pourrait être retenu.

(iii) La méthode *SCAN* appliquée à Y_{4t} (*c'est la variable que nous avons retenu dans la partie (i)*) suggère des modèles $ARMA(0,1)$ ou $ARMA(4,0)$, alors que la méthode *ESACF* suggère des modèles $ARMA(2,3)$ ou $ARMA(0,4)$.

Le modèle $ARMA(0,1)$ appliqué à Y_{4t} correspond au modèle (33) : ce modèle peut être retenu.

Le modèle $ARMA(4,0)$ appliqué à Y_{4t} correspond au modèle estimé slide (106) : les 4 paramètres sont significatifs, et si l'on considère les autocorrélations, on obtient un bruit blanc : ce modèle peut être retenu.

Le modèle $ARMA(2,3)$ appliqué à Y_{4t} correspond au modèle estimé slide (110) : on peut noter que le coefficient en $AR(1)$ est non-significatif. Le slide (111) permet de rester la même équation sans retard d'ordre 1 sur Y_{4t} : tous les coefficients sont là aussi significatifs. Et si l'on considère l'autocorrélogramme des résidus obtenus après régression (slide (11)), on peut noter qu'aucune autocorrelation n'est significativement non nulle : l'hypothèse de bruit blanc est validée également par le Q -test : ce modèle peut être retenu.

Le modèle $ARMA(0,4)$ appliqué à Y_{4t} correspond au modèle estimé slide (115) : les composantes d'ordre 2 et 3 sont ne sont pas significativement non nulles, et si l'on teste le modèle sans les composantes d'ordre 2 et 3, la composante d'ordre 1 n'est plus significative. Enfin, le modèle $MA(4)$ ne marchant pas non plus, ce modèle ne peut pas être retenu.

(iv) Le modèle $AR(4)$ appliquée à Y_{4t} (slide (106)) est valide : il est possible de modéliser la série X_t sans composante MA : le modèle

$$(1 - L)(1 - L^4)(1 - 0.6236L - 0.5452L^2 - 0.5238L^3 - 0.2815L^4)X_t = \varepsilon_t$$

est valide.

(v) Le modèle $(1 - L)(1 - L^4)(1 - 0.8L)X_t = (1 - 0.6L^2)\varepsilon_t$ c'est à dire Y_{4t} modélisé par une modéliser par une modèle $ARMA(1,2)$ est estimé dans le slide (88). Ce modèle est validé : les paramètres sont tous significatifs, et le résidu est un bruit blanc. Ce modèle est retenu.

Les différents modèles possibles sont alors

- [1] - slide (96) $(1 - L)(1 - L^4)X_t = (1 - \beta L)\eta_t$
- [2] - slide (111) $(1 - L)(1 - L^4)(1 - \alpha L^2)X_t = (1 - \beta_1 L - \beta_2 L^2 - \beta_3 L^3)\varepsilon_t$
- [3] - slide (106) $(1 - L)(1 - L^4)(1 - \alpha_1 L - \alpha_2 L^2 - \alpha_3 L^3 - \alpha_4 L^4)X_t = u_t$
- [4] - slide (88) $(1 - L)(1 - L^4)(1 - \alpha L)X_t = (1 - \beta L^2)v_t$
- [5] - slide (131) $(1 - L)(1 - L^4)(1 - \alpha_2 L^2 - \alpha_4 L^4)X_t = (1 - \beta_1 L - \beta_2 L^2 - \beta_5 L^5)e_t$
- [6] - slide (123) $(1 - L)(1 - L^4)(1 - \alpha L^2)X_t = (1 - \beta_1 L - \beta_3 L^3 - \beta_4 L^4)e_t$

où $\eta_t, \varepsilon_t, u_t$ et v_t sont des bruit blancs.

Les indicateurs de choix de modèles donnent les résultats suivants :

	R^2	\bar{R}^2	$\log L$	Akaike	Schwarz	F -stat
[1]	0.2893	0.2893	-101.6273	-0.10115	-0.07025	
[2]	0.3322	0.3032	-94.7231	-0.13313	-0.00763	11.44084
[3]	0.3487	0.3196	-91.8485	-0.13792	-0.01044	11.95962
[4]	0.3273	0.3179	-97.4373	-0.15038	-0.08811	35.02662
[5]	0.9920	0.9916	-92.8359	-0.15744	-0.00056	2117.092
[6]	0.9913	0.9910	-99.1440	-0.08737	0.036229	2709.168

Comme le montre le tableau ci-dessus, le modèle [4] a beau être celui qui a été simulé, ce n'est pas celui qui modélise la mieux la série. En particulier, [5] présente un meilleur critère d'Akaike, et [6] un meilleur critère de Schwarz (*que l'on cherche à minimiser*), ainsi qu'une meilleure F -stat.

4.3 Examen de 2002/2003

Exercice 9 *Le processus $MA(2)$ suivant est-il stationnaire (au second ordre) ?*

$$Y_t = [1 - 2.4L + 0.8L^2]\varepsilon_t \text{ où } \mathbb{E}(\varepsilon_s\varepsilon_t) = \begin{cases} 1 & \text{si } s = t \\ 0 & \text{sinon} \end{cases} \text{ et } \mathbb{E}(\varepsilon_t) = 0 \text{ pour tout } t, s$$

Si oui, calculer sa fonction d'autocovariance $\gamma(h) = \text{cov}(Y_t, Y_{t+h})$.

Un processus $MA(q)$ est toujours stationnaire (*dès lors que q est fini*). Dans le cadre général des processus $MA(q)$, 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

$$\gamma(0) = [1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2] \sigma^2$$

Dans le cas d'un processus $MA(2)$,

$$\gamma(h) = \begin{cases} [1 + \theta_1^2 + \theta_2^2] \sigma^2 & \text{pour } h = 0 \\ \theta_1 [1 + \theta_2] \sigma^2 & \text{pour } h = 1 \\ \theta_2 \sigma^2 & \text{pour } h = 2 \\ 0 & \text{pour } h \geq 3 \end{cases}$$

D'où la fonction d'autocovariance dans le cas particulier $Y_t = [1 - 2.4L + 0.8L^2] \varepsilon_t$ où $V(\varepsilon_t) = 1$,

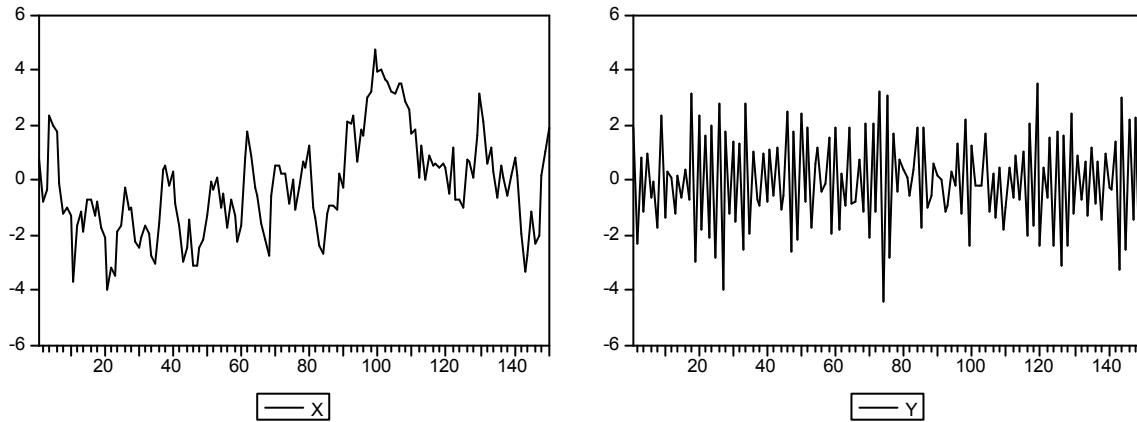
$$\gamma(0) = 1 + 2.4^2 + 0.8^2 = 7.4$$

$$\gamma(1) = 2.4 - .8 * 2.4 = 0.48$$

$$\gamma(2) = 0.8$$

$$\gamma(h) = \text{pour } h \geq 3$$

Exercice 10 Un processus (Z_t) suivant a été simulé suivant un processus $AR(1)$, sans constante, de la forme $Z_t = \rho Z_{t-1} + \varepsilon_t$ où (ε_t) est un bruit blanc gaussien. Il a été simulé une première fois avec $\rho = -0.7$ et une seconde fois $\rho = 0.85$. Sur les sorties ci-dessous, quelle série correspond au cas $\rho = -0.7$? Quelle série correspond au cas $\rho = 0.85$? Justifiez votre réponse.



Le processus de droite semble correspondre au cas où l'autocorrélation est négative : si Z_{t-1} est positif, Z_t semble négative (en moyenne), et réciproquement, ce qui correspond à un cas où $\text{corr}(Z_{t-1}, Z_t) \leq 0$. Le comportement semble le contraire dans le cas de gauche : $\text{corr}(Z_{t-1}, Z_t) \geq 0$. La série de gauche correspond au cas $\rho = 0.85$ et le cas de droite $\rho = -0.7$.

Exercice 11 (i) Dans le cas d'un processus stationnaire $AR(2)$, $X_t = \alpha X_{t-1} + \beta X_{t-2} + \varepsilon_t$ où (ε_t) est un bruit blanc, rappeler les valeurs des $\rho(1)$ et $\rho(2)$ en fonction de α et β .

(ii) Lors de la modélisation d'une série d'observations X_1, \dots, X_T , l'autocorrélogramme suivant a été obtenu,

Autocorr.	Partial Corr.	AC	PAC	Q-Stat	Prob
■■■■■	■■■■■	1 -0.60555	-0.60555	1834.56	0
■■■■■	■■■■■	2 0.68580	0.50387	4188.07	0
■■■■■	■■■■■	3 -0.51155	-0.00083	5497.80	0
■■■■■	■■■■■	4 0.498081	0.00121	6739.72	0
■■■■■	■■■■■	5 -0.40768	-0.00091	7571.90	0

On souhaite alors modéliser la série à l'aide d'un processus $AR(2)$. En utilisant uniquement cette sortie, proposez des estimateurs $\hat{\alpha}$ et $\hat{\beta}$ des coefficients autorégressifs.

(i) Soit (X_t) un processus stationnaire $AR(2)$ de la forme $X_t = \alpha X_{t-1} + \beta X_{t-2} + \varepsilon_t$ où (ε_t) est un bruit blanc, les deux premières équations de Yule-Walker sont

$$\begin{bmatrix} \rho(1) \\ \rho(2) \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix},$$

c'est à dire

$$\begin{cases} \rho(1) = \alpha + \beta\rho(1) \\ \rho(2) = \alpha\rho(1) + \beta \end{cases} \text{ soit } \begin{cases} \rho(1) = \alpha / (1 - \beta) \\ \rho(2) = \alpha^2 / (1 - \beta) + \beta \end{cases}$$

(ii) L'équation de Yule-Walker peut s'inverser, de la façon suivante

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho(1) \\ \rho(2) \end{bmatrix} = \frac{1}{1 - \rho(1)^2} \begin{bmatrix} 1 & -\rho(1) \\ -\rho(1) & 1 \end{bmatrix} \begin{bmatrix} \rho(1) \\ \rho(2) \end{bmatrix},$$

ce qui donne, respectivement

$$\alpha = \frac{\rho(1)[1 - \rho(2)]}{1 - \rho(1)^2} \text{ et } \beta = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2},$$

dont des estimateurs naturels sont

$$\hat{\alpha} = \frac{\hat{\rho}(1)[1 - \hat{\rho}(2)]}{1 - \hat{\rho}(1)^2} \text{ et } \hat{\beta} = \frac{\hat{\rho}(2) - \hat{\rho}(1)^2}{1 - \hat{\rho}(1)^2}.$$

AN : dans l'exemple considéré, $\hat{\rho}(1) \approx -0.60555$ et $\hat{\rho}(1) \approx 0.68580$, et donc

$$\hat{\alpha} \approx \frac{-0.60555 \cdot [1 - 0.68580]}{1 - 0.60555^2} \approx -0.3 \text{ et } \hat{\beta} \approx \frac{0.68580 - 0.60555^2}{1 - 0.60555^2} \approx 0.5.$$

Exercice 12 Les sorties informatiques suivantes ont été obtenues lors de modélisations de modèles $ARIMA(p, d, q)$, suivant la méthodologie proposée par Box et Jenkins, sur trois séries différentes. Le but est de répondre aux questions à l'aide uniquement des sorties proposées.

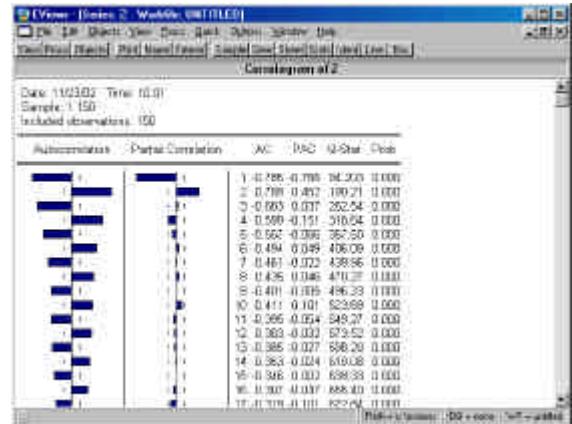
(1) Indiquez si le modèle $ARMA(3,1)$ suivant peut être retenu, sinon, en proposer un autre qu'il pourrait être intéressant de tester

```
Dependent Variable: Z
Method: Least Squares
Sample(adjusted): 4 5000
Included observations: 4997 after adjusting endpoints
Convergence achieved after 7 iterations

Variable Coefficient Std. Error t-Statistic Prob.
AR(1) -0.148403 1.648750 -0.090009 0.9283
AR(2) 0.549109 0.494613 1.110179 0.2670
AR(3) -0.077412 0.831452 -0.093105 0.9258
MA(1) -0.150609 1.649037 -0.091332 0.9272

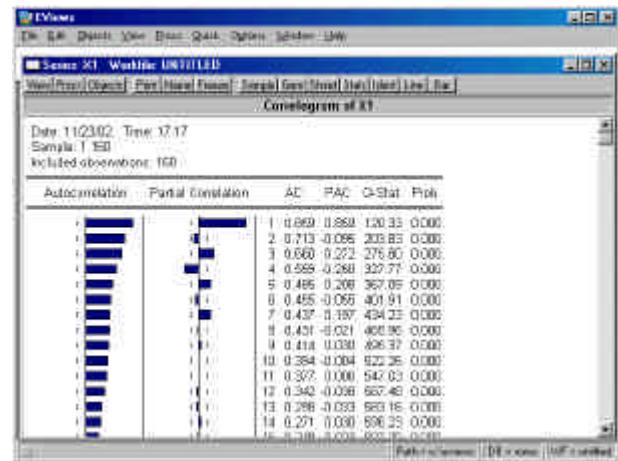
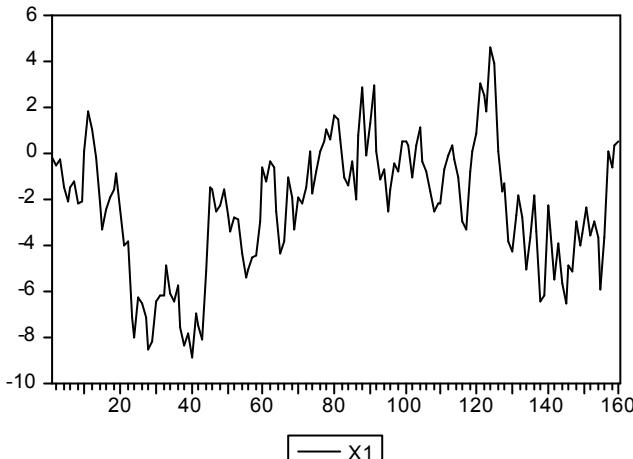
R-squared 0.527571 Mean dependent var -0.013720
Adjusted R-squared 0.527287 S.D. dependent var 1.435987
S.E. of regression 0.987301 Akaike info criterion 2.813116
Sum squared resid 4866.989 Schwarz criterion 2.818332
Log likelihood -7024.570 Durbin-Watson stat 2.000307

Inverted AR Roots .573 .152 -.886
Inverted MA Roots .152
```



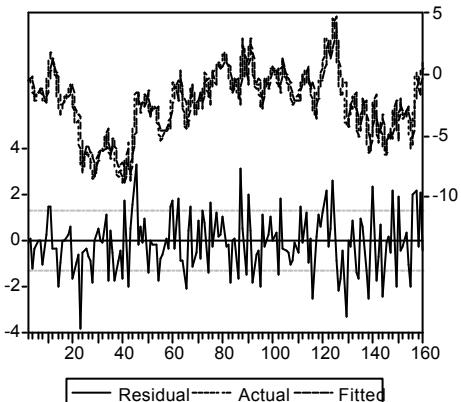
Lors de l'estimation des coefficients, aucun ne semble significatif, donc ce modèle ne peut pas être retenu. En fait, si l'on regarde les racines des polynômes AR et MA (ou l'inverse des racines comme cela est présenté si-dessus), on peut noter que la racine 0.152 apparaît pour les deux polynômes : les polynômes AR et MA ont une racine commune $((1 - \alpha L)(1 - a_1 L)(1 - a_2 L)X_t = (1 - \alpha L)\varepsilon_t)$. L'estimation des modèles $ARMA$ se faisant pour des polynômes minimaux, il est alors normal que la forme $ARMA$ testée ne convienne pas. Toutefois, il pourrait être intéressant de tester un modèle $AR(2)$.

Exercise 13 (2) Indiquez si le modèle ARMA(1,2) suivant peut être retenu,



Dependent Variable: X1
Method: Least Squares
Sample (adjusted): 2 160
Included observations: 159 after adjusting endpoints
Convergence achieved after 7 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.991881	0.025671	37.47017	0.0000
MA(1)	0.242948	0.078655	3.088764	0.0024
MA(2)	-0.498274	0.078661	-6.334439	0.0000
R-squared	0.791638	Mean dependent var		-2.355312
Adjusted R-squared	0.788966	S.D. dependent var		2.783367
S.E. of regression	1.278635	Akaike info criterion		3.348151
Sum squared resid	255.0455	Schwarz criterion		3.406055
Log likelihood	-263.1780	Durbin-Watson stat		2.055100
Inverted AR Roots	.99			
Inverted MA Roots	.59		-.84	



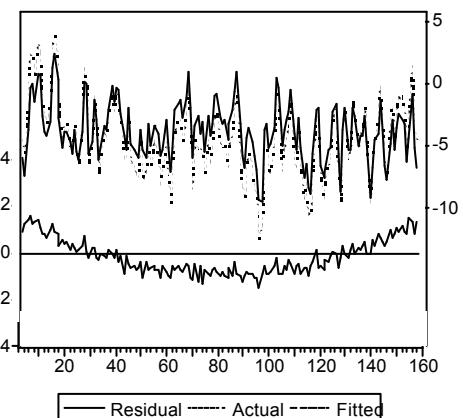
En particulier, il pourra être intéressant de commenter l'hypothèse de stationnarité de la série à l'aide de plusieurs arguments.

Les autocorrelations sont importantes et largement significatives (au moins pour les 15 premiers retards), ce qui, en pratique, laisse présager la présence d'une racine unité. Cette hypothèse est d'ailleurs confirmée lors que l'on regarde l'estimation d'un modèle ARMA(1,2), $(1 - \phi L) X_t = (1 - \alpha L)(1 - \beta L) \varepsilon_t$, on peut noter que $\hat{\alpha} = 0.99$: le pokynôme autorégressif semble avoir une racine unité.

Exercise 14 (3) Indiquez si le modèle ARMA(1,2) suivant peut être retenu, en utilisant la sortie suivante,

Dependent Variable: X2
Method: Least Squares
Sample (adjusted): 2 160
Included observations: 159 after adjusting endpoints
Convergence achieved after 7 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.731881	0.025671	5.547017	0.0003
MA(1)	0.372958	0.078655	13.08764	0.0000
MA(2)	-0.517354	0.078661	-4.334439	0.0061
R-squared	0.791638	Mean dependent var		-2.355312
Adjusted R-squared	0.788966	S.D. dependent var		2.783367
S.E. of regression	1.278635	Akaike info criterion		3.348151
Sum squared resid	255.0455	Schwarz criterion		3.406055
Log likelihood	-263.1780	Durbin-Watson stat		0.975515
Inverted AR Roots	.73			
Inverted MA Roots	.59		-.84	

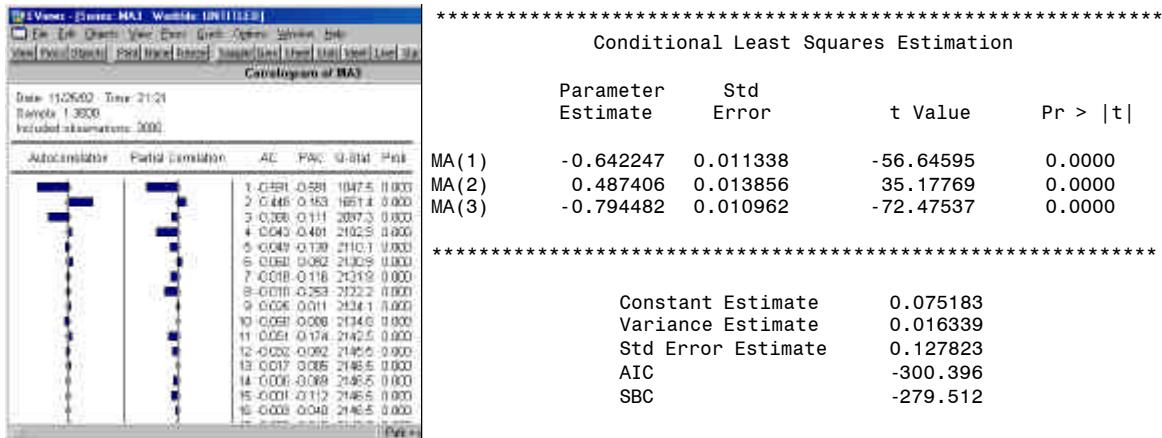


Une des conditions importantes lors d'une modélisation *ARMA* est que les résidus ε_t doivent être un bruit blanc. Dans le cas présenté ci-dessus, les coefficients de la modélisation *ARMA* sont certes significatifs, mais le bruit n'est pas un bruit blanc, comme le montre la représentation graphique ci-dessous. Le test de Durbin-Watson confirme d'ailleurs cette hypothèse, puisqu'il valide l'hypothèse alternative $\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t$ où $\rho \neq 1$.

Exercice 15 Un utilisateur a récupéré un programme de modélisation de processus *ARMA* sur internet, mais sans guide d'utilisation. En particulier, en essayant de modéliser la série par un processus *MA(3)*, il ne sait pas si le modèle testé est le modèle [1] ou le modèle [2],

$$\begin{cases} [1] & X_t = \varepsilon_t - \alpha_1\varepsilon_{t-1} - \alpha_2\varepsilon_{t-2} - \alpha_3\varepsilon_{t-3} \\ [2] & X_t = \eta_t + \beta_1\eta_{t-1} + \beta_2\eta_{t-2} + \beta_3\eta_{t-3} \end{cases} \text{ où } (\varepsilon_t) \text{ et } (\eta_t) \text{ sont des bruits blancs}$$

La sortie obtenue lors de l'estimation des paramètres est la suivante, ainsi que l'autocorrélogramme de (X_t) ,



Précisez, en justifiant votre réponse, quel modèle ([1] ou [2]) est testé à l'aide de ce programme. Autrement, le modèle est-il

$$[1] : X_t = \varepsilon_t + \frac{0.642}{(0.0113)} \varepsilon_{t-1} - \frac{0.487}{(0.0114)} \varepsilon_{t-2} + \frac{0.794}{(0.0109)} \varepsilon_{t-3} \text{ où } (\varepsilon_t) \text{ est un bruit blanc}$$

ou bien

$$[2] : X_t = \varepsilon_t - \frac{0.642}{(0.0113)} \varepsilon_{t-1} + \frac{0.487}{(0.0114)} \varepsilon_{t-2} - \frac{0.794}{(0.0109)} \varepsilon_{t-3} \text{ où } (\varepsilon_t) \text{ est un bruit blanc}$$

Dans le cas d'un processus *MA* (q), écrit sous la forme [1], l'autocovariance à l'ordre q est donnée par

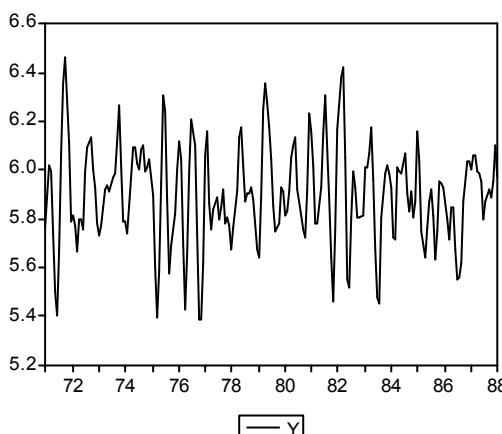
$$\begin{aligned} \gamma(q) &= \text{cov}(X_t, X_{t-q}) = \mathbb{E}(X_t X_{t-q}) \\ &= \mathbb{E}([\varepsilon_t - \theta_1\varepsilon_{t-1} - \dots - \theta_q\varepsilon_{t-q}] [\varepsilon_{t-q} - \theta_1\varepsilon_{t-q-1} - \dots - \theta_q\varepsilon_{t-q-q}]) \\ &= \mathbb{E}(-\theta_q\varepsilon_{t-q}\varepsilon_{t-q}) = -\theta_q\sigma^2 \text{ car } \mathbb{E}(\varepsilon_{t-i}\varepsilon_{t-j}) = \sigma^2 \mathbb{I}(i=j) \end{aligned}$$

L'autocorrélation étant du même signe que l'autocovariance, alors $\rho(q)$ est toujours du signe opposé à θ_q dans un modèle écrit sous forme [1]. Dans l'exemple, la troisième autocorrélation est négative ($\rho(3) \approx -0.385$) et le 3ème coefficient de la régression, correspondant au retard d'ordre 3 de la moyenne mobile est -0.794482 . Autrement dit, le modèle testé ne peut pas être le modèle [1] (*les signes auraient dû être opposés*), c'est donc un modèle de la forme [2] qui a été testé. La modélisation de la série temporelle est alors

$$[2] : X_t = \varepsilon_t - \frac{0.642}{(0.0113)} \varepsilon_{t-1} + \frac{0.487}{(0.0114)} \varepsilon_{t-2} - \frac{0.794}{(0.0109)} \varepsilon_{t-3} \text{ où } (\varepsilon_t) \text{ est un bruit blanc}$$

Exercice 16 Des tests de stationnarité ont été fait sur la série (X_t) présentée ci-dessous, à l'aide de la méthode de Dickey et Fuller. En utilisant uniquement les deux sorties présentées ci-dessous (deux tests de racine unité ont été

faits), expliquez si l'hypothèse de stationnarité est acceptée, rejetée, ou si l'on ne peut pas conclure.



