COURS DE SERIES TEMPORELLES
THEORIE ET APPLICATIONS

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

DESS Actuariat & DESS Mathématiques de la Décision
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 (ECM) ................................. 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 VAR (p) ............................................... 14
    1.3.1 La représentation VAR(p) .......................................... 14
    1.3.2 Les modèles ARMAX ou VARMA .................................... 16
    1.3.3 Estimation des paramètres d’un modèle VAR ........................ 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 VAR ................................ 23
    1.3.7 Tests de bruit blanc des erreurs ................................... 24
    1.3.8 Test de causalité (Granger (1969)) sur des modèles VAR ............ 24
    1.3.9 Les modèles VAR(p) : analyse des chocs ............................ 25
  1.4 Application des modèles VAR ......................................... 28
    1.4.1 Application : investissement, revenu et consommation ................ 28
    1.4.2 Application des modèles VAR : 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 (ADL(p, q)) ....... 35
  1.6 Compléments : modèles multivariés sous SAS .......................... 36
    1.6.1 Estimation d’un modèle VAR ........................................ 36
    1.6.2 Estimation d’un modèle à correction d’erreur ....................... 38
  1.7 Conseils bibliographiques ............................................. 41

2 Les modèles ARCH - Autorégressifs Conditionnellement Hétéroscédastiques 42
  2.1 Notions de stationnarité et notions de linéarité ....................... 43
  2.2 Présentation des modèles ARCH ........................................ 43
    2.2.1 Processus ARCH (1) ................................................ 44
    2.2.2 Processus ARCH (p) ............................................... 45
    2.2.3 Processus GARCH (p, q) ............................................ 45
  2.3 Modèles avec erreurs ARCH .............................................. 46
    2.3.1 Erreurs ARCH (1) .................................................. 46
    2.3.2 Erreurs ARCH (p) .................................................. 48
    2.3.3 Remarque : test de racine unité en présence d’erreurs ARCH ........ 48
  2.4 Estimation et prévision ................................................ 49
    2.4.1 Estimation des paramètres d’un modèle ARCH ....................... 49
    2.4.2 La procédure de Diebold (1987) : test d’autocorrélation en présence d’effet ARCH 49
    2.4.3 Prévision et intervalle de confiance ................................ 50
  2.5 Modèles ARCH 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 BL (p, q, P, Q) ............... 55
    2.6.2 Les modèles autorégressifs exponentiels - EXPAR ................ 55
    2.6.3 Les modèles autorégressifs à seuils - TAR, STAR ou SETAR ....... 56
    2.6.4 Les généralisations des modèles ARCH ............................. 57
  2.7 Les tests de linéarité .................................................. 59
    2.7.1 Les tests du multiplicateur de Lagrange ............................ 59
3 Introduction à la notion de "mémoire longue"

### 3.1 Processus self-similaires
- Accroissements stationnaires de processus self-similaires

### 3.2 Processus FARIMA - ARIMA Fractionnaires
- Processus fractionnaire sans composante AR et MA : processus FARIMA \((0,d,0)\)
- Autocorrélations des processus ARFIMA

### 3.3 Estimation du paramètre \(d\)
- Méthode du log-autocorrélogramme
- Méthode du log-périodogramme
- Méthode de Whittle

### 3.4 Exemples d’applications des processus à mémoire longue
- Applications en finance : rendements d’indices boursier
- Applications en finance : taux de change
- Applications en hydrologie
- Applications en économie

### 3.5 Conseils bibliographiques

4 Compléments : exercices

### 4.1 Exercices avec correction

### 4.2 Examen de 2001/2002

### 4.3 Examen de 2002/2003

### 4.4 Examen de 2003/2004

5 Compléments : simulation de séries temporelles

### 5.1 Simulation de processus en temps discret
- Simulation d’un bruit blanc gaussien
- Simulation d’un processus ARMA
- Simulation d’un processus ARCH