ADF Test Statistic	-0.410240	1% Critical Value*	-2.5757
		5% Critical Value	-1.9413
		10% Critical Value	-1.6165

*MacKinnon critical values for rejection of hypothesis of a unit root

Augmented Dickey-Fuller Test Equation - NONE

Dependent Variable: D(Y)

Method: Least Squares

Sample(adjusted): 1971:03 1988:01

Included observations: 203 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Y(-1)	-0.000765	0.001864	-0.410240	0.6821
D(Y(-1))	0.421666	0.063685	6.621136	0.0000
R-squared	0.179261	Mean dependent var		-0.000231
Adjusted R-squared	0.175178	S.D. dependent var		0.172520
S.E. of regression	0.156682	Akaike info criterion		-0.859391
Sum squared resid	4.934414	Schwarz criterion		-0.826748
Log likelihood	89.22815	Durbin-Watson stat		1.474710

ADF Test Statistic	-15.03881	1% Critical Value*	-3.4639
		5% Critical Value	-2.8758
		10% Critical Value	-2.5743

*MacKinnon critical values for rejection of hypothesis of a unit root

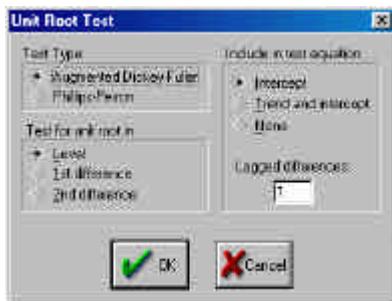
Augmented Dickey-Fuller Test Equation - INTERCEPT

Dependent Variable: D(Y)

Method: Least Squares

Sample(adjusted): 1971:03 1988:01

Included observations: 203 after adjusting endpoints



Variable	Coefficient	Std. Error	t-Statistic	Prob.
Y(-1)	-0.610786	0.040614	-15.03881	0.0000
D(Y(-1))	0.727746	0.048263	15.07888	0.0000
C	3.600772	0.239613	15.02746	0.0000
R-squared	0.614518	Mean dependent var		-0.000231
Adjusted R-squared	0.610663	S.D. dependent var		0.172520
S.E. of regression	0.107647	Akaike info criterion		-1.605248
Sum squared resid	2.317581	Schwarz criterion		-1.556285
Log likelihood	165.9327	F-statistic		159.4153
Durbin-Watson stat	1.765054	Prob(F-statistic)		0.000000

Le test de Dickey-Fuller simple (*test d'un modèle AR (1) sous contrainte*) vise à tester si le coefficient devant la composante AR(1) vaut 1 (*présence de "racine unité"*). Trois tests sont alors possibles : sans constante, avec constante, avec trend et constante, c'est à dire que l'on teste des modèles de la forme suivante, [1] $X_t = aX_{t-1} + \varepsilon_t$ (*sans constante, testé en haut*) ou [2] $X_t = \mu + aX_{t-1} + \varepsilon_t$ (*avec constante, en bas*), par exemple.

Le processus (ε_t) est supposé être un bruit blanc, lors d'une modélisation ARMA, donc d'espérance nulle. Et si l'on cherche l'espérance de (X_t), on trouve (*si le processus (X_t) est effectivement stationnaire*) :

$$[1] \mathbb{E}(X) = a\mathbb{E}(X) + 0, \text{ donc si } a \text{ est différent de 1, } \mathbb{E}(X) = 0,$$

$$[2] \mathbb{E}(X) = \mu + a\mathbb{E}(X) + 0, \text{ donc si } a \text{ est différent de 1, } \mathbb{E}(X) = \mu/(1-a).$$

Le graphique ci-dessus montre que (X_t) "oscille", en moyenne, autour de 6 : on est donc amené à tester un modèle avec constante. Et comme le montre la sortie du bas, le test de Dickey-Fuller pousse à rejeter l'hypothèse $a = 1$. Le modèle sans constante pousse à accepter l'hypothèse $H_0 : a \neq 1$, mais le modèle testé semble manifestement faux.

Au vu du test de Dickey-Fuller, on rejette l'hypothèse de présence de racine unité.

4.4 Examen de 2003/2004

Exercice 17 On considère (Y_t) un processus vérifiant

$$Y_t = X_t + \varepsilon_t, \quad (34)$$

où (ε_t) est un bruit blanc, et où (X_t) vérifie

$$X_t = \phi X_{t-1} + \eta_t + \theta \eta_{t-1}, \quad (35)$$

où (η_t) est un bruit blanc tel que $\mathbb{E}(\varepsilon_t \eta_s) = 0$ pour tout s, t . On suppose de plus que $|\phi| < 1$, $|\theta| < 1$ et $\theta \neq -\phi$. Montrer que (Y_t) est un processus stationnaire et donner sa fonction d'autocorrélation.

Le processus (Y_t) peut s'écrire

$$\begin{aligned} Y_t &= X_t + \varepsilon_t = [\phi X_{t-1} + \eta_t + \theta \eta_{t-1}] + \varepsilon_t \\ &= \phi [Y_{t-1} - \varepsilon_{t-1}] + \eta_t + \theta \eta_{t-1} + \varepsilon_t \text{ car } Y_{t-1} = X_{t-1} + \varepsilon_{t-1}, \\ &= \phi Y_{t-1} + \underbrace{\varepsilon_t - \phi \varepsilon_{t-1}}_{\text{et }} + \underbrace{\eta_t + \theta \eta_{t-1}}_{\text{et }} = \phi Y_{t-1} + u_t. \end{aligned}$$

Le processus (u_t) étant un somme de bruits blancs, il doit être possible de montrer que (u_t) est un processus $MA(q)$. (on peut pas utiliser le fait que u_t est la somme de deux processus stationnaires et que donc, (u_t) est stationnaire, ce résultat est en général faux (cf examen de 2001/2002)). Montrons plus particulièrement que (u_t) est un processus $MA(q)$, en calculant ses autocorrélations, en notant que $\mathbb{E}(u_t) = 0$ (car $\mathbb{E}(\varepsilon_t) = \mathbb{E}(\eta_t) = 0$),

$$\begin{aligned} \mathbb{E}(u_t^2) &= \mathbb{E}([\varepsilon_t - \phi \varepsilon_{t-1} + \eta_t + \theta \eta_{t-1}]^2) \\ &= \mathbb{E}([\varepsilon_t - \phi \varepsilon_{t-1} + \eta_t + \theta \eta_{t-1}] [\varepsilon_t - \phi \varepsilon_{t-1} + \eta_t + \theta \eta_{t-1}]) \\ &= \mathbb{E}(\varepsilon_t^2 + \phi^2 \varepsilon_{t-1}^2 + \eta_t^2 + \theta^2 \eta_{t-1}^2 - 2\phi \varepsilon_t \varepsilon_{t-1} + 2\varepsilon_t \eta_t + 2\theta \varepsilon_t \eta_{t-1} - 2\phi \varepsilon_{t-1} \eta_t - 2\phi \theta \varepsilon_{t-1} \eta_{t-1} + 2\theta \eta_t \eta_{t-1}). \end{aligned}$$

En utilisant la linéarité de l'espérance, et en utilisant le fait que $\mathbb{E}(\varepsilon_t \varepsilon_{t-1}) = 0$ ((ε_t) est un bruit blanc) $\mathbb{E}(\eta_t \eta_{t-1}) = 0$, et

$$\mathbb{E}(\varepsilon_t \eta_t) = \mathbb{E}(\eta_{t-1} \varepsilon_{t-1}) = \mathbb{E}(\varepsilon_t \eta_{t-1}) = \mathbb{E}(\varepsilon_{t-1} \eta_t) = 0,$$

(par hypothèse $\mathbb{E}(\varepsilon_t \eta_s) = 0$ pour tout s, t), cette expression se simplifie en

$$\mathbb{E}(u_t^2) = \mathbb{E}(\varepsilon_t^2) + \phi^2 \mathbb{E}(\varepsilon_{t-1}^2) + \mathbb{E}(\eta_t^2) + \theta^2 \mathbb{E}(\eta_{t-1}^2) = (1 + \phi^2) \sigma_\varepsilon^2 + (1 + \theta^2) \sigma_\eta^2,$$

en notant σ_ε^2 et σ_η^2 , respectivement, les variances des deux bruits blancs (et en détaillant outrageusement les calculs...). De même

$$\begin{aligned} \mathbb{E}(u_t u_{t-1}) &= \mathbb{E}([\varepsilon_t - \phi \varepsilon_{t-1} + \eta_t + \theta \eta_{t-1}] [\varepsilon_{t-1} - \phi \varepsilon_{t-2} + \eta_{t-1} + \theta \eta_{t-2}]) \\ &= \mathbb{E}(\varepsilon_t \varepsilon_{t-1} - \phi \varepsilon_t \varepsilon_{t-2} + \varepsilon_t \eta_{t-1} + \theta \varepsilon_t \eta_{t-2} - \phi \varepsilon_{t-1}^2 + \phi^2 \varepsilon_{t-1} \varepsilon_{t-2} - \phi \varepsilon_{t-1} \eta_{t-1} - \theta \phi \varepsilon_{t-1} \eta_{t-2} \\ &\quad + \varepsilon_{t-1} \eta_t - \phi \varepsilon_{t-2} \eta_t + \eta_t \eta_{t-1} + \theta \eta_t \eta_{t-2} + \theta \varepsilon_{t-1} \eta_{t-1} - \phi \theta \varepsilon_{t-2} \eta_{t-1} + \theta \eta_{t-1}^2 + \theta^2 \eta_{t-1} \eta_{t-2}). \end{aligned}$$

En utilisant encore une fois la linéarité de l'espérance, et le fait que $\mathbb{E}(\varepsilon_s \varepsilon_t) = \mathbb{E}(\eta_s \eta_t) = 0$ pour $s \neq t$ (hypothèse de bruit blanc), et que, par hypothèse, $\mathbb{E}(\varepsilon_t \eta_s) = 0$ pour tout s, t (même égaux), on note qu'il ne reste plus que deux termes non nuls,

$$\mathbb{E}(u_t u_{t-1}) = -\phi \mathbb{E}(\varepsilon_{t-1}^2) + \theta \mathbb{E}(\eta_{t-1}^2) = \theta \sigma_\eta^2 - \phi \sigma_\varepsilon^2.$$

Et de façon plus général, un calcul identique à celui détaillé ci-dessus donne

$$\mathbb{E}(u_t u_{t-h}) = 0 \text{ pour } h \geq 2.$$

On en déduit, tout d'abord que, pour tout h , et pour tout t , $\mathbb{E}(u_t u_{t-h})$ ne dépend pas de t : **le processus (u_t) est effectivement un processus stationnaire**. De plus, compte tenu de la forme particulière des autocorrélations, on peut noter que $\rho_u(h) = \mathbb{E}(u_t u_{t-h}) / \mathbb{E}(u_t^2) = 0$ pour $h \geq 2$: **le processus (u_t) est un processus $MA(q)$, où $q = 1$.**

Remarque : En fait, on sait juste que $q = 1$...il y avait une petite erreur dans l'énoncé : il en mentionné que " $\theta \neq -\phi$ ", ce qui n'a ici aucun intérêt...la condition qui aurait pu être utile était plutôt $\theta \sigma_\eta^2 - \phi \sigma_\varepsilon^2 \neq 0$, qui s'écrirait $\theta \neq \phi$ dans le cas où les bruits ont même variance : dans ce cas, on peut en déduire que (u_t) est un processus $MA(1)$ - sinon (u_t) est un bruit blanc. (tout ce qui était mentionné dans cette remarque n'était, bien entendu, pas demandé).

Aussi, (Y_t) est un processus $ARMA(1,1)$. Notons qu'il s'agit d'un vrai $ARMA(1,1)$ puisque les racines des polynômes autorégressif et moyenne-mobile sont différentes. En effet, il existe un bruit blanc v_t tel que

$$u_t = v_t + \alpha v_{t-1} \text{ où } \alpha \text{ est solution de } \frac{\alpha}{1 + \alpha^2} = \frac{\theta \sigma_\eta^2 - \phi \sigma_\varepsilon^2}{(1 + \phi^2) \sigma_\varepsilon^2 + (1 + \theta^2) \sigma_\eta^2}$$

(ce coefficient étant obtenu en notant que pour un processus $MA(1)$, de la forme $u_t = v_t + \alpha v_{t-1}$, la première autocorrélation $\rho(1)$ est égale à $\alpha / (1 + \alpha^2)$). On obtient ainsi une équation de degré 2, admettant deux racines (l'une étant l'inverse de l'autre - cf exercice 2).

Aussi, (Y_t) suit un processus $ARMA(1,1)$, qui peut d'ailleurs s'écrire

$$Y_t = \phi Y_{t-1} + v_t + \alpha v_{t-1}, \text{ où } (v_t) \text{ est un bruit blanc.}$$

En utilisant la relation donnée dans les notes de cours (page 106), on en déduit la forme des autocorrélations, à savoir

$$\rho(1) = \frac{(1+\phi\alpha)(\phi+\alpha)}{1+\alpha^2+2\phi\alpha}$$

et

$$\rho(h) = \phi^h \rho(1) \text{ pour } h \geq 2.$$

Exercice 18 Monter que les deux processus $MA(1)$ suivant ont la même fonction d'autocorrélation,

$$X_t = \varepsilon_t + \theta \varepsilon_{t-1} \text{ et } Y_t = \eta_t + \frac{1}{\theta} \eta_{t-1}, \quad (36)$$

où (ε_t) et (η_t) sont des bruits blancs au sens L^2 , et où $0 < |\theta| < 1$.

Le cours étant autorisé pendant l'examen (il est aussi possible de retrouver la forme des autocorrélations en moins de 2 pages), on obtient facilement que les autocorrélations pour le processus (X_t) est données par (X_t étant centré, i.e. $\mathbb{E}X_t = 0$)

$$\rho_X(1) = \frac{\mathbb{E}(X_t X_{t-1})}{\mathbb{E}(X_t^2)} = \frac{\theta \mathbb{E}(\varepsilon_t^2)}{(1+\theta^2)\mathbb{E}(\varepsilon_t^2)} = \frac{\theta}{1+\theta^2},$$

et

$$\rho_X(h) = \frac{\mathbb{E}(X_t X_{t-h})}{\mathbb{E}(X_t^2)} = \frac{0}{(1+\theta^2)\mathbb{E}(\varepsilon_t^2)} = 0, \text{ pour } h \geq 2.$$

De façon strictement analogue ((Y_t) étant également un processus $MA(1)$ que l'on notera $Y_t = \eta_t + \kappa \eta_{t-1}$, où $\kappa = 1/\theta$

$$\rho_Y(1) = \frac{\mathbb{E}(Y_t Y_{t-1})}{\mathbb{E}(Y_t^2)} = \frac{\kappa \mathbb{E}(\varepsilon_t^2)}{(1+\kappa^2)\mathbb{E}(\varepsilon_t^2)} = \frac{\kappa}{1+\kappa^2} = \frac{1/\theta}{1+1/\theta^2} = \frac{\theta}{\theta^2+1},$$

(en multipliant en haut et en bas par θ^2) et

$$\rho_Y(h) = \frac{\mathbb{E}(Y_t Y_{t-h})}{\mathbb{E}(Y_t^2)} = \frac{0}{(1+\kappa^2)\mathbb{E}(\eta_t^2)} = 0, \text{ pour } h \geq 2.$$

C'est à dire que pour tout h , $\rho_Y(h) = \rho_X(h)$: les deux processus ont la même fonction d'autocorrélation.

Exercice 19 1) Supposons que (Y_t) soit un processus $AR(1)$ stationnaire, et que (X_t) soit un processus $MA(1)$, soit

$$Y_t = \phi Y_{t-1} + \varepsilon_t \text{ et } X_t = \eta_t - \theta \eta_{t-1}, \quad (37)$$

où (ε_t) et (η_t) sont des bruits blancs. Si pour tout t , $Z_t = X_t + Y_t$, montrez que (Z_t) suit un processus $ARMA(1,q)$ où l'on précisera q ?

2). Soit (X_t) un processus que l'on souhaite modéliser, supposé suivre un processus $ARIMA(p,d,q)$, s'écrivant $(1-L)^d \Phi(L) X_t = \Theta(L) \varepsilon_t$, où (ε_t) est un bruit blanc. Malheureusement, ce processus n'est pas observable directement, et l'on ne peut observer que $X_t^* = X_t + Y_t$ où (Y_t) est une erreur d'observation (erreur de mesure ou bruit exogène) supposé suivre un processus $ARMA(\alpha,\beta)$, dont le bruit associé (η_t) est supposé indépendant de (ε_t) .

Montrer que (X_t^*) suit un processus $ARIMA(p^*,d,q^*)$ où l'on précisera p^* et q^* .

1. Le processus (Z_t) est la somme d'un processus $MA(1)$ et d'un processus $AR(1)$,

$$Z_t = X_t + Y_t = [\eta_t - \theta \eta_{t-1}] + [\phi Y_{t-1} + \varepsilon_t].$$

Comme $Z_{t-1} = X_{t-1} + Y_{t-1}$, et que le terme Y_{t-1} est déjà apparu, faisons apparaître X_{t-1} (ou plus précisément ϕX_{t-1}),

$$Z_t = \eta_t - \theta \eta_{t-1} + \phi Y_{t-1} + \varepsilon_t + \phi X_{t-1} - \phi X_{t-1} = \phi Z_{t-1} + [-\theta \eta_{t-1} + \varepsilon_t - \phi X_{t-1}]$$

or $X_{t-1} = \eta_{t-1} - \theta\eta_{t-2}$, donc, par substitution,

$$Z_t = \phi Z_{t-1} + [-\theta\eta_{t-1} + \varepsilon_t - \phi(\eta_{t-1} - \theta\eta_{t-2})] = \phi Z_{t-1} + u_t,$$

où $u_t = \varepsilon_t - (\theta + \phi)\eta_{t-1} + \phi\theta\eta_{t-2}$. **On retrouve ici une question proche de l'exercice 1**, traitée auparavant : de la même façon, comme il est demandé de montrer que Z_t est un processus $ARMA(1, 1)$, montrons que (u_t) est un processus $MA(1)$. En notant que (u_t) est centré, on en déduit sa fonction d'autocovariance,

$$\begin{aligned}\mathbb{E}(u_t^2) &= \mathbb{E}([\varepsilon_t - \phi\varepsilon_{t-1} + \eta_t + \theta\eta_{t-1}]^2) \\ &= \mathbb{E}([\varepsilon_t - \phi\varepsilon_{t-1} + \eta_t + \theta\eta_{t-1}][\varepsilon_t - \phi\varepsilon_{t-1} + \eta_t + \theta\eta_{t-1}]) \\ &= \mathbb{E}(\varepsilon_t^2 + \phi^2\varepsilon_{t-1}^2 + \eta_t^2 + \theta^2\eta_{t-1}^2 - 2\phi\varepsilon_t\varepsilon_{t-1} + 2\varepsilon_t\eta_t + 2\theta\varepsilon_t\eta_{t-1} - 2\phi\varepsilon_{t-1}\eta_t - 2\phi\theta\varepsilon_{t-1}\eta_{t-1} + 2\theta\eta_t\eta_{t-1}).\end{aligned}$$

En utilisant la linéarité de l'espérance, et en utilisant le fait que $\mathbb{E}(\varepsilon_t\varepsilon_{t-1}) = 0$ ((ε_t) est un bruit blanc) $\mathbb{E}(\eta_t\eta_{t-1}) = 0$, et

$$\mathbb{E}(\varepsilon_t\eta_t) = \mathbb{E}(\eta_{t-1}\varepsilon_{t-1}) = \mathbb{E}(\varepsilon_t\eta_{t-1}) = \mathbb{E}(\varepsilon_{t-1}\eta_t) = 0,$$

(par hypothèse $\mathbb{E}(\varepsilon_t\eta_s) = 0$ pour tout s, t), cette expression se simplifie en

$$\mathbb{E}(u_t^2) = \mathbb{E}(\varepsilon_t^2) + \phi^2\mathbb{E}(\varepsilon_{t-1}^2) + \mathbb{E}(\eta_t^2) + \theta^2\mathbb{E}(\eta_{t-1}^2) = (1 + \phi^2)\sigma_\varepsilon^2 + (1 + \theta^2)\sigma_\eta^2,$$

en notant σ_ε^2 et σ_η^2 , respectivement, les variances des deux bruits blancs (et en détaillant outrageusement les calculs...). De même

$$\begin{aligned}\mathbb{E}(u_t u_{t-1}) &= \mathbb{E}([\varepsilon_t - \phi\varepsilon_{t-1} + \eta_t + \theta\eta_{t-1}][\varepsilon_{t-1} - \phi\varepsilon_{t-2} + \eta_{t-1} + \theta\eta_{t-2}]) \\ &= \mathbb{E}(\varepsilon_t\varepsilon_{t-1} - \phi\varepsilon_t\varepsilon_{t-2} + \varepsilon_t\eta_{t-1} + \theta\varepsilon_t\eta_{t-2} - \phi\varepsilon_{t-1}^2 + \phi^2\varepsilon_{t-1}\varepsilon_{t-2} - \phi\varepsilon_{t-1}\eta_{t-1} - \theta\phi\varepsilon_{t-1}\eta_{t-2} \\ &\quad + \varepsilon_{t-1}\eta_t - \phi\varepsilon_{t-2}\eta_t + \eta_t\eta_{t-1} + \theta\eta_t\eta_{t-2} + \theta\varepsilon_{t-1}\eta_{t-1} - \phi\theta\varepsilon_{t-2}\eta_{t-1} + \theta\eta_{t-1}^2 + \theta^2\eta_{t-1}\eta_{t-2}).\end{aligned}$$

En utilisant encore une fois la linéarité de l'espérance, et le fait que $\mathbb{E}(\varepsilon_s\varepsilon_t) = \mathbb{E}(\eta_s\eta_t) = 0$ pour $s \neq t$ (hypothèse de bruit blanc), et que, par hypothèse, $\mathbb{E}(\varepsilon_t\eta_s) = 0$ pour tout s, t (même égaux), on note qu'il ne reste plus que deux termes non nuls,

$$\mathbb{E}(u_t u_{t-1}) = -\phi\mathbb{E}(\varepsilon_{t-1}^2) + \theta\mathbb{E}(\eta_{t-1}^2) = \theta\sigma_\eta^2 - \phi\sigma_\varepsilon^2.$$

Et de façon plus général, un calcul identique à celui détaillé ci-dessus donne

$$\mathbb{E}(u_t u_{t-h}) = 0 \text{ pour } h \geq 2.$$

On en déduit, tout d'abord que, pour tout h , et pour tout t , $\mathbb{E}(u_t u_{t-h})$ ne dépend pas de t : **le processus (u_t) est effectivement un processus stationnaire**. De plus, compte tenu de la forme particulière des autocorrélations, on peut noter que $\rho_u(h) = \mathbb{E}(u_t u_{t-h})/\mathbb{E}(u_t^2) = 0$ pour $h \geq 2$: le processus (u_t) est un processus $MA(1)$.

Encore une fois, le fait que $q = 1$ (et que (u_t) n'est alors pas un bruit blanc) est assuré en rajoutant l'hypothèse $\theta\sigma_\eta^2 \neq \phi\sigma_\varepsilon^2$. On peut alors conclure que (Y_t) **est un processus $ARMA(1, 1)$** .

2. De façon analogue, supposons que (X_t) suive un processus $ARIMA(p, d, q)$ tel que $(1 - L)^d \Phi(L) X_t = \Theta(L) \varepsilon_t$ et que le bruit d'observation Y_t suive un processus $ARMA(\alpha, \beta)$, tel que $A(L) Y_t = B(L) \eta_t$. On s'intéresse alors à la somme de ces deux processus, $X_t^* = X_t + Y_t$.

Soit $\Phi^*(L) = \Phi(L) A(L)$ de degrés respectifs $p^* = p + \alpha$, et $q^* = q + \beta$. Montrons qu'alors $(1 - L)^d \Phi^*(L) X_t^*$ suit un processus $MA(q^*)$, i.e. (X_t^*) est un processus $ARIMA(p^*, d, q^*)$, ou $ARIMA(p + \alpha, d, q^*)$.

$$\begin{aligned}(1 - L)^d \Phi^*(L) X_t^* &= (1 - L)^d \Phi(L) A(L) [X_t + Y_t] \\ &= A(L) [(1 - L)^d \Phi(L) X_t] + (1 - L)^d \Phi(L) [A(L) Y_t] \\ &= A(L) \Theta(L) \varepsilon_t + (1 - L)^d \Phi(L) B(L) \eta_t \\ &= u_t\end{aligned}$$

On notera $\Theta^{*1}(L) = A(L)\Theta(L)$ et $\Theta^{*2}(L) = (1 - L)^d \Phi(L) B(L)$, de degrés respectifs $q^{*1} = a + q$ et $q^{*2} = d + p + \beta$, et

$$\Theta^{*i}(L) = \sum_{j=0}^{q^{*i}} \theta_j^{*i} L^j \text{ pour } i = 1, 2.$$

Pour montrer que u_t est un processus $MA(q^*)$ - pour un certain q^* - calculons ses autocorrélations, en notant tout d'abord que $\mathbb{E}(u_t) = 0$, et montrons qu'elles s'annulent à partir d'un certain rang.

$$\begin{aligned}\mathbb{E}(u_t^2) &= \mathbb{E}\left(\left[A(L)\Theta(L)\varepsilon_t + (1-L)^d\Phi(L)B(L)\eta_t\right]\left[A(L)\Theta(L)\varepsilon_t + (1-L)^d\Phi(L)B(L)\eta_t\right]\right) \\ &= \sum_{j=0}^{q^{*1}} (\theta_j^{*1})^2 \mathbb{E}(\varepsilon_{t-j}^2) + \sum_{j=0}^{q^{*2}} (\theta_j^{*2})^2 \mathbb{E}(\eta_{t-j}^2)\end{aligned}$$

tous les autres termes étant nuls car ils sont soit de la forme $\mathbb{E}(\varepsilon_{t-j}\varepsilon_{t-k})$ ou $\mathbb{E}(\eta_{t-j}\eta_{t-k})$ avec $j \neq k$, soit $\mathbb{E}(\varepsilon_{t-j}\eta_{t-k})$. Dans le premier cas, les termes sont nuls car (ε_t) et (η_t) sont des bruits blancs (et donc non-autocorrélés), et dans le second cas, les termes sont nuls car les deux bruits sont supposés indépendants (on utilise encore une fois les mêmes résultats que dans la question précédente, ainsi que dans l'exercice 1).

$$\begin{aligned}\mathbb{E}(u_t u_{t-1}) &= \mathbb{E}\left(\left[A(L)\Theta(L)\varepsilon_t + (1-L)^d\Phi(L)B(L)\eta_t\right]\left[A(L)\Theta(L)\varepsilon_{t-1} + (1-L)^d\Phi(L)B(L)\eta_{t-1}\right]\right) \\ &= \sum_{j=1}^{q^{*1}} \theta_j^{*1} \theta_{j-1}^{*1} \mathbb{E}(\varepsilon_{t-j}^2) + \sum_{j=1}^{q^{*2}} \theta_j^{*1} \theta_{j-1}^{*1} \mathbb{E}(\eta_{t-j}^2)\end{aligned}$$

en utilisant encore et toujours les mêmes arguments : bruits blancs indépendants entre eux. De façon plus générale, on en déduit

$$\begin{aligned}\mathbb{E}(u_t u_{t-k}) &= \mathbb{E}\left(\left[A(L)\Theta(L)\varepsilon_t + (1-L)^d\Phi(L)B(L)\eta_t\right]\left[A(L)\Theta(L)\varepsilon_{t-k} + (1-L)^d\Phi(L)B(L)\eta_{t-k}\right]\right) \\ &= \sum_{j=k}^{q^{*1}} \theta_j^{*1} \theta_{j-k}^{*1} \mathbb{E}(\varepsilon_{t-j}^2) + \sum_{j=k}^{q^{*2}} \theta_j^{*1} \theta_{j-k}^{*1} \mathbb{E}(\eta_{t-j}^2)\end{aligned}$$

à condition que ces sommes soient bien définies, c'est à dire $q^{*1} \geq k$ et $q^{*2} \geq k$, sinon, ces sommes valent 0. En particulier,

$$\begin{aligned}\mathbb{E}(u_t u_{t-q^{*1}-1}) &= \mathbb{E}\left(\left[A(L)\Theta(L)\varepsilon_t + (1-L)^d\Phi(L)B(L)\eta_t\right]\left[A(L)\Theta(L)\varepsilon_{t-q^{*1}-1} + (1-L)^d\Phi(L)B(L)\eta_{t-q^{*1}-1}\right]\right) \\ &= 0 + \sum_{j=k}^{q^{*2}} \theta_j^{*1} \theta_{j-k}^{*1} \mathbb{E}(\eta_{t-j}^2)\end{aligned}$$

et de même pour $\gamma(q^{*2} + 1)$: aussi, on en déduit facilement que $\gamma(\max\{q^{*1}, q^{*2}\} + 1) = 0$, et de façon plus générale $\gamma(\max\{q^{*1}, q^{*2}\} + h) = 0$ pour $h \geq 1$. Aussi, (u_t) est un processus $MA(q^*)$ où $q^* = \max\{a + q, d + p + \beta\}$.

Notons que des racines communes peuvent apparaître, aussi de façon plus rigoureuse, on peut juste conclure que (X_t^*) est un processus $ARIMA(p^*, d, q^*)$, où $p^* \leq p + \alpha$ et $q^* \leq \max\{a + q, d + p + \beta\}$.

Exercice 20 Supposons que (X_t) et (Y_t) soient deux processus tels que

$$\begin{cases} X_t = \alpha X_{t-1} + \beta Y_{t-1} + \varepsilon_t \\ Y_t = \delta Y_{t-1} + \gamma X_{t-1} + \eta_t \end{cases} \quad (38)$$

où (ε_t) et (η_t) sont des bruits blancs mutuellement indépendants (i.e. $E\varepsilon_t \eta_{t+h} = 0$ pour tout $h \in \mathbb{Z}$), montrer que (X_t) et (Y_t) suivent tout deux des processus $ARMA(2, 1)$ dont on donnera la forme (en indiquant les hypothèses éventuelles).

(indice : il sera intéressant d'écrire (38) sous forme matricielle $M(L)Z_t = \varepsilon_t$ où M est une matrice de polynômes d'opérateurs retards, puis de noter qu'en multipliant une matrice carrée par une autre matrice carrée (bien choisie) il est possible d'obtenir une matrice diagonale).

En notant $Z_t = (X_t, Y_t)'$, notons que Z_t vérifie

$$Z_t = KZ_{t-1} + u_t \text{ où } K = \begin{pmatrix} \alpha & \beta \\ \delta & \gamma \end{pmatrix} \text{ et } u_t = (\varepsilon_t, \eta_t)',$$

soit

$$M(L)Z_t = u_t \text{ où } M(L) = \begin{pmatrix} 1 - \alpha L & -\beta L \\ -\gamma L & 1 - \delta L \end{pmatrix} \quad (39)$$

Comme le suggère l'indication donnée en fin d'exercice, notons que

$$\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix}.$$

L'équation (39) s'écrit alors

$$\begin{pmatrix} 1 - \alpha L & -\beta L \\ -\gamma L & 1 - \delta L \end{pmatrix} Z_t = u_t$$

et donc

$$\begin{pmatrix} 1 - \delta L & \beta L \\ \gamma L & 1 - \alpha L \end{pmatrix} \begin{pmatrix} 1 - \alpha L & -\beta L \\ -\gamma L & 1 - \delta L \end{pmatrix} Z_t = \begin{pmatrix} 1 - \delta L & \beta L \\ \gamma L & 1 - \alpha L \end{pmatrix} u_t,$$

soit, compte tenu de la relation matricielle décrite ci-dessus,

$$\begin{pmatrix} (1 - \alpha L)(1 - \delta L) + \beta \gamma L^2 & 0 \\ 0 & (1 - \alpha L)(1 - \delta L) - \beta \gamma L^2 \end{pmatrix} Z_t = \begin{pmatrix} 1 - \delta L & \beta L \\ \gamma L & 1 - \alpha L \end{pmatrix} u_t,$$

soit

$$(1 - [\alpha + \delta] L + [\alpha \delta - \beta \gamma] L^2) \begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 1 - \delta L & \beta L \\ \gamma L & 1 - \alpha L \end{pmatrix} u_t,$$

soit

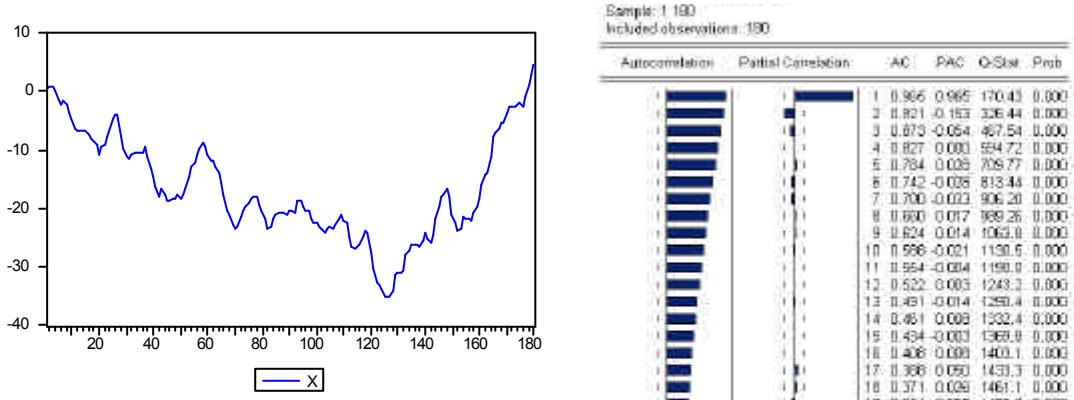
$$\begin{cases} (1 - [\alpha + \delta] L + [\alpha \delta - \beta \gamma] L^2) X_t = (1 - \delta L) \varepsilon_t + \beta L \eta_t \\ (1 - [\alpha + \delta] L + [\alpha \delta - \beta \gamma] L^2) Y_t = \gamma L \varepsilon_t + (1 - \alpha L) \eta_t \end{cases}.$$

En utilisant une nouvelle fois l'approche de l'exercice 1 (et de l'exercice 3), on peut noter que les processus $(1 - \delta L) \varepsilon_t + \beta L \eta_t$ et $\gamma L \varepsilon_t + (1 - \alpha L) \eta_t$ sont des processus $MA(1)$. Et donc, finalement, (X_t) et (Y_t) sont des processus $ARMA(2, 1)$.

L'hypothèse la plus importante pour que ces deux processus aient effectivement une forme $ARMA(2, 1)$ est que $\alpha \delta \neq \beta \gamma$. Il existe en fait deux autres hypothèses, à savoir que les polynômes autorégressifs et moyenne-mobiles n'aient pas de racine commune. Cette condition est toutefois relativement difficile à expliciter de façon analytique.

Exercice 21 Le but de cet exercice est de montrer que les modèles $ARIMA$ sont préférables aux modèles AR avec des racines proches du disque unité, pour faire de la prévision.

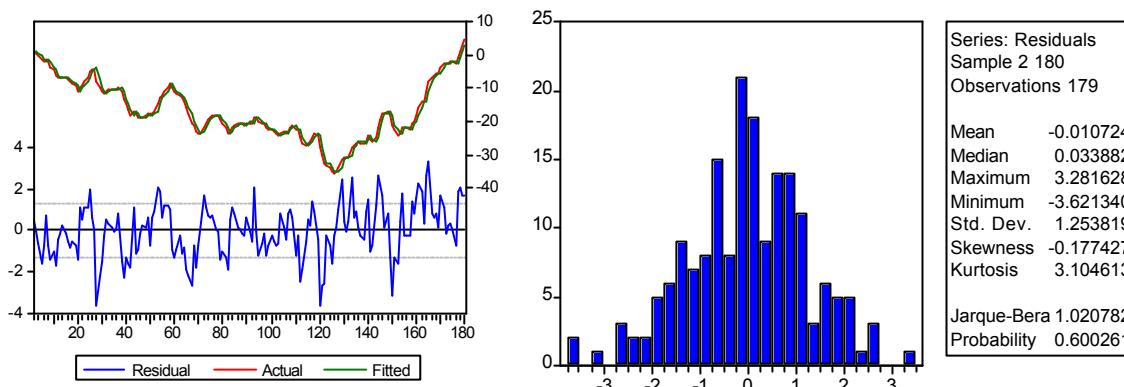
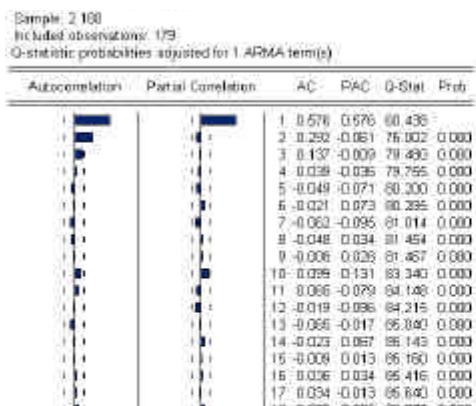
Considérons pour cela la série suivante (X_t) suivante,



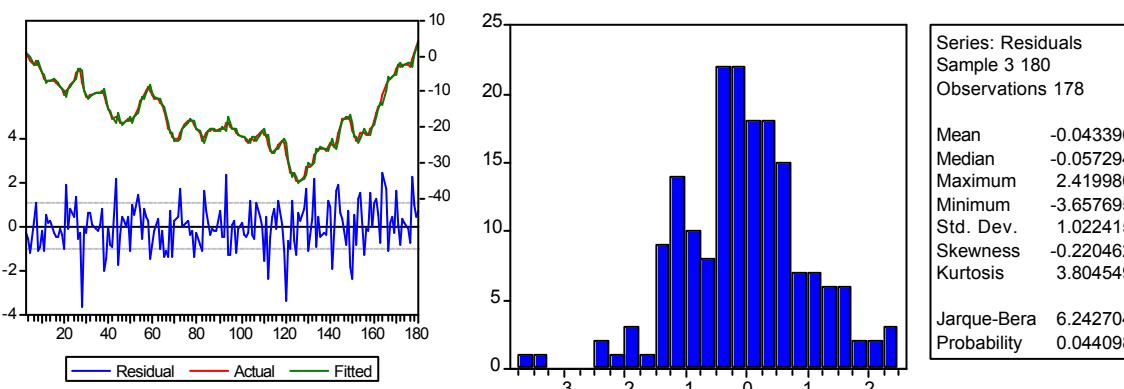
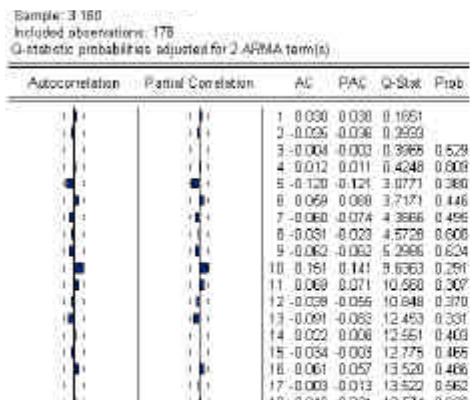
1) Au vue de l'autocorrélogramme, on hésite entre un modèle $AR(1)$ et un modèle $AR(2)$. Justifiez ces deux choix. Commentez les sorties suivantes, et précisez quel modèle vous retiendriez ? (les sorties présentes respectivement l'estimation des coefficients, l'autocorrélogramme des résidus, puis une représentation graphique suivi de l'histogramme

des résidus)

Dependent Variable:	X			
Method:	Least Squares			
Sample(adjusted):	2 180			
Included observations:	179 after adjusting endpoints			
Convergence achieved after 2 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.997954	0.004974	200.6441	0.0000
R-squared	0.980131	Mean dependent var	-16.62738	
Adjusted R-squared	0.980131	S.D. dependent var	8.895314	
S.E. of regression	1.253865	Akaike info criterion	3.295909	
Sum squared resid	279.8476	Schwarz criterion	3.313716	
Log likelihood	-293.9839	Durbin-Watson stat	0.836582	
Inverted AR Roots	1.00			



Dependent Variable:	X			
Method:	Least Squares			
Sample(adjusted):	3 180			
Included observations:	178 after adjusting endpoints			
Convergence achieved after 3 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	1.580110	0.061657	25.62737	0.0000
AR(2)	-0.583467	0.061661	-9.462509	0.0000
R-squared	0.986550	Mean dependent var	-16.72500	
Adjusted R-squared	0.986473	S.D. dependent var	8.823730	
S.E. of regression	1.026244	Akaike info criterion	2.900860	
Sum squared resid	185.3589	Schwarz criterion	2.936610	
Log likelihood	-256.1765	Durbin-Watson stat	1.932620	
Inverted AR Roots	.99	.59		



2) En regardant de plus près l'allure de la série, on souhaite vérifier que la série n'est pas intégrée. L'hypothèse de présence de racine unité est-elle vérifiée (la sortie du haut présentant le test ADF sans constante, en bas avec) ? On

posera $Y_t = X_t - X_{t-1}$ (représentée ci-dessous à droite)

ADF Test Statistic	-0.824171	1% Critical Value*	-2.5771
		5% Critical Value	-1.9415
		10% Critical Value	-1.6166

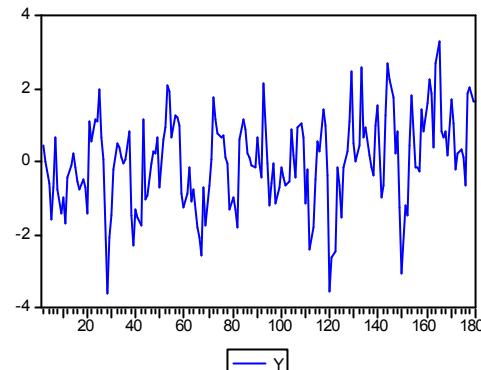
*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(X)

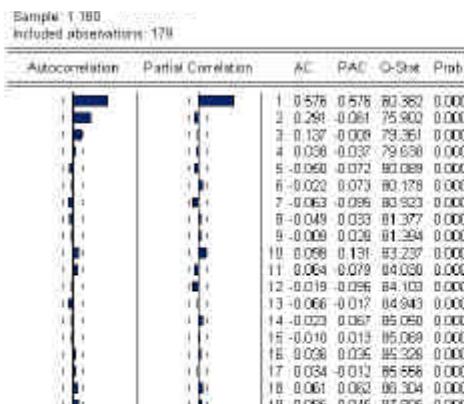
ADF Test Statistic	-1.467257	1% Critical Value*	-3.4682
		5% Critical Value	-2.8777
		10% Critical Value	-2.5753

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(X)



dont on observe l'autocorrélogramme suivant (ainsi que le test ADF avec constante à droite)



ADF Test Statistic	-6.326442	1% Critical Value*	-3.4684
		5% Critical Value	-2.8778
		10% Critical Value	-2.5754

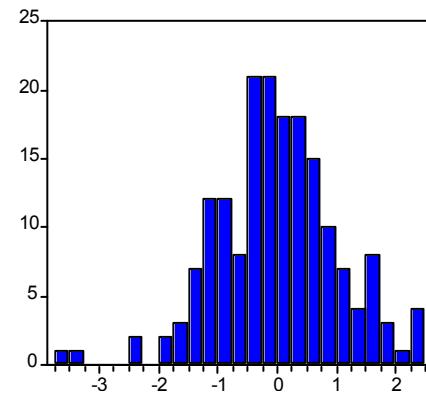
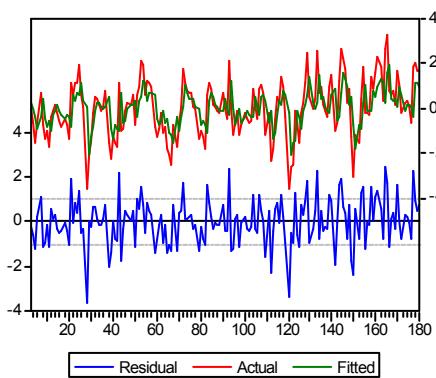
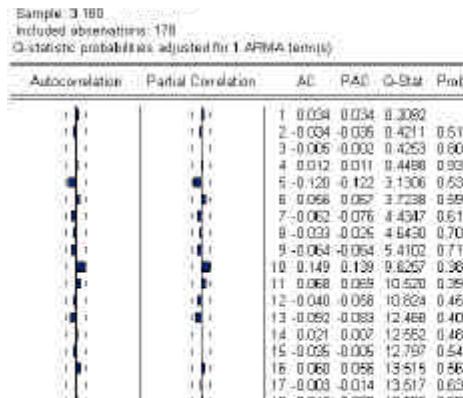
*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(Y)

Quel(s) modèle(s) proposeriez vous pour modéliser la série (Y_t) ?

Après réflexion, on propose de modéliser (Y_t) à l'aide d'un processus AR(1) et l'on obtient les sorties suivantes

Dependent Variable:	Y
Method:	Least Squares
Date:	12/03/03 Time: 12:46
Sample(adjusted):	3 180
Included observations:	178 after adjusting endpoints
Convergence achieved after 2 iterations	
Variable	Coefficient
AR(1)	0.581741
R-squared	0.335086
Adjusted R-squared	0.335086
S.E. of regression	1.025313
Sum squared resid	186.0743
Log likelihood	-256.5193
Inverted AR Roots	.58



Commentez ces sorties.

3) En utilisant ces deux premières questions, proposer deux modèles pour la série (X_t) . On écrira les modèles sous la forme

$$\begin{cases} (1) & (1 - \alpha L)(1 - \beta L) X_t = \varepsilon_t \\ (2) & (1 - L)(1 - \gamma L) X_t = \eta_t \end{cases} \quad (40)$$

(avec $\beta = 0$ si l'on a retenu un modèle AR(1) à la première question, $\beta \neq 1$ sinon). On précisera les valeurs de α , β , γ , ainsi que la variance des bruits.

4) Donnez, dans les deux cas, une méthode pour obtenir une prévision, faite à la date T , pour un horizon h , i.e. $T X_{T+h}$.

5) Donnez, dans les deux cas, la variance de l'erreur de prévision, à horizon 1 puis à horizon 2. Commentez.

1. Il ne s'agissait pas, pour l'instant, de commenter la stationnarité (ou non) de la série. En effet, si l'on considère les autocorrélations partielles de la série, on peut noter que les deux premières (et seulement les deux premières) sont non-nulles Aussi, un modèle AR(2) pourrait être étudié. Néanmoins, la seconde corrélation partielle est à la limite de la borne de l'intervalle de confiance. Statistiquement, on peut donc aussi considérer que cette seconde autocorrélation est nulle, ce qui pousserait à étudier un modèle AR(1).

Les premières sorties correspondent à l'estimation du modèle AR(1) : si le coefficient est significatif, notons que le résidu associé n'est pas un bruit blanc. Il suffit de considérer le test de Ljung-Box (colonne de droite de l'autocorégogramme), même s'il suffisait de noter que les premières autocorrélations sont largement significative ($\hat{\rho}(1) \approx 0,576$). Un modèle AR(1) ne peut pas être retenu.

Les dernières sorties correspondent à l'estimation du modèle AR(2). Les deux coefficients sont significatifs. De plus, l'hypothèse de bruit blanc du résidu peut être ici validée : les autocorrélations sont à l'intervalle de confiance, et le test de Ljung-Box confirme l'hypothèse de bruit blanc. On notera d'ailleurs également que l'hypothèse de normalité est elle aussi validée.

Aussi, le modèle à retenir est ici le modèle AR(2).

2. L'hypothèse de stationnarité semble difficile à valider en considérant l'allure de la série : il est donc légitime de tester la présence, ou non, de racine unité.

La série n'étant pas centrée autour de 0, il convient de tester le modèle avec constante, dans le test de Dickey-Fuller. La sortie du bas confirme la présence de racine unité. On notera également pour la série différenciée $Y_t = X_t - X_{t-1}$ est stationnaire (confirmé par le test de Dickey-Fuller à droite).

En notant que seule la première autocorrélation partielle est significative, unmodèle AR(1) pourrait être testé. Ce modèle est d'ailleurs retenu si l'on considère les sorties : le coefficient autorégressif est significatif et l'hypothèse de bruit blanc du résidu est elle aussi validée.

3. Ces deux premières questions nous ont permis de valider deux modèles,

$$\begin{cases} (1 - 0,99L)(1 - 0,59L) X_t = \varepsilon_t \\ (1 - L)(1 - 0,58L) X_t = \eta_t \end{cases}$$

Compte tenu des différentes sorties, on peut noter que les variances associées aux deux bruits sont respectivement. Dans les deux cas, notons que la variance du bruit est quasiment unitaire.

4. Pour les deux modèles, on a exactement les mêmes prévisions : en effet, que ce soit la forme $AR(Z)$ $(1 - 0,99L)(1 - 0,59L) X_t = \varepsilon_t$ ou la forme AR intégrée, ces deux modèles s'écrivent, de façon plus "claire"

$$X_t = 1,58X_{t-1} - 0,58X_{t-2} + \varepsilon_t.$$

A la date T , les prévisions à horizon $T+1, T+2, T+3\dots$ sont

$$T X_{T+1} = 1,58X_T - 0,58X_{T-1}$$

$$\begin{aligned} T X_{T+2} &= 1,58 T X_{T+1} - 0,58 X_T = 1,58 \times [1,58 X_T - 0,58 X_{T-1}] - 0,58 X_T \\ &= [1,58^2 - 0,58] X_T - 1,58 \times 0,58 X_{T-1} = 1,92 X_T - 0,92 X_{T-1} \end{aligned}$$

$$\begin{aligned}
{}_{T+3}X_{T+3} &= 1,58 {}_T X_{T+2} - 0,58 {}_T X_{T+1} = 1,58 \times [[1,58^2 - 0,58] X_T - 1,58 \times 0,58 X_{T-1}] - 0,58 [1,58 X_T - 0,58 X_{T-1}] \\
&= [1,58 \times [1,58^2 - 0,58] - 0,58 \times 1,58] X_T + [0,58^2 - 1,58^2 \times 0,58] X_{T-1} \\
&= 2,11 X_T - 1,11 X_{T-1},
\end{aligned}$$

puis

$${}_{T+4}X_{T+4} = 1,89 X_T - 1,22 X_{T-1}$$

$${}_{T+5}X_{T+5} = 0,70 X_T - 1,09 X_{T-1}$$

...etc

$${}_{T+9}X_{T+9} = -52,69 X_T + 14,11 X_{T-1}.$$

On retrouve ici le résultat général d'un modèle $AR(Z)$, $X_t = \alpha X_{t-1} + \beta X_{t-2} + \varepsilon_t$,

$${}_{T+1}X_{T+1} = \alpha X_T + \beta X_{T-1}$$

$${}_{T+2}X_{T+2} = \alpha {}_{T+1}X_{T+1} + \beta X_T = \alpha [\alpha X_T + \beta X_{T-1}] + \beta X_T = [\alpha^2 + \beta] X_T + \alpha \beta X_{T-1}$$

$${}_{T+3}X_{T+3} = \alpha {}_{T+2}X_{T+2} + \beta {}_{T+1}X_{T+1} = \alpha [[\alpha^2 + \beta] X_T + \alpha \beta X_{T-1}] + \beta [\alpha X_T + \beta X_{T-1}] = \alpha [\alpha^2 + 2\beta] X_T + \beta [\alpha^2 + \beta] X_{T-1}$$

$$\begin{aligned}
{}_{T+4}X_{T+4} &= \alpha {}_{T+3}X_{T+3} + \beta {}_{T+2}X_{T+2} = \alpha [\alpha [\alpha^2 + 2\beta] X_T + \beta [\alpha^2 + \beta] X_{T-1}] + \beta [[\alpha^2 + \beta] X_T + \alpha \beta X_{T-1}] \\
&= \alpha^2 [\alpha^2 + 3\beta] X_T + \beta \alpha [\alpha^2 + 2\beta] X_{T-1}
\end{aligned}$$

et de façon plus générale

$${}_{T+k}X_{T+k} = \alpha {}_{T+k-1}X_{T+k-1} + \beta {}_{T+k-2}X_{T+k-2} = \alpha^{k-2} [\alpha^2 + (k-1)\beta] X_T + \beta \alpha^{k-3} [\alpha^2 + (k-2)\beta] X_{T-1}$$

(ce qui peut se montrer par récurrence).

5. (i) Calculons l'erreur de prévision pour le modèle (1), c'est à dire le modèle $AR(2)$. De façon générale,

$$(1 - \alpha L)(1 - \beta L) X_t = \varepsilon_t$$

peut s'inverser, de façon à avoir une écriture " $MA(\infty)$ ",

$$\begin{aligned}
X_t &= (1 - \alpha L)^{-1} (1 - \beta L)^{-1} \varepsilon_t = \left(\sum_{i=0}^{\infty} \alpha^i L^i \right) \left(\sum_{j=0}^{\infty} \beta^j L^j \right) \varepsilon_t \\
&= \sum_{i=0}^{\infty} \omega_i \varepsilon_{t-i} \text{ où } \omega_i = \sum_{k=0}^i \alpha^k \beta^{i-k}
\end{aligned}$$

(en utilisant des résultats sur les produits de séries entières). Cette expression permet en particulier d'en déduire l'erreur de prévision (cf cours pour plus de détails).