### 5.2 Introduction à la simulation de processus en temps continu

6 Compléments : Logiciels de séries temporelles

### 6.1 Introduction à EViews

### 6.2 La régression linéaire sous EViews
- Estimation des paramètres \((1)\) et \((2)\)
- Statistique de Student \(F\) \((3)\) et \((4)\)
- Coefficient \(R^2\) \((5)\) et \((6)\)
- Somme des carrés \((7)\) et \((8)\)
- Log-vraisemblance du modèle \((9)\)
- Test du Durbin-Watson \((10)\)
- 'Dependent var' \((11)\) et \((12)\)
- Critère d’Aïkâïké \((13)\) et de Schwarz \((14)\)
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2.9 Statistique de Fisher $F$ (15) – (16)</td>
<td>126</td>
</tr>
<tr>
<td>6.2.10 Statistique sur les résidus</td>
<td>127</td>
</tr>
<tr>
<td>6.2.11 Copier/coller des sorties dans un rapport</td>
<td>128</td>
</tr>
<tr>
<td>6.2.12 La régression linéaire sur d’autres logiciels</td>
<td>130</td>
</tr>
<tr>
<td>6.3 Lecture de l’autocorrélogramme sous Eviews</td>
<td>132</td>
</tr>
<tr>
<td>6.4 Estimation d’un modèle $ARMA$ sous Eviews</td>
<td>133</td>
</tr>
<tr>
<td>6.4.1 Racines des polynômes $AR$ et $MA$ (17)</td>
<td>133</td>
</tr>
<tr>
<td>6.4.2 Tests sur les erreurs $\varepsilon_t$</td>
<td>134</td>
</tr>
<tr>
<td>6.5 Les sorties de la procédure $ARIMA$ sous $SAS$</td>
<td>136</td>
</tr>
<tr>
<td>6.6 Utilisation de $SAS$ avec des séries chronologiques</td>
<td>138</td>
</tr>
<tr>
<td>6.7 Méthode de lissage exponentiel : Holt-Winters</td>
<td>139</td>
</tr>
<tr>
<td>6.8 Modèles $ARIMA$ sous $SAS$</td>
<td>139</td>
</tr>
<tr>
<td>6.8.1 Syntaxe de la procédure $ARIMA$ sous $SAS$</td>
<td>139</td>
</tr>
<tr>
<td>6.8.2 La lecture des sorties $SAS$</td>
<td>140</td>
</tr>
</tbody>
</table>
1 Les séries temporelles multivariées


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 dépendence 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-dessous\(^1\) représente l’évolution du spread des obligations d’État pour l’Argentine, le Brésil et le Mexique, entre 1997 et 2000.

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,

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

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

**Remarque 1** Dans le cas gaussien, \( \text{cov}(X,Y) = 0 \) s’il 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

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

Aussi la loi du rapport figurant dans l’information de Kullback est donnée par

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

d’où finalement (après calculs)

\[ \log \left( \frac{f(y)}{f(y|x)} \right) \]往往会用到的代数运算在上面的表达式中出现。

**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 \( x_{t-1} \) et \( y_{t-1} \)), \( \ell \left( x_t, y_t | x_{t-1}, y_{t-1} \right) \). Dans le cas où les processus \((X_t)\) et \((Y_t)\) sont indépendants, alors

\[ \ell \left( x_t, y_t | x_{t-1}, y_{t-1} \right) = \ell \left( x_t | x_{t-1} \right) \times \ell \left( y_t | y_{t-1} \right). \]

On peut montrer que

\[ \frac{\ell \left( x_t, y_t | x_{t-1}, y_{t-1} \right)}{\ell \left( x_t | x_{t-1} \right) \ell \left( y_t | y_{t-1} \right)} = \ell \left( x_t | x_{t-1} \right) \times \ell \left( y_t | y_{t-1} \right) \times \ell \left( x_t | x_{t-1}, y_{t-1} \right) \times \ell \left( y_t | x_{t-1}, y_{t-1} \right). \]

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

\[ 2 \mathbb{E} \left( \log \frac{\ell \left( x_t, y_t | x_{t-1}, y_{t-1} \right)}{\ell \left( x_t | x_{t-1} \right) \ell \left( y_t | y_{t-1} \right)} \right) = 2 \mathbb{E} \left( \log \frac{\ell \left( x_t | x_{t-1}, y_{t-1} \right)}{\ell \left( x_t | x_{t-1} \right)} \right) + 2 \mathbb{E} \left( \log \frac{\ell \left( y_t | x_{t-1}, y_{t-1} \right)}{\ell \left( y_t | y_{t-1} \right)} \right). \]

Mesure de dépendance \( C_{X,Y} \)

Causalité unidirectionnelle \( C_{X \rightarrow Y} \)

Causalité unidirectionnelle \( C_{Y \rightarrow X} \)

Causalité instantanée \( C_{X \rightarrow Y} \)
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 \((Y_t)\), \(\mathbf{X}_t = \{X_1, X_{t-1}, \ldots\}\) et \(\mathbf{Y}_t = \{Y_1, Y_{t-1}, \ldots\}\). Granger a introduit en 1969 différentes notions de causalité :