L'erreur de prévision à horizon 1, ${}_{T+1}\Delta_{T+1}$ est ε_{T+1} , dont la variance est $\sigma^2 = V(\varepsilon_{T+1}) \approx 1$. À horizon supérieur, les erreurs de prévision sont respectivement

$${}_{T+2}\Delta_{T+2} = X_{T+2} - {}_T X_{T+2} = \varepsilon_{T+2} + [\alpha + \beta] \varepsilon_{T+1}$$

$${}_{T+3}\Delta_{T+3} = X_{T+3} - {}_T X_{T+3} = \varepsilon_{T+3} + [\alpha + \beta] \varepsilon_{T+2} + [\alpha^2 + \alpha\beta + \beta^2] \varepsilon_{T+1}$$

$${}_{T+4}\Delta_{T+4} = X_{T+4} - {}_T X_{T+4} = \varepsilon_{T+4} + [\alpha + \beta] \varepsilon_{T+3} + [\alpha^2 + \alpha\beta + \beta^2] \varepsilon_{T+2} + [\alpha^3 + \alpha^2\beta + \alpha\beta^2 + \beta^3] \varepsilon_{T+1}$$

dont les variances respectives sont

$$V_1 = \sigma^2, V_2 = (1 + [\alpha + \beta]^2) \sigma^2,$$

$$V_3 = (1 + [\alpha + \beta]^2 + [\alpha^2 + \alpha\beta + \beta^2]^2) \sigma^2$$

$$V_4 = (1 + [\alpha + \beta]^2 + [\alpha^2 + \alpha\beta + \beta^2]^2 + [\alpha^3 + \alpha^2\beta + \alpha\beta^2 + \beta^3]^2) \sigma^2.$$

(ii) Dans le cas du modèle $AR(1)$ intégré, il n'est pas possible d'utiliser la forme $MA(\infty)$, puisque le polynôme $(1 - L)$ n'est pas inversible ! Toutefois, en reprenant ce qui a été fait dans les notes de cours, on notera que, pour un processus $ARIMA(1, 1, 0)$, de façon formelle,

$$(1 - L)(1 - \phi L)X_t = \eta_t$$

peut se réécrire

$$(1 - L)X_t = (1 - \phi L)^{-1}\eta_t = \eta_t + \phi\eta_{t-1} + \phi^2\eta_{t-2} + \phi^3\eta_{t-3} + \dots = X_t - X_{t-1}.$$

On a alors

$${}_T X_{T+1} = (1 + \phi)X_T - \phi X_{T-1} = X_T + \phi(X_T - X_{T-1}) = X_T + \phi[\eta_T + \phi\eta_{T-1} + \phi^2\eta_{T-2} + \phi^3\eta_{T-3} + \dots]$$

alors que

$$X_{T+1} = X_T + \eta_{T+1} + \phi\eta_T + \phi^2\eta_{T-1} + \phi^3\eta_{T-2} + \dots$$

Aussi,

$${}_T \Delta_{T+1} = X_{T+1} - {}_T X_{T+1} = \eta_{T+1}.$$

De façon analogue,

$$\begin{aligned} {}_T X_{T+2} &= (1 + \phi){}_T X_{T+1} - \phi X_T = (1 + \phi)[X_T + \phi[\eta_T + \phi\eta_{T-1} + \phi^2\eta_{T-2} + \phi^3\eta_{T-3} + \dots]] - \phi X_T \\ &= X_T + \phi(1 + \phi)[\eta_T + \phi\eta_{T-1} + \phi^2\eta_{T-2} + \phi^3\eta_{T-3} + \dots] \end{aligned}$$

alors que

$$\begin{aligned} X_{T+2} &= X_{T+1} + \eta_{T+2} + \phi\eta_{T+1} + \phi^2\eta_T + \phi^3\eta_{T-1} + \dots \\ &= [X_T + \eta_{T+1} + \phi\eta_T + \phi^2\eta_{T-1} + \phi^3\eta_{T-2} + \dots] + \eta_{T+2} + \phi\eta_{T+1} + \phi^2\eta_T + \phi^3\eta_{T-1} + \dots \\ &= X_T + \eta_{T+2} + (1 + \phi)\eta_{T+1} + \phi(1 + \phi)\eta_T + \phi^2(1 + \phi)\eta_{T-1} + \phi^3(1 + \phi)\eta_{T-2} + \dots \end{aligned}$$

Aussi,

$${}_T \Delta_{T+2} = \eta_{T+2} + (1 + \phi)\eta_{T+1}.$$

En itérant une nouvelle fois afin d'expliciter clairement le résultat,

$$\begin{aligned} {}_T X_{T+3} &= (1 + \phi){}_T X_{T+2} - \phi {}_T X_{T+1} = (1 + \phi)[X_T + \phi(1 + \phi)[\eta_T + \phi\eta_{T-1} + \phi^2\eta_{T-2} + \phi^3\eta_{T-3} + \dots]] \\ &\quad - \phi[X_T + \phi[\eta_T + \phi\eta_{T-1} + \phi^2\eta_{T-2} + \phi^3\eta_{T-3} + \dots]] \\ &= X_T + (1 + \phi + \phi^2)[\phi\eta_T + \phi^2\eta_{T-1} + \phi^3\eta_{T-2} + \phi^4\eta_{T-3} + \dots] \end{aligned}$$

alors que

$$\begin{aligned} X_{T+3} &= X_{T+2} + \eta_{T+3} + \phi\eta_{T+2} + \phi^2\eta_{T+1} + \phi^3\eta_T + \dots \\ &= [X_T + \eta_{T+2} + (1 + \phi)\eta_{T+1} + \phi(1 + \phi)\eta_T + \phi^2(1 + \phi)\eta_{T-1} + \phi^3(1 + \phi)\eta_{T-2} + \dots] \\ &\quad + \eta_{T+3} + \phi\eta_{T+2} + \phi^2\eta_{T+1} + \phi^3\eta_T + \dots \\ &= X_T + \eta_{T+3} + (1 + \phi)\eta_{T+2} + (1 + \phi + \phi^2)\eta_{T+1} + \phi(1 + \phi + \phi^2)\eta_{T-1} + \phi^2(1 + \phi + \phi^2)\eta_{T-2} + \dots \end{aligned}$$

Aussi,

$${}_T \Delta_{T+3} = \eta_{T+3} + (1 + \phi)\eta_{T+2} + (1 + \phi + \phi^2)\eta_{T+1}.$$

De façon générale, on en déduit que

$${}_T \Delta_{T+h} = \eta_{T+h} + (1 + \phi)\eta_{T+h-1} + (1 + \phi + \phi^2)\eta_{T+h-2} + \dots + (1 + \phi + \dots + \phi^{h-1})\eta_{T+1}$$

On en déduit alors les variances respectives sont

$$V_1 = \sigma^2, V_2 = \left(1 + [1 + \phi]^2\right)\sigma^2,$$

$$V_3 = \left(1 + [1 + \phi]^2 + [1 + \phi + \phi^2]^2\right)\sigma^2$$

$$V_4 = \left(1 + [1 + \phi]^2 + [1 + \phi + \phi^2]^2 + [1 + \phi + \phi^2 + \phi^3]^2 \right) \sigma^2.$$

(iii) si l'on effectue les calculs, on notera que les chiffres sont ici grossièrement égaux, dans les deux cas. En fait, on peut noter que l'on est dans un cas où $(1 - \alpha L)(1 - \beta L) \approx (1 - L)(1 - \phi L)$, soit

$$\alpha + \beta \approx 1 + \phi \text{ et } \alpha\beta \approx \phi.$$

Aussi, dans le cas (i) - forme AR(2), si l'on suppose avoir la même variance dans les deux cas (ce qui est d'ailleurs le cas à 10^{-3} près),

$$V_2^{(i)} = \left(1 + [\alpha + \beta]^2 \right) \sigma^2 \approx \left(1 + [1 + \phi]^2 \right) \sigma^2 = V_2^{(ii)}$$

$$\begin{aligned} V_3^{(i)} - V_2^{(i)} &= [\alpha^2 + \alpha\beta + \beta^2]^2 \sigma^2 = ([\alpha + \beta]^2 - \alpha\beta)^2 \sigma^2 \\ &\approx ([1 + \phi]^2 - \phi) \sigma^2 = (1 + \phi + \phi^2) \sigma^2 = V_3^{(ii)} - V_2^{(ii)} \end{aligned}$$

$$\begin{aligned} V_4^{(i)} - V_3^{(i)} &= [\alpha^3 + \alpha^2\beta + \alpha\beta^2 + \beta^3]^2 \sigma^2 = [(\alpha + \beta)^3 - 2\alpha\beta(\alpha + \beta)]^2 \sigma^2 \\ &\approx ([1 + \phi]^3 - 2\phi(1 + \phi)) \sigma^2 = (1 + \phi + \phi^2 + \phi^3) \sigma^2 = V_4^{(ii)} - V_3^{(ii)} \end{aligned}$$

Du point de vue de la prévision, il est donc équivalent de considérer un modèle AR(2) ou de supposer que la série différenciée suit un modèle AR(1) : les prévisions et les intervalles de confiance associés sont quasiment identiques.

Exercice 22 On dispose de T observations d'une série temporelle, X_1, \dots, X_T . On suppose que l'on se trouve à la date T et que l'on souhaite prévoir la valeur X_{T+h} . On notera \hat{X}_{T+h} une telle prévision, effectuée à la date T pour un horizon $h \in \mathbb{N}^*$. Une méthode proposée pour fournir une prévision est la suivante : soit β une constante, appelée constante de lissage, appartenant à $]0, 1[$, et considérons

$$\hat{X}_{T+1} = (1 - \beta) \sum_{j=0}^{T-1} \beta^j X_{T-j}. \quad (44)$$

- 1) Donner une interprétation du coefficient β , en particulier que signifie β proche de 0 et β proche de 1 ?
- 2) On supposera alors que $\hat{X}_{T+h} = \hat{X}_{T+1}$. Montrer que la prévision \hat{X}_{T+h} peut s'écrire comme une moyenne pondérée entre la prévision faite à la date précédente (en $T-1$, i.e. \hat{X}_{T-1}) et la dernière observation X_T . Réécrire cette expression comme la dernière prévision corrigée d'un terme proportionnel à la dernière erreur de prévision,

$$\hat{X}_{T+h} = \hat{X}_{T-1} + \alpha(X_T - \hat{X}_{T-1}). \quad (45)$$

Donner l'expression du coefficient α .

- 3) On considérera μ^* défini par

$$\mu^* = \arg \min_{\mu \in \mathbb{R}} \sum_{j=0}^{T-1} \beta^j (X_{T-j} - \mu)^2. \quad (46)$$

Donner l'expression de μ^* . Montrer en particulier que

$$\hat{X}_{T+h} = \frac{1}{1 - \beta^T} \mu^*. \quad (47)$$

De cette dernière relation, on notera que pour T suffisamment grand (c'est à dire si l'on a beaucoup d'observations), $\hat{X}_{T+h} \approx \mu^*$, c'est à dire que \hat{X}_{T+h} peut être interprété comme la constante qui s'ajuste le mieux à la série "au voisinage" de X_T .

4) Comme on peut le voir, cette méthode de prévision pose un gros problème : le choix de la constante de lissage β . On dira que la prévision est "optimale" si $\varepsilon_T = X_T - \hat{X}_{T-1}$ est l'innovation du processus (X_t). Montrer que sous cette hypothèse, en utilisant la relation (45),

$$(1 - L) X_{T+1} = (1 - \beta L) \varepsilon_{T+1},$$

c'est à dire que (X_t) est un processus ARIMA(0, 1, 1).

1. β est ici un coefficient permettant de pondérer plus ou moins les observations plus ou moins lointaines. Aussi, si $\beta \rightarrow 0$, on notera que toutes les observations, hormis la première ($j = 0$) seront très peu pondérées : dans ce cas, on priviliera les toutes dernières observations, au détriment des plus anciennes, dont le poids sera très très faible. De façon analogue, si $\beta \rightarrow 1$, même les anciennes observations auront un poids non négligeable.

2. A la date $T - 1$, une prévision à horizon h a été effectuée, en posant

$${}_{T-1}X_{T+h} = (1 - \beta) \sum_{j=0}^{T-2} \beta^j X_{T-1-j} = (1 - \beta) [X_{T-1} + \beta X_{T-2} + \beta^2 X_{T-3} + \dots].$$

De façon analogue, la prévision faite à la date T , en considérant

$$\begin{aligned} {}_T X_{T+h} &= (1 - \beta) \sum_{j=0}^{T-1} \beta^j X_{T-j} = (1 - \beta) [X_T + \beta X_{T-1} + \beta^2 X_{T-2} + \beta^3 X_{T-3} + \dots] \\ &= (1 - \beta) X_T + \underbrace{\beta(1 - \beta) [X_{T-1} + \beta X_{T-2} + \beta^2 X_{T-3} + \dots]}_{={}_{T-1}X_{T+h}} \end{aligned}$$

aussi, ${}_T X_{T+h}$ est obtenu comme moyenne pondérée, avec respectivement des poids $(1 - \beta)$ et β , de la dernière observation X_T , et de l'ancienne prévision ${}_{T-1}X_{T+h}$. On retrouve d'ailleurs ici que si $\beta \rightarrow 0$, on ne prend en compte que la dernière observation et pas la dernière prévision (on n'intègre plus le passé "lointain"), alors que si $\beta \rightarrow 1$, on est assez peu sensible à la toute dernière observation, compte tenu du poids important du passé "lointain".

On peut d'ailleurs réécrire

$$\begin{aligned} {}_T X_{T+h} &= (1 - \beta) X_T + \beta {}_{T-1}X_{T+h} = (1 - \beta) X_T + [1 - (1 - \beta)] {}_{T-1}X_{T+h} \\ &= {}_{T-1}X_{T+h} + (1 - \beta) [X_T - {}_{T-1}X_{T+h}], \end{aligned}$$

soit $\alpha = 1 - \beta$: $1 - \beta$ est alors le poids donné à la correction pour erreur de prédiction.

3. On note ici

$$\mu^* = \arg \min_{\mu \in \mathbb{R}} \sum_{j=0}^{T-1} \beta^j (X_{T-j} - \mu)^2.$$

On sait (cf cours d'économétrie de licence) que la moyenne est la constante minimisant l'erreur quadratique : en effet, si

$$h : \mu \mapsto \sum_{j=0}^{T-1} \beta^j (X_{T-j} - \mu)^2,$$

alors

$$\frac{dh(\mu)}{d\mu} = -2 \sum_{j=0}^{T-1} \beta^j (X_{T-j} - \mu).$$

Aussi, $dh(\mu)/d\mu = 0$ si et seulement si

$$\sum_{j=0}^{T-1} \beta^j (X_{T-j} - \mu) = 0 \text{ soit } \mu \left(\sum_{j=0}^{T-1} \beta^j \right) = \sum_{j=0}^{T-1} \beta^j X_{T-j}$$

d'où finallement

$$\mu^* = \frac{1 - \beta^T}{1 - \beta} \sum_{j=0}^{T-1} \beta^j X_{T-j}.$$

On peut déduire de cette expression que

$${}_T X_{T+h} = \frac{1}{1 - \beta} \sum_{j=0}^{T-1} \beta^j X_{T-j} = \frac{1}{1 - \beta^T} \mu^*,$$

aussi, avec un long historique (T grand), on note que ${}_T X_{T+h} \approx \mu^*$ (d'autant plus que β est petite) : cette prévision est alors la constante qui s'ajuste au mieux (au sens de l'erreur quadratique) à la série au voisinage de X_T .

4. Sous l'hypothèse de prévision optimale, l'erreur de prévision $\varepsilon_T = X_T - \hat{X}_{T-1} X_T$ est l'innovation du processus (X_t). Et donc, en utilisant (8),

$$\hat{X}_{T+h} = \hat{X}_{T-1} X_{T+h} + (1 - \beta) [X_T - \hat{X}_{T-1} X_{T+h}].$$

En remplaçant $\hat{X}_{T-1} X_T$ par $X_T - \varepsilon_T$, ou, de façon plus générale, $\hat{X}_{T+h} = \hat{X}_T X_{T+1}$ par $X_{T+1} - \varepsilon_{T+1}$, on peut réécrire cette expression sous la forme

$$\hat{X}_{T+1} = \hat{X}_{T-1} X_T + (1 - \beta) [X_T - \hat{X}_{T-1} X_T] \text{ ou } X_{T+1} - \varepsilon_{T+1} = X_T - \varepsilon_T + (1 - \beta) \varepsilon_T$$

soit

$$X_{T+1} - X_T = \varepsilon_{T+1} - \beta \varepsilon_T, \text{ soit } (1 - L) X_{T+1} = (1 - \beta L) \varepsilon_{T+1}$$

aussi, le processus (X_t) est un processus $ARMA(0, 1, 1)$.

5 Compléments : simulation de séries temporelles

...

5.1 Simulation de processus en temps discret

Avant de parler de simulation de processus, il convient de rappeler quelques méthodes de simulations de variables aléatoires, en particulier des lois gaussiennes. On supposera ici qu'il est possible de générer les nombres aléatoires (ou pseudo aléatoires) suivant une loi uniforme sur $[0, 1]$ (cf Fishman (1996))

5.1.1 Simulation d'un bruit blanc gaussien

La méthode la plus simple (*et la plus efficace*) pour simuler des variables gaussiennes indépendantes consiste à utiliser l'algorithme de Box-Müller, reposant sur le résultat suivant

Lemme 1 Si U et V sont des variables aléatoires uniformes sur $[0, 1]$ et indépendantes, alors $X = \sqrt{-2 \log U} \cos(2\pi V)$ et $Y = \sqrt{-2 \log U} \sin(2\pi V)$ suivent des loi gaussienne $\mathcal{N}(0, 1)$ indépendantes.

Ceci permet de simuler aisément un vecteur $\varepsilon \sim N(0, \mathbb{I}) \in \mathbb{R}^d$ en simulant d variables uniformes indépendantes. Pour information, si l'on cherche à simuler un vecteur $X \sim N(m, \Sigma)$ où la matrice Σ est une matrice de covariance (i.e. définie positive) que l'on supposera inversible, l'algorithme est le suivant :

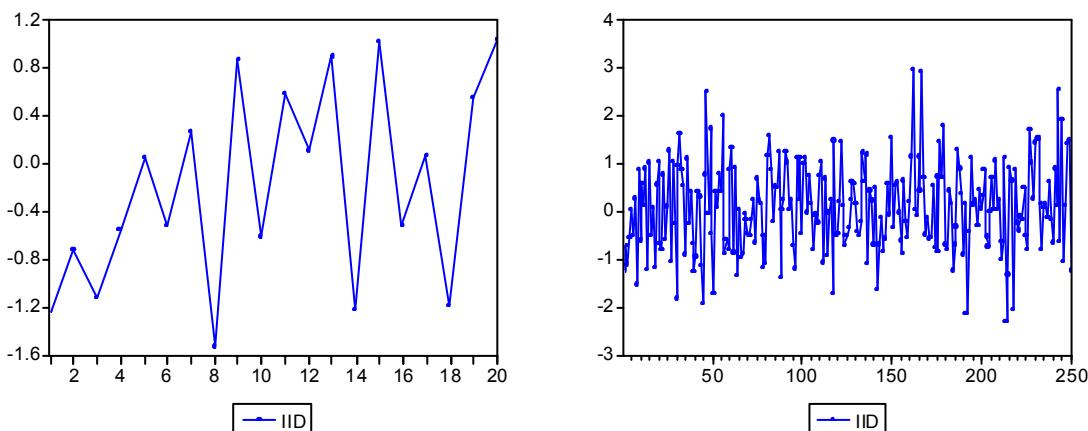
(i) trouver une matrice racine carrée de Σ , i.e. T telle que $TT' = \Sigma$ (ex: décomposition de Cholesky) : comme Σ est inversible, T aussi,

On peut noter que $Z = T^{-1}(X - m) \sim N(0, \mathbb{I})$, et donc

(ii) simuler n variables gaussiennes centrées réduites et indépendantes, que l'on notera Z

(iii) poser $X = m + TZ$, et alors $X \sim N(m, \Sigma)$.

Exemple 29 Le graphique ci-dessous correspond à la simulation d'un bruit blanc, respectivement pour 20 observations, et pour 250

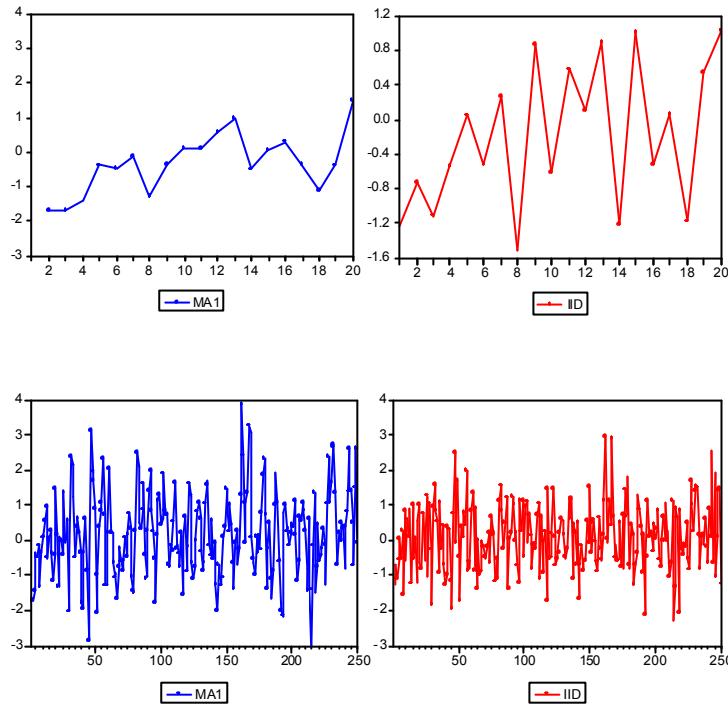


5.1.2 Simulation d'un processus ARMA

Simulation d'un processus $MA(q)$ Nous avons vu dans la partie précédente comment simuler une échantillon i.i.d. gaussien, $\varepsilon_1, \dots, \varepsilon_n$ correspondant à notre bruit blanc. De façon générale, pour simuler un processus $MA(q)$ l'algorithme est le suivant :

- (i) simuler $q + n$ valeurs initiales d'un bruit blanc $\varepsilon_{-q+1}, \dots, \varepsilon_0, \varepsilon_1, \dots, \varepsilon_n$
- (ii) à l'étape $i + 1$, on pose $x_{i+1} = \varepsilon_{i+1} + \theta_1\varepsilon_i + \theta_2\varepsilon_{i-1} + \dots + \theta_q\varepsilon_{i-q+1}$.

Exemple 30 Les graphiques suivant correspondent à la simulation d'un processus $MA(1)$ de coefficient $\theta_1 = 0.8$, avec $n = 20$ en haut et $n = 250$ en bas, et avec x_t à gauche, et ε_t à droite,



Simulation d'un processus $AR(p)$ Un processus $AR(p)$ étant un processus Markovien, simuler un processus $AR(p)$ revient à simuler une chaîne de Markov (*d'ordre p*)

Lemme 2 Une chaîne de Markov d'ordre 1 sur \mathbb{R} (ou plus généralement sur E , un espace mesurable) est une suite de variables aléatoires (X_t) , $t \in \mathbb{N}$ telle qu'il existe

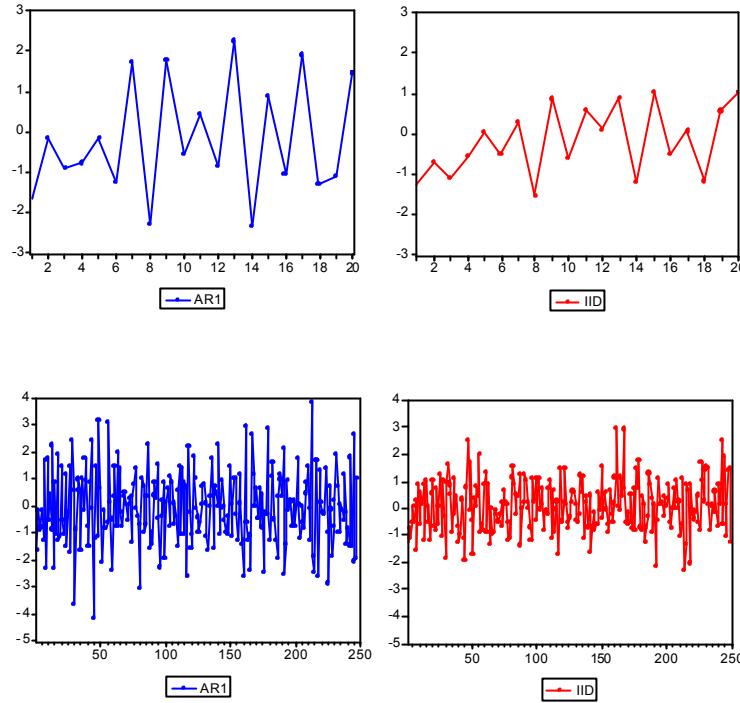
- (i) une suite (ε_t) de variables indépendantes et de même loi à valeurs
- (ii) une application mesurable Φ telle que pour tout $t \in \mathbb{N}$, $X_{t+1} = \Phi(t, X_t, \varepsilon_t)$

Pour simuler un processus $AR(p)$ l'algorithme est le suivant :

- (i) choisir ou simuler une valeur initiale x_0 et simuler un bruit blanc $\varepsilon_1, \dots, \varepsilon_n$
- (ii) à l'étape $i + 1$, on pose $x_{i+1} = \phi_1 x_i + \varepsilon_{i+1}$.

Exemple 31 Les graphiques suivant correspondent à la simulation d'un processus $AR(1)$ de coefficient $\phi_1 = -0.8$,

avec $n = 20$ en haut et $n = 250$ en bas, et avec x_t à gauche, et ε_t à droite,



De façon générale, pour simuler un processus $AR(p)$ l'algorithme est le suivant :

- (i) choisir ou simuler p valeurs initiales $x_0, x_{-1}, \dots, x_{-p+1}$ et simuler un bruit blanc $\varepsilon_1, \dots, \varepsilon_n$
- (ii) à l'étape $i + 1$, on pose $x_{i+1} = \phi_1 x_i + \phi_2 x_{i-1} + \dots + \phi_p x_{i-p+1} + \varepsilon_{i+1}$.

Simulation d'un ARMA(p, q) Un processus $ARMA(p, q)$, de la forme $\Phi(L)X_t = \Theta(L)\varepsilon_t$ peut se voir comme un processus $AR(p)$ où le 'bruit' associé suit un processus $MA(q)$: il convient alors de simuler un processus $MA(q)$ tout d'abord, puis d'utiliser l'algorithme de simulation d'un processus $AR(p)$ en remplaçant le bruit blanc par le processus $MA(q)$.

5.1.3 Simulation d'un processus $ARCH$

[A INSERER]

5.2 Introduction à la simulation de processus en temps continu

Les processus de diffusion qu'il peut être intéressant de simuler, en particulier en finance, sont généralement de la forme

$$dX_t = b(X_t)dt + a(X_t)dW_t, X_0 = x. \quad (48)$$

On suppose que b et a sont lipschitziennes, i.e. il existe $K > 0$ tel que $\|a(x) - a(y)\| + \|b(x) - b(y)\| \leq K\|x - y\|$ pour tout x, y . Par exemple, dans le cadre du modèle de Black et Scholes, la dynamique des prix, sous la probabilité risque neutre s'écrit

$$dX_t = rX_t dt + X_t \sigma dW_t$$

dont la solution est

$$X_t = X_0 \exp \left(\left[r - \frac{1}{2}\sigma^2 \right] t + \sigma W_t \right)$$

Pour simuler X_t , il suffit alors de simuler la valeur de (W_t) à la date t , i.e. une loi normale $\mathcal{N}(0, t)$.

Pour simuler (X_t) , on a recours à une approximation de (X_t) , correspondant à une discréétisation de l'équation (48). Pour cela, considérons une partition de $[0, T]$ en n subdivision, de taille T/n , et notons $t_i = iT/n$. Soit $\delta_n = T/n$ le pas de temps, et on noter $\Delta W_n^i = W_{t_i} - W_{t_{i-1}}$. La discréétisation d'Euler consiste à écrire

$$X_{t_i} = X_{t_{i-1}} + \int_{t_{i-1}}^{t_i} b(X_s)ds + \int_{t_{i-1}}^{t_i} a(X_s)dW_s \approx X_{t_{i-1}} + b(X_{t_{i-1}})\delta_n + a(X_{t_{i-1}})\Delta W_n^i$$

ce qui conduit à considérer le processus suivant

$$\begin{cases} \tilde{X}_0 = X_0 \\ \tilde{X}_{t_i} = \tilde{X}_{t_{i-1}} + b(\tilde{X}_{t_{i-1}}) \delta_n + a(\tilde{X}_{t_{i-1}}) \Delta W_n^i \end{cases} \quad (49)$$

et simuler (\tilde{X}_t) revient à simuler les accroissement $\Delta W_n^i \sim \mathcal{N}(0, \delta_n)$, i.e ce qui revient à simuler les lois gaussienne indépendantes, centrées, de variance δ_n .

Sous l'hypothèse où a et b sont lipschitziennes, il est alors possible de montrer (*application du lemme de Gronwall*) que, pour $p \geq 1$,

$$\mathbb{E} \left(\sup_{t \in [0, T]} \|X_t - \tilde{X}_t\|^p \right)^{1/p} \leq \frac{C}{\sqrt{n}}$$

c'est à dire que l'on a une convergence uniforme, dans L^p , du processus discréte (\tilde{X}_t) vers le processus (X_t) en $1/\sqrt{n}$. Ou de façon plus générale, en rajoutant des conditions de régularité sur a, b et g , on a que

$$\mathbb{E} \|g(X_T) - g(\tilde{X}_T)\| \leq \frac{C}{n}$$

(en utilisant le théorème d'Ito).

Remarque 23 L'approximation utilisée ici, correspondant au schéma d'Euler, repose sur l'approximation

$$\int_{t_{i-1}}^{t_i} a(X_s) dW_s \approx a(X_{t_{i-1}}) \Delta W_n^i$$

mais une approximation à un ordre supérieur est également possible. En utilisant le fait que $\int_{t_{i-1}}^{t_i} \Delta W_n^i dW_s = \frac{1}{2} [(\Delta W_n^i)^2 - \delta_n]$, peut approximer

$$\int_{t_{i-1}}^{t_i} a(X_s) dW_s \approx a(X_{t_{i-1}}) \Delta W_n^i + \frac{1}{2} a'(X_{t_{i-1}}) a(X_{t_{i-1}}) [(\Delta W_n^i)^2 - \delta_n]$$

Aussi, le schéma d'approximation (49) peut se réécrire

$$\begin{cases} \tilde{X}_0 = X_0 \\ \tilde{X}_{t_i} = \tilde{X}_{t_{i-1}} + b(\tilde{X}_{t_{i-1}}) \delta_n + a(\tilde{X}_{t_{i-1}}) \Delta W_n^i + \frac{1}{2} a'(X_{t_{i-1}}) a(X_{t_{i-1}}) [(\Delta W_n^i)^2 - \delta_n] \end{cases}$$

(en fait, le terme de dérive b ayant une contribution inférieure dans l'erreur d'approximation par rapport au terme de diffusion, il est inutile de le corriger). On peut alors noter que dans le cas de cette approximation, on a, pour a et b deux fois continuellement dérivable, et pour $p \geq 1$

$$\max_{i=1, \dots, n} \mathbb{E} \left(\|X_{t_i} - \tilde{X}_{t_i}\|^p \right)^{1/p} \leq \frac{C}{n}$$

c'est à dire que la convergence L^p se fait ici en $1/n$, et non plus en $1/\sqrt{n}$ (dans le cas du schéma d'Euler). Les références sur ce sujet sont Peter E. Kloeden et Eckhard Platen, 'Numerical Solution of Stochastic Differential Equations', Springer-Verlag (1995) ou encore Peter E. Kloeden, Eckhard Platen et Henri Schurz, 'Numerical Solution of SDE Through Computer Experiment', Springer-Verlag (1997).

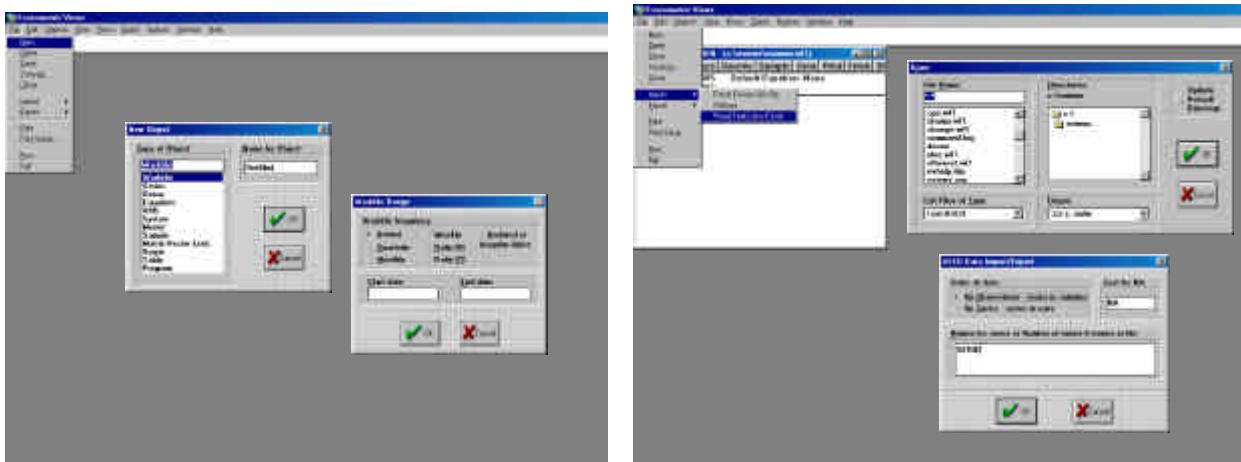
6 Compléments : Logiciels de séries temporelles

6.1 Introduction à EViews

La version 3.1 de EViews semble pouvoir être téléchargée gratuitement sur le site <http://ysdxf.myetang.com/soft.htm>. La page étant en chinois, je décline toute responsabilité sur l'usage de cette version n'est pas autorisée par les éditeurs.

6.1.1 Les données sous EViews

La "work file" La première étape consiste à créer un *work file*, qui contiendra l'ensemble des séries chronologiques. La démarche à suivre est de cliquer sur **File** puis **New**, puis de donner un nom à la base. Il est alors nécessaire d'indiquer le format des données des séries qui seront utilisées (*comme cela est présenté ci-dessous à gauche*) : données annuelles (annual : 1950 1994), données trimestrielles (quarterly 1950:1 1994:4), données mensuelles (monthly 1950:01 1994:12), données journalières (daily 3:25:1950 12:7:1994), ou sans date (undated or irregular 1 174). Par défaut, deux séries seront alors créées, *c* qui correspondra, par défaut à la constante, et *resid* qui correspondra (*lors de régressions*) à la série des résidus.



Une fois créée cette *work file*, deux alternatives existent pour travailler avec des données : importer des données (*sous format texte ou excel, par exemple*), ou les générer. L'importation est relativement simple, à condition que les séries soient présentées sous forme de colonnes. La procédure d'importation est présentée ci-dessus à droite : il suffit de cliquer sur **File** puis **Import**, puis d'indiquer le format de la base à importer. Dans l'exemple présenté, il s'agit d'importer une série au format **txt** : on choisit **Read Text Lotus-Excel** puis on va chercher la base dans le répertoire où elle se trouve (*par défaut dans le répertoire où est installé Eviews*). Une fois sélectionné le fichier, une autre fenêtre demande de nommer les séries dans l'ordre où elle apparaissent : dans l'exemple considéré, il n'y avait qu'une série, que nous avons appelée '**Série**'¹⁸.

Pour générer une série, deux alternatives sont possibles :

- générer une série à partir d'une autre série X_t : ceci arrive lorsque l'on souhaite créer une série différencier. En cliquant sur **Generate Series**, il suffit de taper alors ' $Z=X-X(-1)$ ' pour générer la série $Z_t = X_t - X_{t-1}$, encore ' $Z=2X-3$ ' pour générer la série $Z_t = 2X_t - 3$. On peut alors générer, de façon générale, toute série $Z_t = \Phi(L)X_t$ où Φ est un polynôme.

- générer une série aléatoire : il est assez facile de simuler un échantillon gaussien, centré, de variance unitaire.

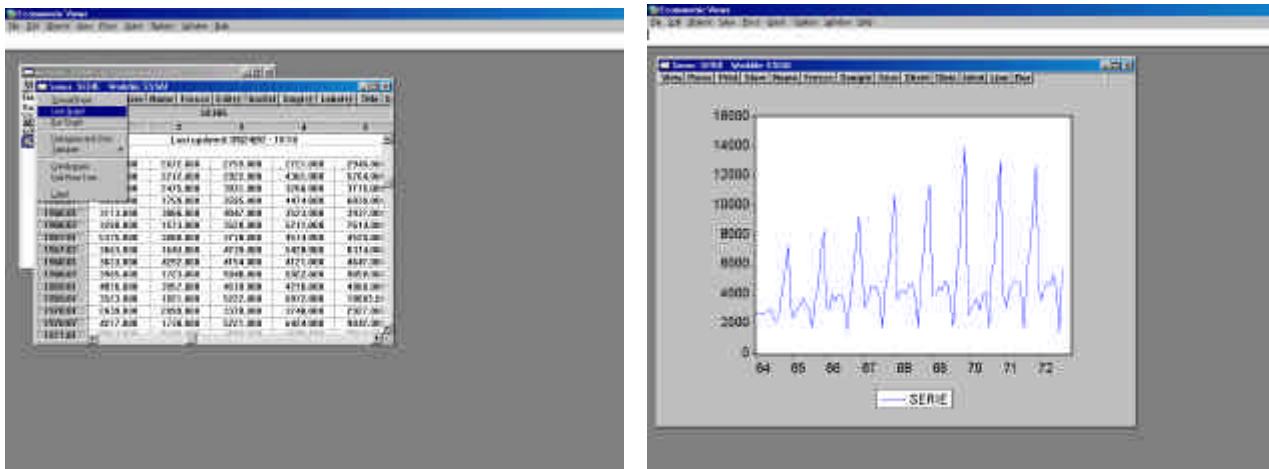
A partir de ces deux méthodes, il est possible de simuler et de générer toutes sortes de séries. Une fois générée la série, il est possible de la visualiser en cliquant dessus.

6.1.2 Les graphiques sous Eviews

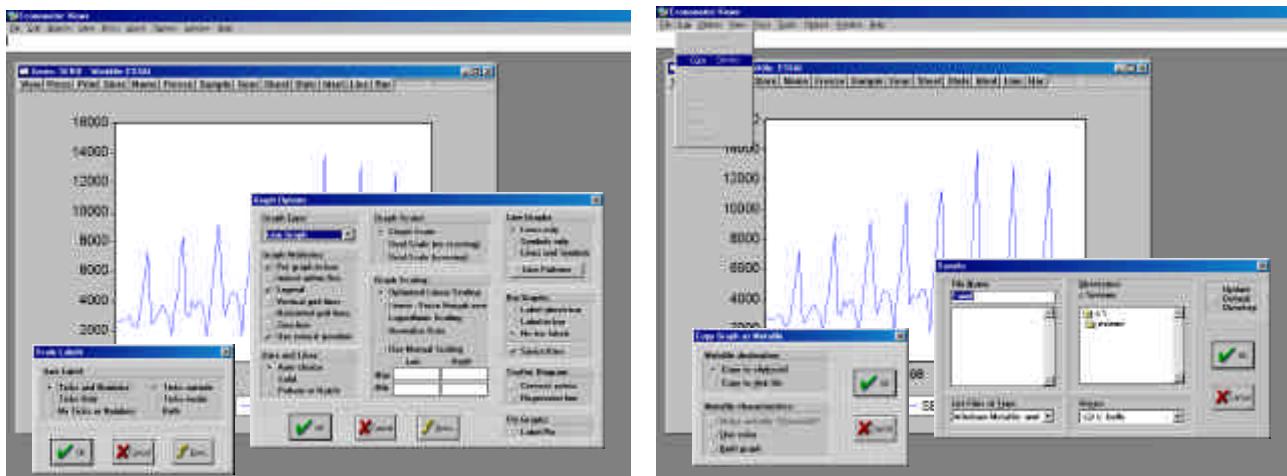
Plusieurs types de représentations sont possibles sous Eviews. Pour générer le graphique (*standard*) d'une série temporelle, il suffit de cliquer sur "**Line Graph**" (*ci-dessous à gauche*). Une fois le graphique obtenu (*ci-dessous au centre*), il est possible d'obtenir la fenêtre associée au module graphique, permettant de changer les différents

¹⁸Dans le cas de l'importation de séries d'un fichier MSExcel, Eviews demande à quelle cellule comme les bases : en commençant en **B2** par exemple, il prendra les séries **B2:Bn**, **C2:Cn**, **D2:Dn**,... où n correspond au nombre de valeurs des séries de la workfile (*nombre des valeurs+1 en fait*). Là encore les séries sont toujours en colonne.

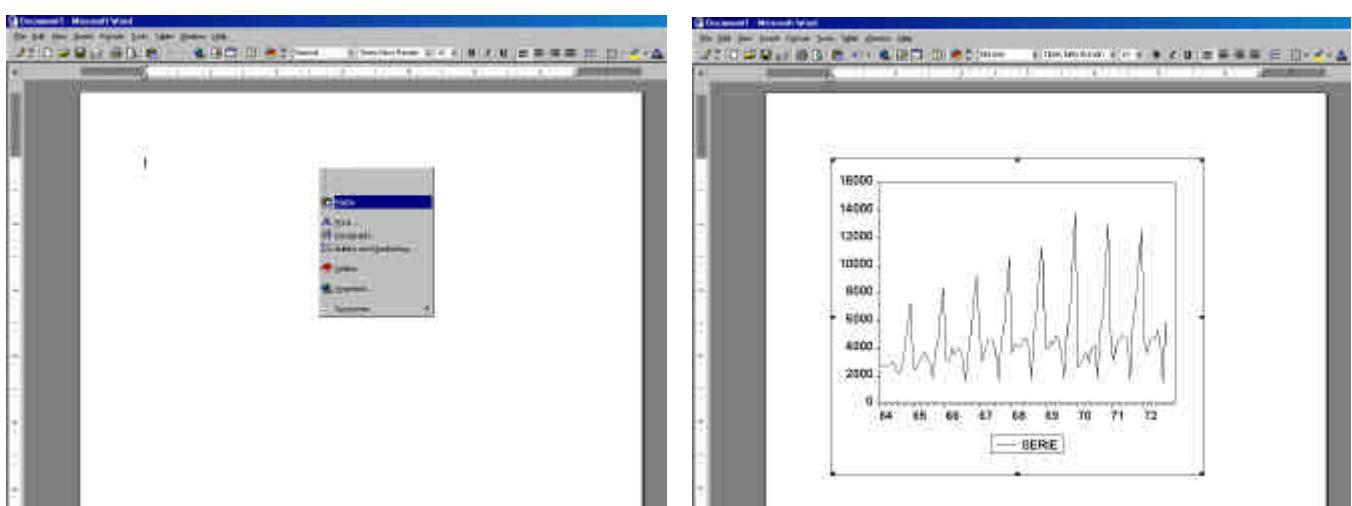
paramètres,



En particulier, le type de graphique peut être modifié (*lignes simples, lignes avec points pour les valeurs, histogramme...etc*), les échelles...etc



Il peut être alors intéressant de copier un graphique dans un document Word (ou tout autre traitement de texte). Le plus simple est de cliquer sur **Edit** puis **Copy** afin de copier le graphique en mémoire (plusieurs options sont néanmoins possibles : copier en couleur, copier en noir et blanc - cette dernière option étant la plus intéressante si l'on est amené à imprimer en noir et blanc, les traits de différentes couleurs étant alors transformés en traits de forme différentes : traits plein, pointillés, différentes épaisseurs...). Il est alors soit d'enregistrer le fichier en tant que fichier wmf, soit de le copier, avant de le coller dans un document externe,



Il suffit, une fois copier le graphique, de retourner sur le document Word et d'y coller l'image.

6.2 La régression linéaire sous Eviews

Considérons la sortie suivante, obtenue sous EViews,

```
LS // Dependent Variable is Y
Sample: 1 14
Included observations: 14

Variable Coefficient Std. Error T-Statistic Prob.
          (1)          (2)          (3)          (4)
C      32.89132  11.66331  2.820068  0.0182
X1     0.801901  0.298436  2.687012  0.0228
X2    -0.381362  0.156581 -2.435564  0.0351
X3    -0.037132  0.052023 -0.713768  0.4917

R-squared      0.702687 (5)   Mean dependent var  17.71429 (11)
Adjusted R-squared 0.613493 (6)  S.D. dependent var  4.177385 (12)
S.E. of regression 2.597069 (7) Akaike info criterion 2.143723 (13)
Sum squared resid 67.44767 (8) Schwartz criterion 2.326311 (14)
Log likelihood    -30.87120 (9) F-statistic        7.878181 (15)
Durbin-Watson stat 3.186886 (10) Prob(F-statistic) 0.005452 (16)
```

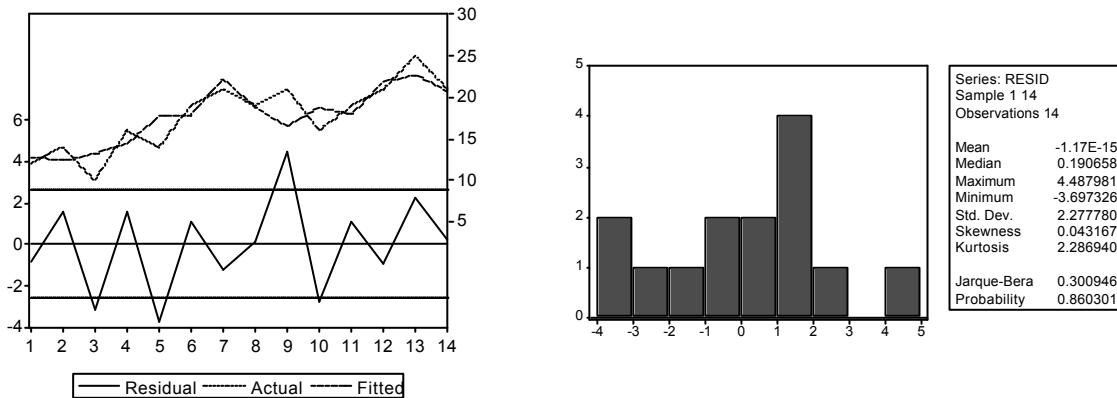
Ce type de sortie correspond à des sorties obtenues en économétrie, lors de l'estimation des paramètres du modèle linéaire

$$Y_i = \beta^1 X_i^1 + \beta^2 X_i^2 + \beta^3 X_i^3 + \beta^4 X_i^4 + \varepsilon_i$$

qui se réécrit de façon matricielle $Y_i = X_i \beta + \varepsilon_i$. Supposons que nous avons un échantillon Y_1, \dots, Y_n et X_1^i, \dots, X_n^i de n observations. On supposera l'hypothèse (H) sur les résidus ε_t , de stationnarité au second ordre :

- les résidus sont centrés $\mathbb{E}(\varepsilon_i) = 0$
 - la variance des résidus est constante : $V(\varepsilon_i) = 0$ (*homoscédasticité*)
 - les résidus sont non-autocorrélés : $\text{cov}(\varepsilon_i, \varepsilon_{i-h}) = 0$ pour tout $h \neq 0$
- Dans les procédures de tests, nous ajouterons une hypothèse supplémentaire :
- les résidus sont gaussiens : $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$

Dans l'exemple développé ci-dessus, les résidus, et la courbe \widehat{Y}_i ajustée sont représentés ci-dessous à gauche. La sortie de droite donne différentes statistiques sur $\widehat{\varepsilon}_i = Y_i - \widehat{Y}_i$,



Pour information, les observations étaient les suivantes,

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Y	12	14	10	16	14	19	21	19	21	16	19	21	25	21
X_1	2	1	3	6	7	8	8	5	5	8	4	9	12	7
X_2	45	43	43	47	42	41	32	33	41	38	32	31	35	29
X_3	121	132	154	145	129	156	132	147	128	163	161	172	174	180
\widehat{Y}	11.16	15.61	6.82	17.61	10.30	20.12	19.80	19.14	25.49	13.24	20.08	20.10	27.29	21.24
$\widehat{\varepsilon}$	-0.84	1.61	-3.18	1.61	-3.70	1.12	-1.20	0.14	4.49	-2.76	1.08	-0.90	2.29	0.24

(Source : Bourbonnais, 1998).

6.2.1 Estimation des paramètres (1) – (2)

L'estimation par la méthode des moindres carrés ordinaires donne

$$\hat{\beta} = (X'X)^{-1} X'Y$$

La valeur de cette estimation est donnée en (1).

Propriété 11 $\hat{\beta}$ est un estimateur sans biais, linéaire, de β

Preuve. $\mathbb{E}(\hat{\beta}) = (X'X)^{-1} X' \mathbb{E}(Y)$ puisque la matrice X est supposée constante, et que l'espérance est un opérateur linéaire. Et puisque $\mathbb{E}(\varepsilon_t) = 0$ par hypothèse, alors $\mathbb{E}(Y) = X\beta$ et donc $\mathbb{E}(\hat{\beta}) = (X'X)^{-1} X' X \beta = \beta$ ■

Théorème 3 Gauss-Markov - $\hat{\beta}$ est l'estimateur, linéaire en Y , sans biais de β , dont la variance est minimale

Cet estimateur sera qualifié de *BLUE* - Best Linear Unbiased Estimator (*meilleur estimateur linéaire sans biais*). Sous l'hypothèse où (ε_t) est gaussien, soit $\varepsilon_t \sim N(0, \sigma^2)$, on a le résultat suivant

Propriété 12 $\hat{\beta}$ est un estimateur gaussien, centré en β de variance $(X'X)^{-1} \sigma^2$

La variance σ^2 des résidus peut être estimée par

$$s^2 = \frac{1}{n-p-1} \|Y - X\hat{\beta}\|^2 = \frac{1}{n-p-1} \hat{\varepsilon}' \hat{\varepsilon}$$

où p est le nombre de paramètres. Cet estimateur est alors sans biais. L'écart-type des estimateurs, donné en (2) est alors estimé par

$$s(\hat{\beta}) = \sqrt{\frac{1}{n-p-1} (X'X)^{-1} \|Y - X\hat{\beta}\|^2}$$

On notera par la suite $\hat{Y}_t = X_t \hat{\beta}$.

Exemple 32 Dans l'exemple présenté dans l'introduction, on a

$$X = \begin{pmatrix} 1 & 2 & 45 & 121 \\ 1 & 1 & 43 & 132 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 7 & 29 & 180 \end{pmatrix}, X'X = \begin{pmatrix} 14 & 85 & 532 & 2094 \\ 85 & 631 & 3126 & 13132 \\ 532 & 3126 & 2066 & 78683 \\ 2094 & 13132 & 78683 & 317950 \end{pmatrix}$$

dont l'inverse est

$$(X'X)^{-1} = \begin{pmatrix} 20.17 & 0.015 & -0.231 & -0.076 \\ 0.015 & 0.013 & 0.001 & -0.001 \\ -0.231 & 0.012 & 0.003 & 0.001 \\ -0.076 & -0.001 & 0.001 & 0.000 \end{pmatrix}$$

Aussi

$$\underbrace{\begin{pmatrix} 20.17 & 0.015 & -0.231 & -0.076 \\ 0.015 & 0.013 & 0.001 & -0.001 \\ -0.231 & 0.012 & 0.003 & 0.001 \\ -0.076 & -0.001 & 0.001 & 0.000 \end{pmatrix}}_{(X'X)^{-1}} \underbrace{\begin{pmatrix} 248 \\ 1622 \\ 9202 \\ 37592 \end{pmatrix}}_{X'Y} = \underbrace{\begin{pmatrix} 32.891 \\ 0.802 \\ -0.381 \\ -0.037 \end{pmatrix}}_{\hat{\beta}}$$

Exemple 33 Toujours dans cet exemple, la variance des résidus est estimée par

$$s^2 = \frac{1}{14-3-1} \sum_{i=1}^{14} \hat{\varepsilon}_i^2 = \frac{67.45}{10} = 6.745$$

La matrice de variance-covariance des estimateurs est alors estimée par

$$\hat{\Sigma} = s^2 (X'X)^{-1} = \begin{pmatrix} 11.66 & * & * & * \\ * & 0.29 & * & * \\ * & * & 0.15 & * \\ * & * & * & 0.05 \end{pmatrix}$$

les valeurs sur la diagonale correspondant aux variances des coefficients de la régression.

6.2.2 Statistique de Student F (3) – (4)

La statistique de Student Le but est de tester ici le caractère significatif des coefficients de la régression. On teste, pour chaque coefficient, $H_0 : \beta_i = 0$ contre l'hypothèse alternative $\beta_i \neq 0$. Pour cela, utilisons le théorème de Pythagore, correspondant de la décomposition de la variance

$$\underbrace{\frac{1}{n} \sum (Y_i - \bar{Y})^2}_{\text{variance totale}} = \underbrace{\frac{1}{n} \sum (Y_i - \hat{Y}_i)^2}_{\text{variance résiduelle}} + \underbrace{\frac{1}{n} \sum (\hat{Y}_i - \bar{Y})^2}_{\text{variance expliquée}}$$

qui peut aussi s'écrire

$$\|\hat{\varepsilon}_t\|^2 = \|Y - X\hat{\beta}\|^2 + \|X\beta - X\hat{\beta}\|^2$$

On peut alors montrer (*théorème de Cochran*) que

$$\frac{\|Y - X\hat{\beta}\|^2}{\sigma^2} \text{ suit un } \chi^2(n-p-1)$$

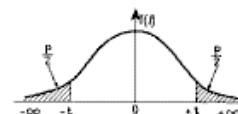
Donc, en remplaçant σ par son estimateur, on peut montrer, toujours sous l'hypothèse de normalité de ε_t que

$$t_{n-p-1} = \frac{(\beta_i - \hat{\beta}_i)}{\sqrt{\sum (Y_i - \hat{Y}_i)^2 [(X'X)_{i,i}^{-1}]}} \sqrt{n-p-1}$$

suit un t de Student à $n-p-1$ degrés de liberté. En effet, sous H_0 , $\hat{\beta}_i$ suit une loi normale centrée, de variance $\sigma\sqrt{[(X'X)_{i,i}^{-1}]}$.

La statistique t_{n-p-1} doit alors être comparée à la valeur donnée dans la table de Student.

Valeurs de T ayant la probabilité P d'être dépassées en valeur absolue



v	$P = 0.90$	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01
1	1.058	0.325	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.823	63.657
2	0.142	0.289	0.445	0.617	0.816	1.064	1.386	1.886	2.920	4.503	6.665	9.925
3	0.137	0.277	0.424	0.584	0.765	0.978	1.250	1.638	2.355	3.182	4.541	5.841
4	0.134	0.271	0.414	0.569	0.741	0.941	1.190	1.593	2.135	2.776	3.747	4.694
5	0.132	0.267	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.131	0.265	0.404	0.555	0.718	0.906	1.134	1.440	1.945	2.447	3.143	3.707
7	0.130	0.263	0.402	0.549	0.711	0.896	1.119	1.415	1.899	2.365	2.998	3.499
8	0.130	0.262	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.806	3.355
9	0.129	0.261	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.263	2.821	3.250
10	0.129	0.260	0.397	0.542	0.700	0.879	1.093	1.372	1.812	2.224	2.764	3.169
11	0.129	0.260	0.396	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.128	0.259	0.395	0.539	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.128	0.259	0.394	0.538	0.694	0.870	1.079	1.350	1.771	2.166	2.650	3.012
14	0.128	0.258	0.393	0.537	0.692	0.868	1.076	1.343	1.761	2.143	2.624	2.977
15	0.128	0.258	0.395	0.536	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.128	0.258	0.392	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.128	0.257	0.392	0.534	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.127	0.257	0.392	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.127	0.257	0.391	0.533	0.686	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.127	0.257	0.391	0.533	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.127	0.257	0.391	0.532	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.127	0.256	0.390	0.532	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.127	0.256	0.390	0.532	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.127	0.256	0.390	0.531	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.127	0.256	0.390	0.531	0.684	0.857	1.058	1.316	1.708	2.060	2.485	2.787
26	0.127	0.256	0.390	0.531	0.684	0.857	1.058	1.315	1.706	2.056	2.479	2.779
27	0.127	0.256	0.389	0.531	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.127	0.256	0.389	0.530	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750
30	0.12566	0.25315	0.38532	0.52440	0.67449	0.84162	1.03643	1.28155	1.64485	1.99966	2.32634	2.57582

Dans le cas où t_{n-p-1} est supérieur à la valeur indiquée dans le tableau, on rejette H_0 : le coefficient est significativement non nul.

Exemple 34 En reprenant l'exemple de l'introduction, la T -stat étant le ratio de la valeur estimée sur l'écart-type estimé, on obtient

$$t^{(1)} = \left| \frac{32.89132}{11.66331} \right| = 2.820068 \text{ et } t^{(2)} = \left| \frac{0.801901}{0.298436} \right| = 2.687012$$

$$t^{(3)} = \left| \frac{-0.381362}{0.156581} \right| = 2.435564 \text{ et } t^{(4)} = \left| \frac{-0.037132}{0.052023} \right| = 0.713768$$

Ces valeurs sont alors à comparer avec le quantile de la loi de Student à $n - p - 1 = 10$ degrés de liberté. D'après la table ci-dessus, $t_{10}^{5\%} = 2.228$: si les 3 premiers coefficients (la constante, X_1 et X_2) sont significatifs, le dernier (X_3) ne semble pas l'être. Il conviendrait alors d'enlever cette variable du modèle, et de refaire une estimation sans cette variable

Puissance d'un test, ou p -value La variable (4) correspond à la p -value (ou puissance) du test. Considérons deux hypothèses, dont une seule est vérifiée, H_0 (*hypothèse à tester*) et H_1 (*hypothèse alternative*). Il y a 4 cas possibles à distinguer, suivant la décision prise (*valider H_0 ou rejeter H_0*) et la 'vérité' (H_0 était, ou n'était pas, vérifiée) :

- (i) valider H_0 et H_0 était vraie : bonne décision
 - (ii) valider H_0 alors que H_0 n'était pas vraie : mauvaise décision
 - (iii) rejeter H_0 (*et valider H_1*), alors que H_0 était vraie : mauvaise décision
 - (iv) rejeter H_0 et H_0 n'était pas vérifiée : bonne décision
- Deux types d'erreurs sont alors possible :
- rejeter à tort H_0 : cas (iii), erreur de première espèce - probabilité α
 - accepter à tort H_0 : cas (ii), erreur de seconde espèce - probabilité β

Définition 14 On appellera puissance du test, ou p -value, la probabilité $1 - \beta$, probabilité de rejeter H_0 en ayant raison.

On peut montrer que α et β varie en sens inverse, aussi, plus $1 - \beta$ sera petit, plus α sera petit : miniser la puissance revient à minimiser le risque de première espèce.

Dans le cas de la régression linéaire, et du test de significativité des coefficients, une très faible p -value signifie que l'on a peu de chance de rejeter H_0 à tort.

Exemple 35 En reprenant l'exemple de l'introduction, les T -stats sont

$$t^{(1)} = 2.820068, t^{(2)} = 2.687021, t^{(3)} = 2.435564 \text{ et } t^{(4)} = 0.713768$$

Ces valeurs sont à comparer avec les quantiles de la loi de Student à $n - p - 1 = 10$ degrés de liberté :

ν	0.60	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01
10	0.542	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169

Pour le dernier coefficient $t = 0.713768$: ceci correspond à une probabilité comprise entre 0.4 ($t = 0.879$) et 0.5 ($t = 0.7$). t étant proche de 0.7, on peut effectivement s'attendre à avoir une p -value proche de 0.5. Pour le premier coefficient (la constante), $t = 2.82$: ceci correspond à une probabilité comprise entre 0.01 ($t = 3.169$) et 0.02 ($t = 2.764$). t étant proche de 0.276, on peut effectivement s'attendre à avoir une p -value proche de 0.02.

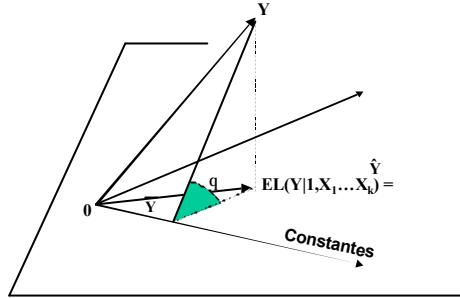
Remarque 24 La p -value peut servir pour savoir si l'on accepte ou si l'on rejette l'hypothèse H_0 : une faible valeur de p pousse à rejeter H_0 . Pour savoir si "c'est bien" ou pas d'avoir une faible valeur, il est important de rappeler l'hypothèse H_0 :

- pour le test de Student (T -stat), H_0 correspond à la nullité du coefficient : si p est faible, le coefficient peut être jugé comme significatif
- pour le test de Fisher (F -stat), H_0 correspond à la nullité de tous les coefficients : si p est faible, le modèle peut être jugé comme significatif
- pour le test de Box-Pierce, H_0 correspond à l'hypothèse de bruit blanc : si p est faible, on rejette l'hypothèse de bruit blanc des résidus

6.2.3 Coefficient R^2 (5) – (6)

Coefficient R^2 Le coefficient R est le coefficient de corrélation entre la série Y_1, \dots, Y_n et la série estimée $\hat{Y}_1, \dots, \hat{Y}_n$. Le carré de ce coefficient s'interprète alors en termes de variance expliquée :

$$R^2 = \frac{\text{Variance expliquée par la régression}}{\text{Variance totale de } Y} = \frac{\sum (Y_i - \bar{Y})^2 - \sum (Y_i - \hat{Y}_i)^2}{\sum (Y - \bar{Y})^2}$$



Le coefficient R^2 est le cosinus de l'angle formé par $Y - \bar{Y}$ et $\hat{Y} - \bar{Y}$ dans \mathbb{R}^n , où \bar{Y} est le vecteur $\bar{Y} \times \mathbb{I}$.

Remarque 25 Il est toujours possible d'augmenter 'artificiellement' un R^2 : considérons $p+1$ variables X_t^i , alors on a toujours

$$R^2(X_t^1, \dots, X_t^p, X_t^{p+1}) \geq R^2(X_t^1, \dots, X_t^p)$$

Pour augmenter un R^2 , il suffit de rajouter une variable explicative. Si l'on se fixe uniquement un critère de maximisation de R^2 dans le cadre d'une modélisation, le 'meilleur' modèle prendra en compte toutes les variables possibles.

Le coefficient R^2 peut être calculé à l'aide de la relation suivante

$$R^2 = 1 - \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

Il faut toutefois prendre des précautions lors de l'utilisation du R^2 : un coefficient ne mesure que de la dépendance linéaire.

Exemple 36 Considérons $X_i = i$ et $Y_i = i^2$ pour $i = 1, \dots, n$. Pour $n = 30$, on obtient le modèle

$$Y_i = 5.92 + 0.03 X_i \quad \text{avec un } R^2 = 0.94$$

Ce modèle peut paraître très significatif ($R^2 = 94\%$) mais ne peut pas être utilisé pour autant : Durbin Watson : $DW = 0.057$.

Coefficient R^2 ajusté - \bar{R}^2 Le R^2 ajusté, noté \bar{R}^2 , permet de prendre en compte le nombre de variables utilisées dans le modèle,

$$\bar{R}^2 = \frac{(n-1) R^2 - p}{n-p} = 1 - \frac{n-1}{n-p} \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

où n est le nombre d'observations, et p le nombre de variables explicatives (y compris la constante). Ce coefficient ajusté est plus intéressant que le R^2 puisqu'il prend en compte le nombre de variables explicatives.

Exemple 37 En reprenant l'exemple de l'introduction, en notant que $\sum_{i=1}^n \hat{\varepsilon}_i = 67.45$ et que $\sum_{i=1}^n (Y_i - \bar{Y})^2 = 226.86$, on alors

$$R^2 = 1 - \frac{67.45}{226.86} = 0.702$$

et le R^2 ajusté est

$$\bar{R}^2 = 1 - \frac{n-1}{n-p-1} (1 - R^2) = 1 - \frac{14-1}{14-4} (1 - 0.702) = 0.613$$

6.2.4 Somme des carrés (7) – (8)

La variable (8) est la somme des carrés des résidus,

$$SQR = SCR = s^2 = \sum_{i=1}^n \hat{\varepsilon}_i^2$$

La variable (7) est l'écart-type 'de la régression', c'est à dire σ , écart-type des erreurs ε_i . Cette valeur est estimé par

$$SER = \sqrt{s^2} \text{ où } s^2 = \frac{1}{14 - 3 - 1} \sum_{i=1}^{14} \hat{\varepsilon}_i^2 = \frac{67.45}{10} = 6.745$$

Exemple 38 En reprenant l'exemple de l'introduction, on a $SCR = \sum_{i=1}^n \hat{\varepsilon}_i = 67.45$ et donc

$$SER = s \text{ où } s^2 = \frac{1}{14 - 3 - 1} \sum_{i=1}^{14} \hat{\varepsilon}_i^2 = \frac{67.45}{10} = 6.745$$

donc $SER = \sqrt{6.745} = 2.597$.

6.2.5 Log-vraisemblance du modèle (9)

La valeur donnée en (9) correspond à la log-vraisemblance du modèle. On suppose là encore que les erreurs ε_i sont normalement distribuées, centrées, de variance σ^2 . L'idépendance des ε_i implique que $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$ suit une vecteur gaussien, $\mathcal{N}(0, \sigma^2 \mathbb{I}_{n \times n})$. Aussi, Y suit un vectuer gaussien $Y \sim \mathcal{N}(X\beta, \sigma^2 \mathbb{I}_{n \times n})$. Aussi, conditionnellement aux vecteurs $X_j, j = 1, \dots, p$, aux vecteurs β des coefficients de la régression, et à σ^2 , la distribution des Y_i est normale :

$$f(Y_i | X_{1,i}, \dots, X_{p,i}, \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(Y_i - X_i'\beta)^2}{2\sigma^2} \right]$$

en notant $X_i = (X_{1,i}, \dots, X_{p,i})'$. Les observations étant supposées indépendantes, la distribution jointe des Y_i est

$$\begin{aligned} f(Y_1, Y_2, \dots, Y_n) &= f(Y_1) \times f(Y_2) \times \dots \times f(Y_n) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left[-\frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^2} \right] \end{aligned}$$

Aussi, la vraisemblance du modèle est

$$\ell(\beta, \sigma^2 | Y, X) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left[-\frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^2} \right]$$

et la log-vraisemblance s'écrit alors

$$\log \ell(\beta, \sigma^2 | Y, X) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^2}$$

Aussi, la log-vraisemblance empirique du modèle, donnée dans (9), est

$$L = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log s^2 - \frac{(Y - X\beta)'(Y - X\beta)}{2s^2} \text{ où } s^2 = \frac{1}{n - p - 1} \sum_{i=1}^n \hat{\varepsilon}_i^2 = \frac{1}{n - p - 1} (Y - X\beta)'(Y - X\beta)$$

Exemple 39 En reprenant l'exemple de l'introduction, on a $s^2 = 6.745$ et donc

$$L = -\frac{14}{2} \log 2\pi - \frac{14}{2} \log 6.745 - \frac{14 - 3 - 1}{2} = -30.871$$

Remarque 26 Lors d'une comparaison entre modèles, un critère de choix peut être basé sur cette vraisemblance : le 'meilleur' modèle aura une plus grande vraisemblance, ce qui revient à minimiser $-L$ (L étant toujours négatif par construction).

6.2.6 Test du Durbin-Watson (10)

Le test du Durbin-Watson permet de détecter une autocorrélation des résidus d'ordre 1, sous la forme

$$\varepsilon_t = \rho \varepsilon_{t-1} + \eta_t \text{ où } \eta_t \sim N(0, \sigma_\eta^2)$$

et le test d'hypothèse s'écrit

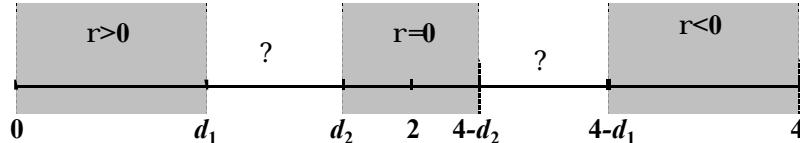
$$H_0 : \rho = 0 \text{ contre } H_1 : \rho \neq 0$$

Pour tester l'hypothèse H_0 , la statistique de Durbin Watson utilisée est

$$DW = \frac{\sum_{t=2}^n [\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1}]^2}{\sum_{t=1}^n \hat{\varepsilon}_t^2}$$

où les $\hat{\varepsilon}_t$ sont les résidus de l'estimation du modèle.

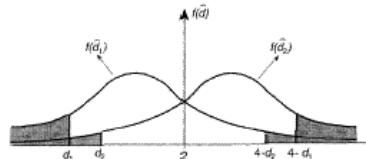
De part sa construction, cette statistique est comprise entre 0 et 4. On peut aussi montrer que $DW = 2$ lorsque $\hat{\rho} = 0$ ($\hat{\rho}$ étant le ρ observé). Afin de tester H_0 , Durbin et Watson ont tabulé les valeurs critiques de DW au seuil de 5% en fonction de la taille de l'échantillon de la taille de l'échantillon n et du nombre de variables explicatives p (k dans le tableau ci-dessous). La lecture de la table permet de déterminer deux valeurs d_1 et d_2 comprises entre 0 et 2 qui peuvent délimiter les régions suivantes :



Selon la position de DW on peut conclure :

- si $d_2 \leq DW \leq 4 - d_2$: on accepte l'hypothèse $H_0 : \rho = 0$
- si $0 \leq DW \leq d_1$: on rejette $H_0 : \rho > 0$
- si $4 - d_1 \leq DW \leq 4$: on rejette $H_0 : \rho < 0$

Dans le cas où $d_1 < DW < d_2$ ou $4 - d_2 < DW < 4 - d_1$, nous sommes dans un cas d'indétermination : il est impossible de conclure dans un sens ou dans un autre. Ces règles de décision sont illustrées ci-dessous,



Les coefficients de Durbin & Watson sont indiqués ci-dessous

n	k = 1		k = 2		k = 3		k = 4		k = 5	
	d ₁	d ₂								
15	1,08	1,36	0,95	1,34	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
31	1,36	1,50	1,30	1,57	1,23	1,65	1,16	1,74	1,09	1,83
32	1,37	1,50	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82
33	1,38	1,51	1,32	1,58	1,26	1,65	1,19	1,73	1,13	1,81
34	1,39	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,15	1,81
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80
36	1,41	1,52	1,35	1,59	1,29	1,65	1,24	1,73	1,18	1,80
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,80
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,21	1,79
39	1,43	1,54	1,38	1,60	1,33	1,66	1,27	1,72	1,22	1,79
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
55	1,53	1,60	1,49	1,64	1,45	1,68	1,41	1,72	1,38	1,77
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,50	1,70	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,52	1,70	1,49	1,74	1,46	1,77
75	1,60	1,65	1,57	1,68	1,54	1,71	1,51	1,74	1,47	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
85	1,62	1,67	1,60	1,70	1,57	1,72	1,55	1,75	1,52	1,77
90	1,63	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78
95	1,64	1,69	1,62	1,71	1,60	1,73	1,58	1,75	1,56	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78

k est le nombre de variables exogènes (constante exclue).
n est la taille de l'échantillon.

Toutefois, 2 règles doivent être vérifiées (*voir construction et démonstration du test : Gourieroux & Monfort*),

- le modèle doit comporter un terme constant (*sinon d'autres tables existent*)

- la variable à expliquer ne doit pas figurer parmi les variables explicatives (*en tant que variable retardée*)

Cette statistique ne permet de tester que la forme AR(1) des résidus. Là aussi, d'autres tests existent pour tester une forme $\varepsilon_t = \rho\varepsilon_{t-h} + \eta_t$ pour $h \neq 1$.

6.2.7 'Dependent var' (11) – (12)

Les valeurs (11) et (12) sont des informations sur la variable Y à expliquer : sa moyenne empirique (11) et son écart-type (12).

6.2.8 Critère d'Aikaïké (13) et de Schwarz (14)

Le critère d'Akaïke - AIC Ce critère est utilisé pour choisir un modèle, et comparer des modèles. Ce critère mesure la distance entre les densités de probabilité 'observée' et estimée. Dans le cas d'une estimation par moindre carrés, ce critère mesurera l'écart entre les résidus et la distribution gaussienne.

$$AIC = \log \left(\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 \right) + \frac{2p}{n}$$

Le critère de Schwarz - BIC ou SC Le critère de Schwarz est un critère légèrement différent de celui introduit par Akaike,

$$SC = \log \left(\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 \right) + \frac{p \log n}{n}$$

De la même façon que pour la vraisemblance, un critère de choix peut être de minimiser SC ou AIC.

6.2.9 Statistique de Fisher F (15) – (16)

Ce test est une généralisation des tests de Student, vus dans la partie (6.2.2) : on va tester ici la nullité de l'ensemble des coefficients de la régression : $H_0 : \beta_1 = \dots = \beta_p = 0$ contre $H_1 : \text{il existe } j \text{ tel que } \beta_j \neq 0$.

Ce test est un cas particulier des tests de relation linéaire entre les coefficients, de la forme $A\beta = \theta$ où A est une matrice de rang q . On pose \widehat{Y} la solution des moindres carrés ordinaires, soit $\widehat{Y} = X(X'X)^{-1}X'Y$, et \widehat{Y}_0 la solution des moindres carrés ordinaires contraints (*contrainte* $A\beta = \theta$).

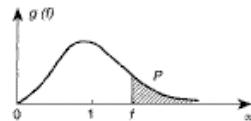
On peut alors montrer que si H_0 est vraie,

$$\frac{n-p}{p-1} \frac{\|Y - \widehat{Y}_0\|^2 - \|Y - \widehat{Y}\|^2}{\|Y - \widehat{Y}\|^2} = F(p-1, n-p)$$

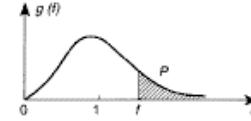
où p est le nombre de variables, incluent la constante. Ce test permet en particulier de faire un test simultané de tous les coefficients de la régression :

Les valeurs F sont alors à comparer à

Valeurs de F ayant la probabilité P d'être dépassées ($F = s_t^2/\bar{s}_t^2$)



Valeurs de F ayant la probabilité P d'être dépassées ($F = s_t^2/\bar{s}_t^2$)



v_2	$v_1 = 1$		$v_1 = 2$		$v_1 = 3$		$v_1 = 4$		$v_1 = 5$	
	$P = 0.05$	$P = 0.01$								
1	161.4	4052	199.5	4999	215.7	5403	224.6	5625	230.2	5764
2	18.51	98.49	19.00	99.00	19.16	99.17	19.25	99.25	19.30	99.30
3	10.13	34.12	9.55	30.81	9.28	29.46	9.12	28.71	9.01	28.24
4	7.71	21.20	6.94	18.03	6.59	16.69	6.39	15.98	6.26	15.53
5	6.61	16.26	5.79	13.27	5.41	12.96	5.19	12.39	5.05	10.97
6	5.99	13.74	5.14	10.91	4.76	9.78	4.53	9.15	4.39	8.75
7	5.59	12.25	4.74	9.55	4.35	8.45	4.12	7.85	3.97	7.45
8	5.32	11.26	4.46	8.65	4.07	7.59	3.84	7.01	3.69	6.63
9	5.12	10.56	4.26	8.02	3.86	6.99	3.63	6.42	3.48	6.06
10	4.96	10.04	4.10	7.56	3.71	6.55	3.48	5.99	3.33	5.64
11	4.84	9.65	3.99	7.28	3.59	6.22	3.36	5.67	3.20	5.32
12	4.75	9.33	3.88	6.93	3.49	5.95	3.26	5.41	3.11	5.06
13	4.67	9.07	3.80	6.70	3.41	5.74	3.18	5.20	3.02	4.86
14	4.60	8.86	3.74	6.51	3.34	5.56	3.11	5.03	2.96	4.69
15	4.54	8.68	3.68	6.36	3.29	5.42	3.06	4.89	2.90	4.56
16	4.49	8.53	3.63	6.23	3.24	5.29	3.01	4.77	2.85	4.44
17	4.45	8.40	3.59	6.11	3.20	5.18	2.96	4.67	2.81	4.34
18	4.41	8.28	3.55	6.01	3.16	5.09	2.93	4.58	2.77	4.25
19	4.38	8.18	3.52	5.93	3.13	5.01	2.90	4.50	2.74	4.17
20	4.35	8.10	3.49	5.85	3.10	4.94	2.87	4.43	2.71	4.10
21	4.32	8.02	3.47	5.77	3.07	4.87	2.84	4.37	2.68	4.04
22	4.30	7.94	3.44	5.72	3.05	4.82	2.82	4.31	2.66	3.99
23	4.28	7.88	3.42	5.66	3.03	4.76	2.80	4.26	2.64	3.94
24	4.26	7.82	3.40	5.61	3.01	4.73	2.78	4.22	2.62	3.90
25	4.24	7.77	3.38	5.57	2.99	4.68	2.76	4.18	2.60	3.86
26	4.22	7.72	3.37	5.53	2.98	4.64	2.74	4.14	2.59	3.82
27	4.21	7.68	3.35	5.49	2.96	4.60	2.73	4.11	2.57	3.78
28	4.20	7.64	3.34	5.45	2.95	4.57	2.71	4.07	2.56	3.75
29	4.18	7.60	3.33	5.42	2.93	4.54	2.70	4.04	2.54	3.73
30	4.17	7.56	3.32	5.39	2.92	4.51	2.69	4.02	2.53	3.70
40	4.08	7.31	3.23	5.18	2.84	4.31	2.61	3.83	2.45	3.51
60	4.00	7.08	3.15	4.98	2.76	4.13	2.52	3.65	2.37	3.34
120	3.92	6.85	3.07	4.79	2.68	3.95	2.45	3.48	2.29	3.17
∞	3.84	6.64	2.99	4.60	2.60	3.78	2.37	3.32	2.21	3.02

v_2	$v_1 = 6$		$v_1 = 8$		$v_1 = 12$		$v_1 = 24$		$v_1 = \infty$	
	$P = 0.05$	$P = 0.01$	$P = 0.05$	$P = 0.01$						
1	234.0	5.899	238.9	5.981	243.9	6.106	249.0	6.234	254.3	6.366
2	19.33	99.53	19.37	99.36	19.41	99.42	19.45	99.46	19.50	99.50
3	8.94	27.91	8.84	27.49	8.74	27.05	8.64	26.60	8.53	26.12
4	6.16	15.21	6.04	14.80	5.91	14.37	5.77	13.93	5.63	13.46
5	4.95	10.67	4.82	10.27	4.68	9.89	4.53	9.47	4.36	9.02
6	4.28	8.47	4.15	8.10	4.00	7.72	3.84	7.31	3.67	6.88
7	3.87	7.19	3.73	6.84	3.57	6.47	3.41	6.07	3.23	5.63
8	3.58	6.37	3.44	6.03	3.28	5.67	3.12	5.28	2.99	4.86
9	3.37	5.80	3.23	5.47	3.07	5.11	2.90	4.73	2.71	4.31
10	3.22	5.39	3.07	5.06	2.91	4.71	2.74	4.33	2.54	3.91
11	3.09	5.07	2.95	4.74	2.79	4.40	2.61	4.02	2.40	3.60
12	3.00	4.82	2.85	4.50	2.69	4.16	2.50	3.78	2.30	3.36
13	2.92	4.62	2.77	4.30	2.60	3.96	2.42	3.59	2.21	3.16
14	2.85	4.46	2.70	4.14	2.53	3.80	2.35	3.43	2.13	3.00
15	2.79	4.32	2.64	4.00	2.48	3.67	2.29	3.29	2.07	2.87
16	2.74	4.20	2.59	3.89	2.42	3.55	2.24	3.18	2.01	2.75
17	2.70	4.10	2.55	3.79	2.38	3.45	2.19	3.08	1.96	2.65
18	2.66	4.01	2.51	3.71	2.34	3.37	2.15	3.00	1.92	2.57
19	2.63	3.94	2.48	3.65	2.31	3.30	2.11	2.92	1.88	2.49
20	2.60	3.87	2.45	3.56	2.28	3.23	2.08	2.86	1.84	2.42
21	2.57	3.81	2.42	3.51	2.28	3.17	2.05	2.80	1.81	2.36
22	2.55	3.76	2.40	3.45	2.23	3.12	2.03	2.75	1.78	2.31
23	2.53	3.71	2.38	3.41	2.20	3.07	2.00	2.70	1.76	2.26
24	2.51	3.67	2.36	3.36	2.18	3.03	1.98	2.66	1.73	2.21
25	2.49	3.63	3.34	3.32	2.16	2.99	1.96	2.62	1.71	2.17
26	2.47	3.59	3.32	3.29	2.15	2.96	1.95	2.58	1.69	2.13
27	2.46	3.56	3.30	3.26	2.13	2.93	1.93	2.55	1.67	2.10
28	2.44	3.53	3.29	3.23	2.12	2.90	1.91	2.52	1.65	2.06
29	2.43	3.50	3.28	3.20	2.10	2.87	1.90	2.49	1.64	2.03
30	2.42	3.47	3.27	3.17	2.09	2.84	1.89	2.47	1.62	2.01
40	2.34	3.29	2.18	2.99	2.00	2.66	1.79	2.29	1.51	1.80
60	2.25	3.12	2.10	2.82	1.92	2.50	1.70	2.12	1.39	1.60
120	2.17	2.96	2.01	2.66	1.83	2.34	1.61	1.95	1.25	1.38
∞	2.09	2.80	1.94	2.51	1.75	2.18	1.52	1.79	1.00	1.00

Exemple 40 En reprenant l'exemple de l'introduction, la statistique de Fisher, permettant de tester la significativité globale de la régression à 3 variables¹⁹ (X_1, X_2, X_3), est donnée par

$$F = \frac{n-p}{p-1} \frac{R^2}{1-R^2} = \frac{10}{3} \frac{0.702}{1-0.702} = 7.878$$

Cette valeur est à comparer avec le F de Fisher à 3 et 10 degré de liberté : $t_{3,10}^{5\%} = 3.71$. F étant plus grande que $t_{3,10}^{5\%}$, nous rejetons l'hypothèse de nullité de tous les coefficient : la régression est globalement significative.

La puissance du test est obtenue de la même façon que pour le test de Student : elle permet de confirmer que la régression est globalement significative.

6.2.10 Statistique sur les résidus

Dans la partie initiale, nous avons montré des sorties statistiques sur les résidus. Parmi les différentes statistiques, en plus des 4 premiers moments (moyenne, variance/dispersion, skewness/asymétrie, kurtosis/épaisseur des queues), un test de normalité est donné. La variance $\hat{\sigma}^2$ n'est pas ici exactement la même que s^2 . En effet,

$$s^2 = \frac{1}{n-p} \sum_{i=1}^n \hat{\varepsilon}_i^2 = \frac{1}{14-3-1} \sum_{i=1}^{14} \hat{\varepsilon}_i^2 = \frac{67.45}{10} = 6.745$$

¹⁹Ce test inclus toujours la constante comme variable explicative.

alors que

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n \hat{\varepsilon}_i^2 = \frac{1}{14-1} \sum_{i=1}^{14} \hat{\varepsilon}_i^2 = \frac{67.45}{13} = 5.188$$

Test de Jarque-Berra Un test possible est celui de Bera et Jarque (1984), basé sur le skewness (*coefficent d'asymétrie de la distribution*) et la kurtosis (*aplatissement - épaisseur des queues*).

En notant μ_k le moment d'ordre k de la distribution, $\mu_k = \mathbb{E}([X - \mathbb{E}(X)]^k)$, on appelle skewness le coefficient $s = \mu_3/\mu_2^{3/2}$ et kurtosis $k = \mu_4/\mu_2^2$. Sous des hypothèses de normalités, on a normalité des estimateurs du skewness et de la kurtosis,

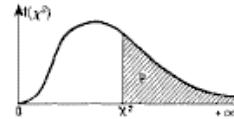
$$s \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sqrt{6/n}\right) \text{ et } k \xrightarrow{\mathcal{L}} \mathcal{N}\left(3, \sqrt{24/n}\right) \text{ quand } n \rightarrow \infty$$

Le test de Bera & Jarque repose sur le fait que, si la distribution suit une loi normale, alors la quantité

$$BJ = \frac{n}{6} s^2 + \frac{n}{24} [k - 3]^2$$

suit asymptotiquement une loi du χ^2 à 2 degrés de liberté. Aussi, si $BJ \geq \chi^2_{1-\alpha}(2)$ on rejette l'hypothèse H_0 de normalité des résidus au seuil α .

3. TABLE DE LA LOI DE CHI-DEUX
Valeurs de χ^2 ayant la probabilité P d'être dépassées



v	$P = 0.90$	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	11,079	13,388	15,086
6	2,200	3,070	3,828	5,348	7,231	8,558	10,645	12,594	15,033	16,812
7	2,833	3,823	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	12,894	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	24,998	28,259	30,578
16	9,312	11,152	12,624	15,338	18,418	20,465	23,542	26,298	29,633	32,000
17	10,086	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	14,844	17,187	19,021	22,337	26,018	28,429	32,007	35,172	38,868	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	22,473	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	40,236	43,773	47,962	50,892

Lorsque $v > 30$ on peut admettre que la quantité $\sqrt{2} \chi^2 - \sqrt{2} v - 1$ suit la loi normale réduite.

Exemple 41 En reprenant une dernière fois l'exemple développé dans l'introduction, le skewness empirique vaut 0.043167 et la kurtosis empirique 2.28694. La statistique de Bera-Jarque est alors

$$BJ = \frac{14}{6} 0.043167^2 + \frac{14}{24} [2.28694 - 3]^2 = 0.0043 + 0.2966 = 0.3009$$

Cette valeur est alors à comparer au quantile de la loi du χ^2 .

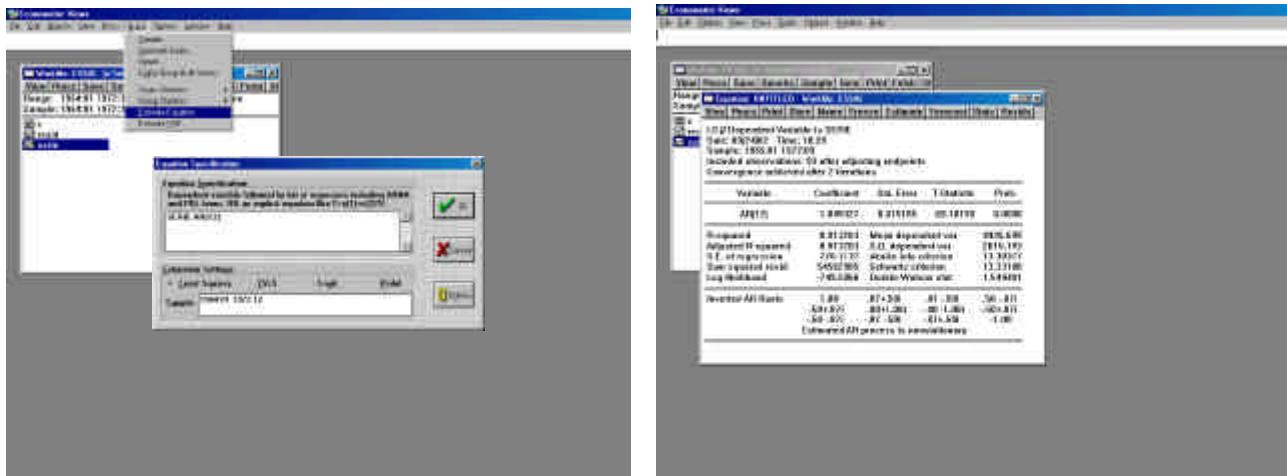
6.2.11 Copier/coller des sorties dans un rapport

Les écrans présentés ci-dessous montrer comment estimer un modèle AR(12) (*sans constante : pour rajouter la constante, il suffit de rajouter c dans l'équation à estimer*). Il suffit de cliquer sur Quick et Estimate Equation. La méthode est toujours la même : on écrit l'équation linéaire économétrique que l'on cherche à estimer ($Z = \alpha Y + \beta X + \gamma + \varepsilon$), avec comme première variable la variable à expliquer (Z), puis la liste de toutes les autres variables (Y , X et la constante) : on tape alors Z Y X c. De la même façon en série temporelle, pour estimer un modèles

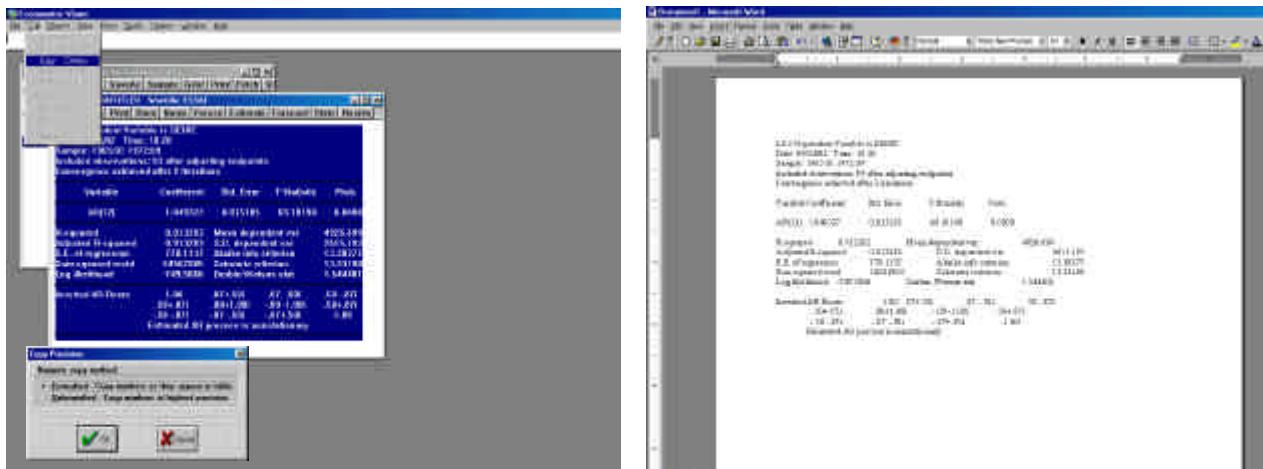
ARMA pour une série (X_t), tape X_t puis les composantes *AR* et *MA* que l'on souhaite prendre en compte : par exemple, l'écriture X AR(3) AR(5) MA(1) MA(3) permettra de tester et d'identifier le modèle

$$X_t = \alpha X_{t-3} + \beta X_{t-5} + \varepsilon_t + \gamma \varepsilon_{t-1} + \delta \varepsilon_{t-3}$$

Dans le cas présenté ci-dessous (**SERIE AR(12)**) on estime un modèle *AR* (12), ne prenant en compte que le 12ème retard, sur la série '**Serie**' (*présentée dans l'introduction sur EViews*)

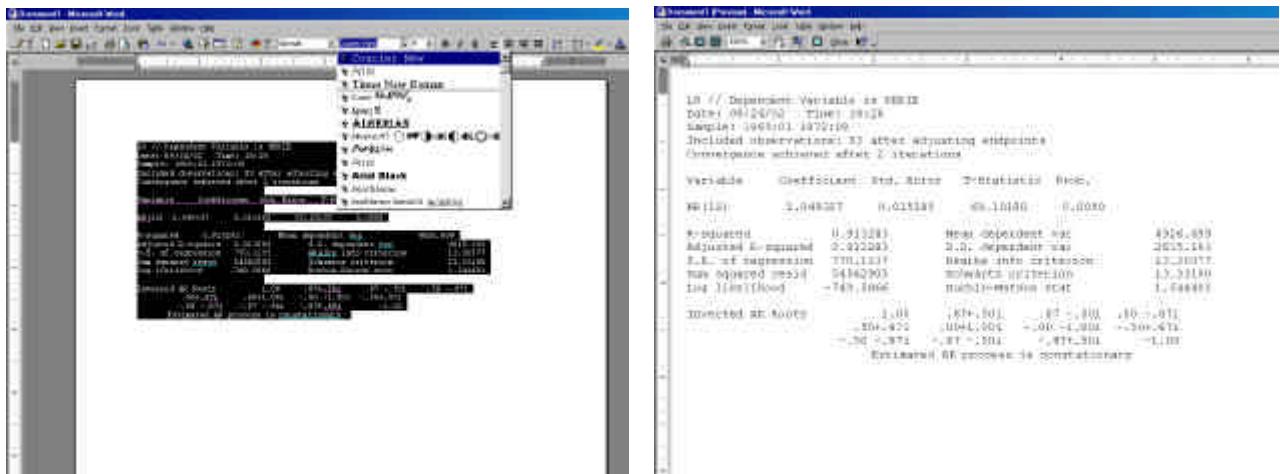


Il est possible, de la même façon que nous avions vu comment coller un graphique (*quel qu'il soit*), de coller des sorties chiffrées (*sortie d'une régression dans l'exemple qui nous intéresse, mais aussi sorties de différents tests proposés sous EViews*). La méthode la plus simple est alors de sélectionner le texte de la sortie (*ci-dessous à gauche*), de le copier en cliquant sur **Edit** et **Copy**, et de sélectionner les options adéquates proposées dans la fenêtre représentée en bas,



L'option la plus utile est ici **Formatted** (*à moins de souhaiter les valeurs des paramètres et des statistiques avec 12 chiffres significatifs*). Une fois copiée la sortie, il suffit là encore de retourner sous le traitement de texte, et d'y copier les informations copiées (*ci-dessus à droite*). Afin d'avoir une sortie sous un format proche de celui obtenu sous EViews, il est alors généralement possible de changer la police de caractère (*en général 'Time New Roman' par défaut*) et de sélectionner une police plus proche de celle utilisée sous Eviews (*'Courier New' par exemple*). La sortie reformattée est présentée ci-dessous. Il est toutefois possible que les espaces ne soient pas ajustés de façon optimale, et il suffit de

quelques suppression d'espaces pour avoir une sortie présentable.



6.2.12 La régression linéaire sur d'autres logiciels

Les parties précédentes ont montré comment lire des sorties d'estimations de modèles linéaires, obtenues sous Eviews. Les sorties obtenues sont relativement générales, et tous les logiciels fournissent les mêmes informations en sortie. Nous allons présenter tout d'abord les sorties sous *S-Plus*, qui peuvent être obtenues, comme sous EVViews, sans ligne de programmation, en utilisant uniquement les menus déroulants. Nous évoquerons ensuite les sorties sous *SAS* et sous *RATS*.

La régression linéaire sous *S-Plus* Les sorties sous *S-Plus* se présentent (*toujours sur l'exemple initial*) sous la forme suivante

```
*** Linear Model ***

Call: lm(formula = Y ~ X1 + X2 + X3, data = data.exemple.reg, na.action
na.omit)
Residuals:
    Min      1Q Median      3Q     Max
-3.697 -1.126  0.1907  1.485  4.488

Coefficients:
            Value Std. Error t value Pr(>|t|)
(Intercept) 32.8913 11.6633   2.8201  0.0182
X1          0.8019  0.2984   2.6870  0.0228
X2         -0.3814  0.1566  -2.4356  0.0351
X3         -0.0371  0.0520  -0.7138  0.4917

Residual standard error: 2.597 on 10 degrees of freedom
Multiple R-Squared:  0.7027
F-statistic: 7.878 on 3 and 10 degrees of freedom, the p-value is
0.005452
```

La variable **Intercept** correspond (*comme sous SAS*) à la constante.

La première partie donne des informations quantitatives sur le résidu de la régression (*moyenne, quartiles, médiane...*). L'estimation est la même que celle sous EVViews, avec l'estimateur des coefficients, son écart-type estimé, la statistique de Student, et sa *p*-value.

Enfin, *S-Plus* donne la '*residual standard error*', le *R*² du modèle, la statistique de Fisher, et sa *p*-value.

La régression linéaire sous *SAS* La régression linéaire se fait à l'aide de la commande suivante

```
PROC REG; MODEL Y = X1 X2 X3;
```

SAS fournit alors la sortie suivante

Dependent Variable: Y					
Analysis of Variance					
Source	DF	Squares	Sum of Square	Mean F Value	Prob>F
Model	3	159.40948	53.13649	7.878	0.0055
Error	10	67.44767	6.74477		
C Total	13	226.85714			
Root MSE		2.59707	R-square	0.7027	
Dep Mean		17.71429	Adj R-sq	0.6135	
C.V.		14.66087			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	32.891324	11.66331015	2.820	0.0182
X1	1	0.801901	0.29843584	2.687	0.0228
X2	1	-0.381362	0.15658069	-2.436	0.0351
X3	1	-0.037132	0.05202312	-0.714	0.4917

Là encore, la sortie est assez proche de celles obtenues à l'aide des autres logiciels. En particulier, le tableau du bas correspond à ceux déjà obtenus (*estimateur des coefficients, son écart-type estimé, la statistique de Student, et sa p-value*). Tout un ensemble d'indicateurs est toutefois donné auparavant, dans la partie 'Analyse de la Variance' :

- la somme des carrés de erreurs (*SCR*) : ligne 'Error', colonne 'Squares' = 67.44767
- la statistique de Fisher : F Value = 7.878, ainsi que la p-value associée (Prob>F) = 0.0055
- l'écart-type de la régression : Root MSE = 2.59707
- la moyenne de la variable Y : Dep Mean = 17.71429
- le 'coefficient de variation' : C.V. = 14.66087. Ce coefficient correspond au rapport Root MSE/Dep Mean exprimé en pourcentage, $CV = 100 \times 2.59707 / 17.1429$
- le R^2 : R-square = 0.7027
- le R^2 ajusté, \bar{R}^2 : Adj R-sq = 0.6135

La régression linéaire sous RATS La régression linéaire se fait à l'aide de la commande suivante

```
linreg Y residus; #constant X1 X2 X3;
```

RATS fournit alors la sortie suivante

Dependent Variable Y - Estimation by Least Squares					
Usable Observations	14	Degrees of Freedom	10		
Total Observations	24	Skipped/Missing	10		
Centered R**2	0.702687	R Bar **2	0.613493		
Uncentered R**2	0.985401	T x R**2	13.796		
Mean of Dependent Variable	17.714285714				
Std Error of Dependent Variable	4.177385480				
Standard Error of Estimate	2.597068848				
Sum of Squared Residuals	67.447666031				
Regression F(3,10)		7.8782			
Significance Level of F	0.00545230				
Durbin-Watson Statistic	3.186886				
Q(6-0)	15.898588				
Significance Level of Q	0.01430876				
Variable	Coeff	Std Error	T-Stat	Signif	
1. Constant	32.89132428	11.66331015	2.82007	0.01815860	
2. X1	0.80190069	0.29843584	2.68701	0.02281643	
3. X2	-0.38136236	0.15658069	-2.43556	0.03511440	
4. X3	-0.03713244	0.05202312	-0.71377	0.49169355	

Là encore, la sortie est assez proche de celles obtenues à l'aide des autres logiciels. En particulier, le tableau du bas correspond à ceux déjà obtenus (*estimateur des coefficients, son écart-type estimé, la statistique de Student, et sa p-value*). Tout un ensemble d'indicateurs est toutefois donné auparavant :

- le R^2 : Centred $R^{**2} = 0.702687$
 - le R^2 ajusté, \bar{R}^2 : R Bar $^{**2} = 0.613493$
 - un R^2 non centré : Uncentred $R^{**2} = 0.985401$
 - la variable $N \times R^2$: T x $R^{**2} = 0.985401 * 14 = 13.769$
 - la moyenne de la variable Y : Mean of Dependent Variable = 17.71429
 - l'écart-type de la variable Y : Std Error of Dependent Variable = 4.177385
 - l'écart-type de la régression : Standard Error of Estimate = 2.59707 (=s - indiqué en (7))
 - la somme des carrés de erreurs (SCR) : Sum of Squared Residuals = 67.447666031
 - la statistique de Fisher : Regression F(3,10) = 7.878, ainsi que la p-value associée (Significance Level of F) = 0.00545230
 - la statistique de Durbin Watson : Durbin Watson Statistic = 3.186886
 - la statistique Q de : $Q(6-0) = 15.898588$, ainsi que la p-value associée (Significance Level of Q) = 0.01430876
- Cette statistique Q correspond à celle de Box-Pierce,

Statistique de Box-Pierce Q , 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 $\text{cov}(\varepsilon_t, \varepsilon_{t-k}) = 0$ pour tout k , soit $\rho(k) = 0$ pour tout k . 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 :

$$Q'_h = T(T+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{T-k}$$

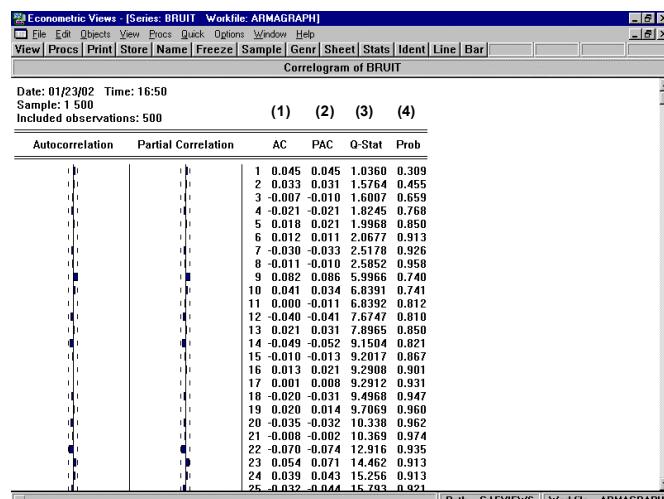
qui suit asymptotiquement, sous H_0 une loi du χ^2 à h degrés de liberté. Ces tests sont appelés par les anglo-saxons 'portmanteau tests', soit littéralement tests 'fourre-tout'.

RATS effectue le test pour $h = 6$.

Exemple 42 Dans l'exemple précédent, $Q_6 = 15.898588$ qui est à comparer avec la quantile de la loi de χ^2 à 6 degrés de liberté : 10.645, ce qui conduit à rejeter H_0 , hypothèse de bruit blanc des résidus. Ce rejet est confirmé par le faible niveau de la p-value.

6.3 Lecture de l'autocorrélogramme sous Eviews

Sous EViews, les autocorrélogrammes se présentent sous la forme suivante



Les courbes à gauche correspondent à l'autocorrélogramme de la série, et à son autocorrélogramme partiel. Les valeurs des séries sont données respectivement en (1) et en (2). L'autocorrélogramme (1) correspond à $\rho(h) = \text{corr}(X_t, X_{t-h})$, estimé à partir de l'échantillon par

$$\widehat{\rho}_T(h) = \frac{\widehat{\gamma}_T(h)}{\widehat{\gamma}_T(0)} \text{ où } \widehat{\gamma}_T(h) = \frac{1}{T-h} \sum_{t=1}^{T-h} (X_t - \bar{X}_T)(X_{t-h} - \bar{X}_T)$$

La fonction d'autocorrélation partielle (2) correspond à $\psi_X(h)$ défini par

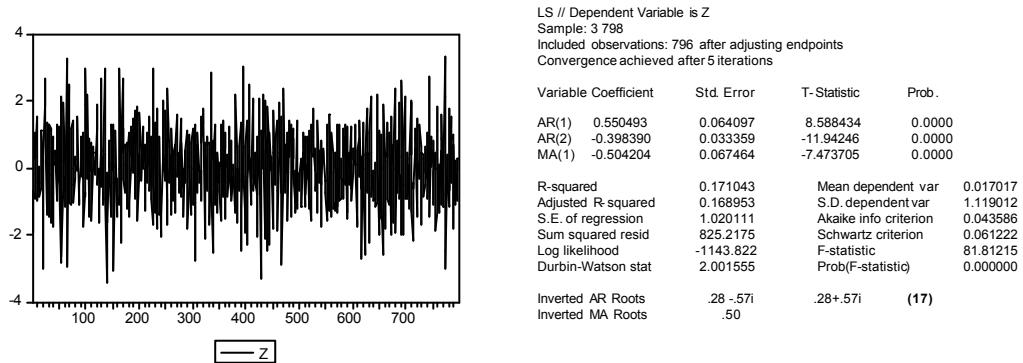
$$\psi_X(h) = \text{corr}(X_{t-h} - EL(X_{t-h}|X_{t-1}, \dots, X_{t-h+1}), X_t - EL(X_t|X_{t-1}, \dots, X_{t-h+1}))$$

Le test (3) est la valeur du test de Ljung-Box, et la valeur (4) étant la puissance de ce test.

Remarque 27 Contrairement aux sorties de régression et aux graphiques, il est assez difficile de copier les données de l'autocorrélogramme, qui ne correspondent ni à un image (comme pouvait l'être un graphique : les autocorrélations ou la Q-stat sont présentés sous format "texte" : il est possible de copier dans un document la valeur exacte de l'autocorrélation), ni à du texte (les graphiques de gauche - autocorrélogrammes - seraient alors copiés sous une forme difficilement exploitable). La méthode la plus simple semble alors être d'utiliser la copie d'écran. Pour cela, il suffit généralement de cliquer simultanément sur les touches 'XXX' et 'Impr écran' sur le clavier, et de coller l'image dans le traitement de texte²⁰.

6.4 Estimation d'un modèle ARMA sous Eviews

Considérons la série Z_t (représentée à gauche), et considérons la modélisation ARMA(2,1), dont l'estimation donne les résultats ci-dessous à droite



Cette sortie (*obtenue sous EVViews*) est très proche de la sortie d'une régression linéaire que nous avions déjà commenté en (6.2). Seule la ligne (17) n'apparaît pas lors d'une simple régression linéaire

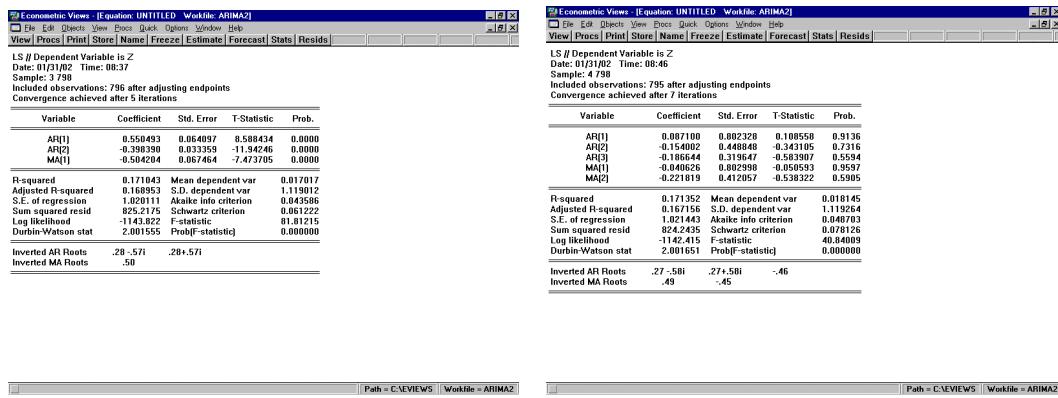
6.4.1 Racines des polynômes AR et MA (17)

L'information supplémentaire, correspondant à la sortie (17) correspond à l'inverse des racines du polynôme ARMA. Dans l'exemple que nous avons considéré ici, les inverses des racines du polynôme AR sont $0.28 \pm 0.57i$, et l'inverse de la racine de la composante MA est 0.50. Deux résultats importants sont à noter :

- les racines sont à l'extérieur du disque unité (*les inverses étant à l'intérieur*), et ne sont pas de module égal à 1
- les racines du polynôme AR et du polynôme MA sont distinctes

La sortie ci-dessous à gauche correspond à l'estimation d'un modèle ARMA(2,1), et à droite, d'un modèle ARMA(3,2)

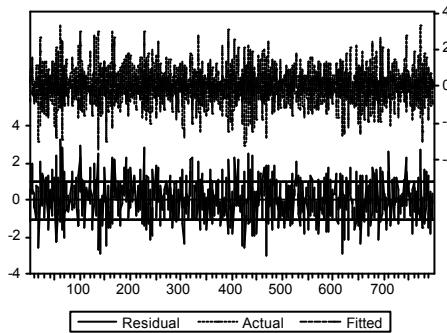
²⁰Toute autre suggestion de méthode sera la bienvenue.



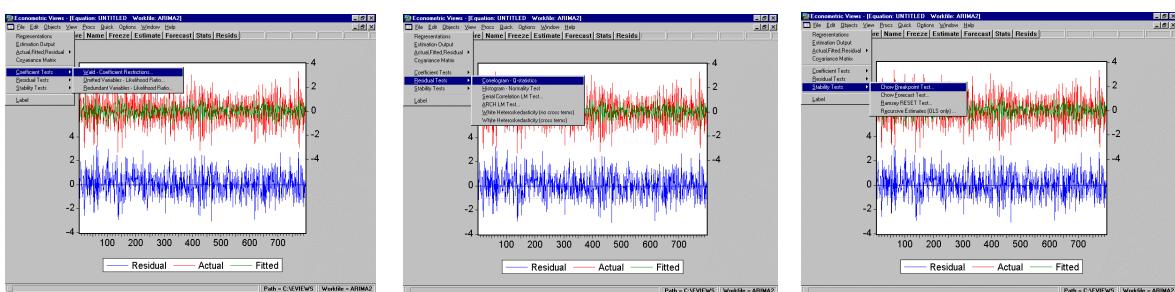
Comme nous pouvons le noter, nous retrouvons des racines très proches de celle obtenu dans le cas $ARMA(2,1)$, mais nous obtenons une racine commune (-0.45) aux deux polynômes : le processus $ARMA$ n'est alors plus sous la forme minimale. Dans ce cas, aucun des coefficient n'apparaît comme significatif, même si les R^2 et la statistique de Fisher sont identiques pour les 2 modèles. En fait, toutes les statistiques sont identiques pour les deux modèles : nous obtenons exactement le même modèle.

6.4.2 Tests sur les erreurs ε_t

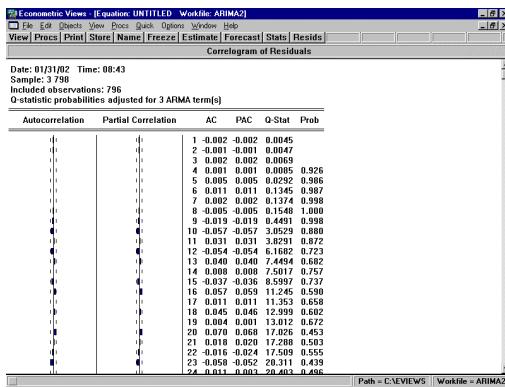
Un grand nombre de tests et d'information peuvent être obtenus sur le processus ε_t ,



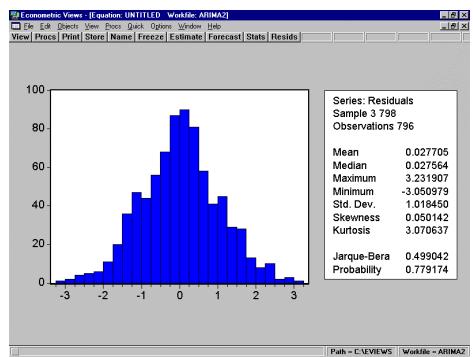
Les écrans ci-dessous montrent l'ensemble des tests qu'il est possible de faire sur les résidus.



Autocorrélogramme des erreurs L'autocorrélogramme permet de tester l'hypothèse de bruit blanc des erreurs, ou plutôt le caractère non-autocorrélé des erreurs. Dans notre exemple, les 3 premières valeurs de la p -value de la statistique Q de Ljung Box n'apparaît pas, correspondant aux 3 termes du modèles $ARMA(2,1)$ (2 coefficients autorégressifs et 1 coefficient moyen mobile). Sur l'exemple ci-dessous, la p -value (toujours très élevée) permet de valider l'hypothèse de bruit-blanc de l'errreur ε_t ,

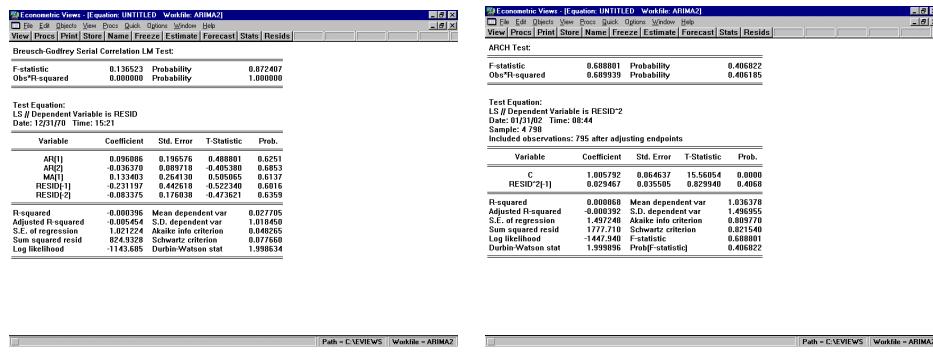


Distribution (marginale) des erreurs



Les statistiques de la distribution des erreurs permettent de tester l'hypothèse de normalité des erreurs ε_t . Encore une fois, si cette hypothèse ne fait pas partie des hypothèses de base des modèles $ARMA$, cette hypothèse est nécessaire pour faire tous les tests de validation du modèle (*Student, Fisher, vraisemblance*).

Tests LM et modélisation $ARCH$ des résidus



Test LM - test de Breusch-Godfrey Ce test est (*de même que le test de Durbin Watson*) un test d'autocorrélation des erreurs ε_t . Toutefois, cette procédure permet de tester l'autocorrélation à des ordres plus élevés que l'ordre 1, comme c'était le cas pour le test de Durbin Watson.

Supposons que nous ayons testé le modèle $\Phi(L)X_t = \Theta(L)\varepsilon_t$, et que l'on souhaite tester une autocorrélation des erreurs de la forme

$$\varepsilon_t = \sum_{i=1}^m \psi_i \varepsilon_{t-i} = \Psi(L) \varepsilon_t$$

Le test du multiplicateur de Lagrange mis en place par Breusch et Godfrey vise à tester l'hypothèse $H_0 : \psi_i = 0$ pour tout i . Pour cela, on considère le modèle (*) $\Phi(L)X_t = \Theta(L)\Psi(L)\varepsilon_t$: sous l'hypothèse H_0 , alors $(n-m)R^2$ suit un chi-deux à m degrés de liberté.

En l'occurrence, *EViews* teste une forme $MA(2)$ pour les erreurs ε_t . Sur l'exemple étudié, on peut noter que la statistique de Fisher du modèle (*) vaut 0.136, avec une p -value de 87%, et le la statistique $(n-m)R^2$ est nulle. Ce test permet de rejeter l'hypothèse où les erreurs ε_t suivent un modèle $MA(2)$.

Modélisation ARCH(1) des erreurs Il est également possible de tester si les erreurs ε_t suivent un modèle ARCH(1).

La sortie obtenue sous EViews présentent la statistique de Fisher et le R^2 du modèle $\Phi(L) X_t = \Theta(L) \varepsilon_t$ où ε_t suit un processus ARCH(1). De plus, EViews donne les résultats de l'estimation des paramètres du modèle ARCH(1) ajusté sur la série ε_t ,

$$\varepsilon_t \sim \eta_t \sqrt{c + \alpha \varepsilon_{t-1}^2} \text{ où } \eta_t \text{ est un bruit blanc}$$

Si la constante est significative sur l'exemple présenté, le paramètre α l'est nettement moins (*statistique de Student valant 0.83, correspondant à une p-value de 40%*) : l'hypothèse ARCH(1) des erreurs est rejetée.

6.5 Les sorties de la procédure ARIMA sous SAS

Les sorties ci-dessous correspondent à l'estimation à l'aide de la procédure ARIMA de SAS d'un modèle de type

$$(1 - L)(1 - L^4)(1 - \alpha L)Z_t = (1 - \beta L - \gamma L^2)\varepsilon_t$$

pour une variable Z . La première sortie correspond à l'autocorrélogramme de la série Z_t comme cela est rappelé par la ligne (1). Les lignes (2) et (3) rappellent la moyenne et l'écart type empiriques que la série Z_t . Enfin (5) donne le nombre d'observations *utilisables* : si l'on part de 200 observations et que l'on considère la série différenciée $Y_t = (1 - L)(1 - L^4)$, on ne pourra travailler que sur 195 observations. La colonne de gauche (4) indique le retard h (*allant ici de 0 à 15*), la colonne (5) donne les autocovariances empiriques $\hat{\gamma}(h)$ et la colonne (6) les autocorrelations empiriques $\hat{\rho}(h) = \hat{\gamma}(h) / \hat{\gamma}(0)$. Les points (7), "..." donnent l'intervalle de confiance de la série des autocorrelations empiriques : ils donnent la limite de la région d'acceptation du test $\rho(h) = 0$. L'écart-type de l'estimateur $\hat{\rho}(h)$ est donné colonne (9). Enfin, les barres (8), "****" représentent les autocorrelations.

ARIMA Procedure																												
Name of variable = Z. (1)																												
Mean of working series = 0.042511 (1) Standard deviation = 1.112798 (2) Number of observations = 75 (3)																												
Autocorrelations																												
(4)	(5)	(6)	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	(9)	Std			
Lag	Covariance	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1		0			
0	1.238320	1.00000																										
1	-0.443324	-0.35800																										0.115470
2	-0.120113	-0.09700																										0.129426
3	-0.219486	-0.17724																										0.130392
4	0.087635	0.07077																										0.133566
5	0.279599	0.22579																										0.134065
6	-0.133568	-0.10786																										0.139042
7	-0.072910	-0.05888																										0.140154
8	-0.137041	-0.11067																										0.140483
9	0.245910	0.19858																										0.141641
10	0.046701	0.03771																										0.145306
11	-0.225334	-0.18197																										0.145436
12	-0.101882	-0.08227																										0.148441
13	0.139227	0.11243																										0.149047
14	0.188290	0.15205																										0.150174
15	-0.0045016	-0.00364																										0.152213

".." marks two standard errors

La sortie suivante donne l'autocorrélogramme inverse (10) et l'autocorrélogramme partiel (11) de la série Z_t ,

ARIMA Procedure																							
Inverse Autocorrelations																							
(10)	Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1
1	0.77158	.																					
2	0.61314	.																					
3	0.50585	.																					
4	0.38195	.																					
5	0.28273	.																					
6	0.26963	.																					
7	0.27145	.																					
8	0.28372	.																					
9	0.23720	.																					

Partial Autocorrelations																							
(11)	Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1
1	-0.35800	.																					
2	-0.25827	.																					
3	-0.38516	.																					
4	-0.29795	.																					
5	0.02150	.																					
6	-0.05608	.																					
7	-0.06272	.																					
8	-0.15030	.																					
9	0.04093	.																					

Le tableau suivant donne le test de Portmanteau (*autocorrelation check for white noise*) permettant de tester l'hypothèse de bruit blanc des résidus. La statistique considérée est

$$Q_K = N(N+2) \sum_{h=1}^K \frac{\hat{\rho}^2(h)}{N-h}$$

qui suit un χ^2 à $k-p-q$ degrés de liberté sous l'hypothèse H_0 de bruit blanc. La colonne (12) donne la valeur de K , nombre de retards considérés, et la colonne (13) donne

La colonne (14) indique le nombre de degrés de liberté associé au test du χ^2 . Enfin, la colonne (15) donne la probabilité que le $\chi^2(K-p-q)$ soit supérieur à Q_K . Enfin, (16) rappelle les autocorrélations empiriques (soit (6)),

ARIMA Procedure							
Autocorrelation Check for White Noise							
(12)	(13)	(14)	(15)	Autocorrelations (16)			
Lag	To	Chi Square	DF	Prob			
6	18.85	6	0.004	-0.358	-0.097	-0.177	0.071
12	27.39	12	0.007	-0.059	-0.111	0.199	0.038
						-0.182	-0.082

ARIMA Estimation Optimization Summary (17)							
Estimation Method:							Unconditional Least Squares
Parameters Estimated:							4
Termination Criteria:							Maximum Relative Change in Estimates
Iteration Stopping Value:							0.001
Criteria Value:							2.01533823
Maximum Absolute Value of Gradient:							31.2375772
R-Square (Relative Change in Regression SSE) from Last Iteration							
Step:							0.99999906
Objective Function:							Sum of Squared Residuals
Objective Function Value:							59.010527
Marquardt's Lambda Coefficient:							0.00001
Numerical Derivative Perturbation Delta:							0.001
Iterations:							4
Warning Message:							Estimates may not have converged.

La colonne (17) [A FINIR]

ARIMA Procedure							
Unconditional Least Squares Estimation							
(19)							
(18)	Parameter	Approx.	(20)	(21)	(22)		
MU	MU	0.04580	0.0077016	5.92	0		
MA1,1	MA1,1	1.06967	0.41311	2.59	1		
MA1,2	MA1,2	-0.06969	0.32291	-0.22	2		
AR1,1	AR1,1	0.43898	0.31266	1.40	1		
Constant Estimate = 0.02558089							
Variance Estimate = 0.83113418							
Std Error Estimate = 0.91166561							
AIC = 206.350923							
SBC = 215.620875							
Number of Residuals= 75							
Correlations of the Estimates (29)							
Parameter	MU	MA1,1	MA1,2	AR1,1			
MU	1.000	0.052	-0.051	0.055			
MA1,1	0.052	1.000	-0.843	0.786			
MA1,2	-0.051	-0.843	1.000	-0.929			
AR1,1	0.055	0.786	-0.929	1.000			

```

ARIMA Procedure

To      Chi           Autocorrelations (30)
Lag     Square DF   Prob
6       8.98  3  0.030 -0.099 -0.069 -0.131  0.109  0.253 -0.054
12      17.54  9  0.041 -0.068 -0.091  0.168  0.017 -0.203 -0.115
18      25.73 15  0.041  0.087  0.144 -0.028 -0.230 -0.049  0.003
24      30.16 21  0.089  0.048 -0.073 -0.147  0.070  0.028 -0.079

Model for variable Z

Estimated Mean = 0.04559679 (31)

Autoregressive Factors (32)
Factor 1: 1 - 0.43898 B**(1)

Moving Average Factors (33)
Factor 1: 1 - 1.0697 B**(1) + 0.069688 B**(2)

```

6.6 Utilisation de SAS avec des séries chronologiques

Nous allons reprendre ici 2 exemples (*sample programs*) de *SAS*, le premier montrant comment mettre en place la méthode de Holt-Winters, et la seconde montrant comment faire de la modélisation à l'aide de processus *ARIMA*,

```

title1 "Intervention Data for Ozone Concentration";
title2 "(Box and Tiao, JASA 1975 P.70)";

data air;
input ozone @@;
label ozone = 'Ozone Concentration';
x1 = 'Intervention for post 1960 period';
summer = 'Summer Months Intervention';
winter = 'Winter Months Intervention';
date = intnx('month', '31dec54'd, _n_);
format date monyy.;
month = month(date);
year = year(date);
x1 = year >= 1960;
summer = (5 < month < 11) * (year > 1965);
winter = (year > 1965) - summer;
cards;
2.7 2.0 3.6 5.0 6.1 5.9 5.0 6.4 7.4 8.2 3.9
4.1 4.5 5.5 3.8 4.8 5.6 6.3 5.9 8.7 5.3 5.7 5.7
3.0 3.4 4.9 4.5 4.0 5.7 6.3 7.1 8.0 5.2 5.0 4.7
3.7 3.1 2.5 4.0 4.1 4.6 4.2 5.1 4.6 4.4 4.0
2.9 2.4 4.7 5.1 4.0 7.5 7.7 6.3 5.3 5.7 4.8 2.7
1.7 2.0 3.4 4.0 4.3 5.0 5.5 5.0 5.4 3.8 2.4 2.0
2.2 2.5 2.6 3.3 2.9 4.3 4.2 4.2 3.9 3.9 2.5 2.2
2.4 1.9 2.1 4.5 3.3 3.4 4.1 5.7 4.8 5.0 2.8 2.9
1.7 3.2 2.7 3.0 3.4 3.8 5.0 4.8 4.9 3.5 2.5 2.4
1.8 2.3 2.5 3.1 3.5 4.5 5.7 5.0 4.8 4.8 2.1 1.4
2.1 2.9 2.7 4.2 3.9 4.1 4.6 5.8 4.4 6.1 3.5 1.9
1.8 1.9 3.7 4.4 3.8 5.6 5.7 5.1 5.6 4.8 2.5 1.5
1.8 2.5 2.1 1.8 3.7 3.7 4.1 5.1 3.7 5.4 3.0 1.8
2.1 2.6 2.2 3.2 3.5 3.5 4.9 4.2 4.7 3.7 3.2 1.8
2.0 1.7 2.8 3.2 4.4 3.4 3.1 5.5 3.8 3.2 2.3 2.2
1.3 2.3 2.7 3.3 3.7 3.0 3.4 4.7 4.6 2.9 1.7 1.3
1.8 2.0 2.2 3.0 2.4 3.5 3.5 3.3 2.7 2.5 1.6 1.2
1.5 2.0 3.1 3.0 3.5 3.4 4.0 3.8 3.1 2.1 1.6 1.3
;
proc arima data=air;
identify var=ozone(12) crosscorr=(x1(12) summer winter
) noprint;

estimate q=(1) (12) input=(x1 summer winter )
noconstant method=ml tprint;

forecast lead=12 id=date interval=month;

run;
quit;

```

```

title1 "Sales of Passenger Cars";
symbol1 i=spline v=plus;
proc gplot data=sashelp.usecon;
plot vehicles * date = 1 /
haxis='1jan80'd to '1jan92'd by year;
where date >= '1jan80'd;
format date year4.;
run;

proc forecast data=sashelp.usecon interval=month
method=winters seasons=month lead=12
out=out outfull outresid outest=est;
id date;
var vehicles;
where date >= '1jan80'd;
run;

proc print data=est;
run;

title2 "The OUT= Data Set";
proc print data=out;
run;

proc gplot data=out;
plot vehicles * date = 1 / vref=0
haxis='1jan80'd to '1jan92'd by year;
where _type_ = 'RESIDUAL';
format date year4.;
run;

title2 "Plot of Residuals";
symbol1 i=needle;
proc gplot data=out;
plot vehicles * date = 1 / vref=0
haxis='1jan80'd to '1jan93'd by year;
where _type_ = 'RESIDUAL';
format date year4.;
run;

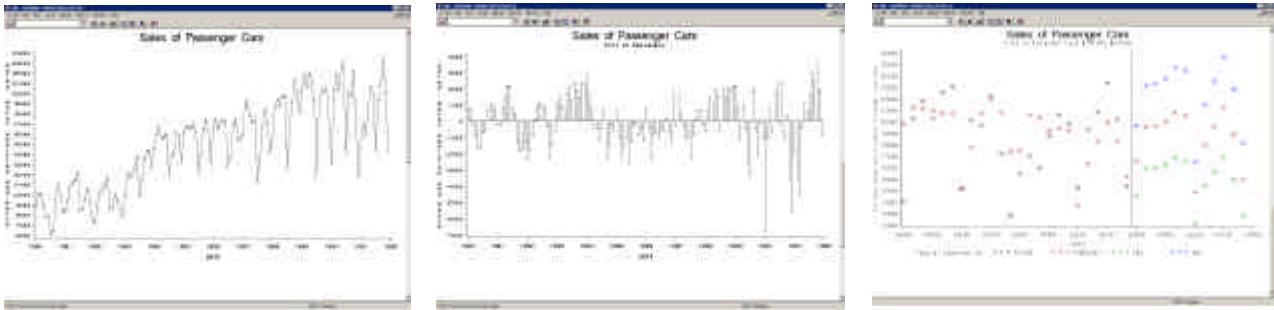
title2 "Plot of Forecast from WINTERS Method";
symbol1 i=none vstar b=2; /* for _type_=ACTUAL */
symbol2 i=spline v=plus b=2; /* for _type_=FORECAST */
symbol3 i=spline l=3; /* for _type_=L95 */
symbol4 i=spline l=3; /* for _type_=U95 */

proc gplot data=out;
plot vehicles * date = _type_ /
href='15dec91'd
haxis='1jan90'd to '1jan93'd by qtr;
where _type_ ^= 'RESIDUAL' & date >= '1jan90'd;
run;

```

6.7 Méthode de lissage exponentiel : Holt-Winters

Le programme ci-dessus donne les sorties suivantes



6.8 Modèles ARIMA sous SAS

6.8.1 Syntaxe de la procédure ARIMA sous SAS

La procédure *ARIMA* contient trois phases successives,

1/ la phase d'*identification* (*IDENTIFY*) pour définir la série à étudier et obtenir des statistiques descriptives (*en particulier les autocorrélogrammes*),

2/ la phase d'*estimation* (*ESTIMATE*) pour effectuer l'estimation des paramètres intervenant une fois la spécification du modèle effectuée,

3/ la phase de *prévision* (*FORECAST*) pour effectuer la prévision de la série considérée à l'aide du modèle estimé dans la phase 2.

La synthaxe générale de cette procédure est de la forme suivante

```
proc ARIMA data=nom_base;
  identify var=N_serie nlag=24;
  estimate p=1 q=1;
  forecast lead=24 interval=month id=date out=resultats;
run;
```

Pour créer une série en différenciant la série autant de fois qu'on le souhaite

<i>identify var=N_serie;</i>	on travaille sur la série X_t
<i>identify var=N_serie(1);</i>	on travaille sur la série $Y_t = L X_t = X_{t-1}$
<i>identify var=N_serie(12);</i>	on travaille sur la série Z_t
<i>identify var=N_serie(1,1);</i>	on travaille sur la série U_t
<i>identify var=N_serie(1,12);</i>	on travaille sur la série V_t

La forme du modèle *ARIMA* à estimer peut être assez complexe. Pour cela, différentes notations sont possibles sous *SAS*

<i>estimate P=1;</i>	on considère le polynôme $AR \Phi(L) = (1 - \phi_1 L)$
<i>estimate P=4;</i>	on considère le polynôme $AR \Phi(L) = (1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \phi_4 L^4)$
<i>estimate P=(1 4);</i>	on considère le polynôme $AR \Phi(L) = (1 - \phi_1 L - \phi_4 L^4)$
<i>estimate P=(1)(4);</i>	on considère le polynôme $AR \Phi(L) = (1 - \phi_1 L)(1 - \phi_4 L^4)$
<i>estimate P=(1)(1 4);</i>	on considère le polynôme $AR \Phi(L) = (1 - \phi L)(1 - \phi_1 L - \phi_4 L^4)$

De plus, un grand nombre d'options sont disponibles pour chacunes des étapes

<i>identify CENTER</i>	la série est centrée
<i>identify NOMISS</i>	les valeurs manquantes sont exclues
<i>identify ALPHA=</i>	permet de spécifier le niveau de significativité des tests
<i>identify ESACF</i>	affiche les résultats de la procédure <i>ESACF</i>
<i>identify SCAN</i>	affiche les résultats de la procédure <i>SCAN</i>
<i>identify P=</i>	détermine la dimension <i>AR</i> dans les méthodes <i>ESACF</i> et <i>SCAN</i>
<i>identify Q=</i>	détermine la dimension <i>MA</i> dans les méthodes <i>ESACF</i> et <i>SCAN</i>
<i>identify NLAG=</i>	permet de spécifier le nombre des retards dans les autocorrélogrammes
<i>identify STATIONARITY</i>	permet d'effectuer des tests de stationnarité sur la série

estimate METHOD= détermine la méthode utilisée pour l'estimation
estimate PLOT affiche le graphique des résidus
estimate AR= permet de spécifier la valeur initiale du modèle *AR*
estimate MA= permet de spécifier la valeur initiale du modèle *MA*
estimate MU= permet de spécifier la constante

Tous les détails sur la procédure *ARIMA* sont disponibles dans la doc *SAS* ou sur internet.

6.8.2 La lecture des sorties *SAS*

Les sorties *SAS* sont les suivantes

ARIMA Procedure							ARIMA Estimation Optimization Summary						
Conditional Least Squares Estimation							Maximum Likelihood						
Iteration	SSE	MA1,1	MA2,1	NUM1	NUM2	NUM3	Lambda	R Crit	Estimation Method:	Maximum Likelihood			
0	154.53372	-0.2924	0.4074	-1.1349	-0.1173	0.05581	0.00001	1	Parameters Estimated:	5			
Iter	Loglike	MA1,1	MA2,1	NUM1	NUM2	NUM3	Lambda	R Crit	Termination Criteria:	Maximum Relative Change in Estimates			
									Iteration Stopping Value:	0.001			
									Criteria Value:	0.00019544			
									Alternate Criteria:	Relative Change in Objective Function			
									Alternate Criteria Value:	1.24707E-8			
									Maximum Absolute Value of Gradient:	0.00712034			
									R-Square (Relative Change in Regression SSE) from Last Iteration				
									Step:	0.00007323			
									Objective Function:	Log Gaussian Likelihood			
									Objective Function Value:	-245.88481			
									Marguardt's Lambda Coefficient:	1E-9			
									Numerical Derivative Perturbation Delta:	0.001			
									Forecasts for variable OZONE				
									Obs	Forecast	Std Error	Lower 95%	Upper 95%
									217	1.4205	0.7966	-0.1407	2.9817
Parameter	Estimate	Std Error	T Ratio	Lag	Variable	Shift			218	1.8446	0.8244	0.2287	3.4604
MA1,1	-0.26684	0.06710	-3.98	1	OZONE	0			219	2.4567	0.8244	0.8408	4.0725
MA2,1	0.76665	0.05973	12.83	12	OZONE	0			220	2.8590	0.8244	1.2431	4.4748
NUM1	-1.33062	0.19236	-6.92	0	X1	0			221	3.1501	0.8244	1.5342	4.7659
NUM2	-0.23936	0.05952	-4.02	0	SUMMER	0			222	2.7211	0.8244	1.1053	4.3370
NUM3	-0.08021	0.04978	-1.61	0	WINTER	0			223	3.3147	0.8244	1.6989	4.9306
									224	3.4787	0.8244	1.8629	5.0946
									225	2.9405	0.8244	1.3247	4.5564
									226	2.3587	0.8244	0.7429	3.9746
									227	1.8588	0.8244	0.2429	3.4746
									228	1.2898	0.8244	-0.3260	2.9057
Maximum Likelihood Estimation													
									Forecasts for variable OZONE				
									Obs	Forecast	Std Error	Lower 95%	Upper 95%
									217	1.4205	0.7966	-0.1407	2.9817
Approx.									218	1.8446	0.8244	0.2287	3.4604
Parameter	Estimate	Std Error	T Ratio	Lag	Variable	Shift			219	2.4567	0.8244	0.8408	4.0725
MA1,1	-0.26684	0.06710	-3.98	1	OZONE	0			220	2.8590	0.8244	1.2431	4.4748
MA2,1	0.76665	0.05973	12.83	12	OZONE	0			221	3.1501	0.8244	1.5342	4.7659
NUM1	-1.33062	0.19236	-6.92	0	X1	0			222	2.7211	0.8244	1.1053	4.3370
NUM2	-0.23936	0.05952	-4.02	0	SUMMER	0			223	3.3147	0.8244	1.6989	4.9306
NUM3	-0.08021	0.04978	-1.61	0	WINTER	0			224	3.4787	0.8244	1.8629	5.0946
Variance Estimate = 0.63450562									225	2.9405	0.8244	1.3247	4.5564
Std Error Estimate = 0.79655861									226	2.3587	0.8244	0.7429	3.9746
AIC = 501.769629									227	1.8588	0.8244	0.2429	3.4746
SBC = 518.360229									228	1.2898	0.8244	-0.3260	2.9057
Number of Residuals=	204												