**Définition 1** \((i)\)  *Y cause X à la date \(t\)* si et seulement si
\[
E(X_t|\mathbf{X}_{t-1}, \mathbf{Y}_{t-1}) \neq E(X_t|\mathbf{X}_{t-1}).
\]

\((ii)\)  *Y cause X instantanément* à la date \(t\) si et seulement si
\[
E(X_t|\mathbf{X}_{t-1}, \mathbf{Y}_t) \neq E(X_t|\mathbf{X}_{t-1}, Y_{t-1}).
\]

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
\[
\begin{align*}
X_t &= \varepsilon_t + \alpha Z_t + \beta Z_{t-1} \\
Y_t &= Z_t,
\end{align*}
\]
doù \((\varepsilon_t)\) et \((Z_t)\) sont des bruits blancs indépendants.

\((i)\)  \(E(X_t|\mathbf{X}_{t-1}, \mathbf{Y}_{t-1}) = \beta Z_{t-1} = \beta Y_{t-1} : \text{Y cause X à la date \(t\)}*

\((ii)\)  \(E(X_t|\mathbf{X}_{t-1}, Y_t) = \alpha Z_t + \beta Z_{t-1} = \alpha Y_t + \beta Y_{t-1} : \text{il y a causalité instantanée de Y vers X si et seulement si} \alpha \neq 0\).

1.2 La notion de cointégration


Une série est intégrée d’ordre \(d\) s’il convient de la différencier \(d\) fois avant de la stationnariser.

**Définition 2** La série \((X_t)\) est dite *intégrée d’ordre \(d\)* \((d \geq 1)\) si \(\Delta^{d-1}X_t\) n’est pas stationnaire et \(\Delta^dX_t\) 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

with 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 cointégrées.

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

ce qui revient à décomposer un processus stationnaire \((\Delta Y_t)\) en une somme de deux processus stationnaires \((\Delta X_t\) et \(Y_{t-1} - \beta X_{t-1})\). De façon plus générale que \((1)\), ces modèles s’écrivent
\[
\Delta Y_t = \mu + \sum_{i=1}^{p} \alpha_i \Delta Y_{t-i} + \sum_{j=0}^{q} b_j \Delta X_{t-j} + c [Y_{t-1} - \beta X_{t-1}] + \eta_t, \tag{2}
\]
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


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 \(C_t = kR_t\), 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. \(c_t - c_{t-1} = c\). On suppose qu’il existe une relation dynamique liant consommation et revenu de la forme suivante,
\[
A(L)c_t = \mu + B(L) + \varepsilon_t \tag{3}
\]
où \(A\) et \(B\) sont ici supposés de degré \(p\) et \(q\) respectivement, et où \((\varepsilon_t)\) est un bruit blanc. Or, d’après la formule de Taylor, il est possible d’écrire
\[
A(L) = A(1) + (1 - L) A^*(L) \quad o\`\`u \quad A^*(L) = \sum_{k=1}^{p} \frac{(-1)^k}{k!} A^{(k)} (1 - L)^{k-1},
\]
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

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

(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 $A(L) = 1 + a_1 L$ et $B(L) = b_0 + b_1 L$, alors (3) s’écrit

$$c_t + a_1 c_{t-1} = \mu + b_0 r_t + b_1 r_{t-1} + \varepsilon_t,$$
d’où finalement

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

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 $\Delta c_t$ 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 $Z_t = (X_t, Y_t)^t$. D’après le théorème de Wold, on peut représenter $(Z_t)$ sous la forme $(1 - L) Z_t = \Theta(L) \varepsilon_t$ où $(\varepsilon_t)$ est un bruit blanc (en dimension 2), $\Theta(0) = I$ 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 $\Phi(L) \Delta Z_t = \varepsilon_t$.

(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.

Comme nous l’avons dit plus haut, le vecteur $(Z_t)$ est une forme de cointégration. La relation de long-terme est donc

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

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

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

Le coefficient $\mu$ 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 Dikey & 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

\[
Y_t = A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + \varepsilon_t \quad \text{ou} \quad \Delta Y_t = \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-2} + \ldots + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t,
\]

où \(\Pi = (A_1 + \ldots + A_p) - I\) et \(\Gamma_i = -(A_{i+1} + \ldots + A_p)\). Le théorème de représentation de Granger affirme alors que le coefficient de la matrice \(\Pi\) a un rang réduit \(r < k\), nombre de variables cointégrées, et donc, il existe des matrices \(U\) et \(V\) de rang \(r\), telles que \(\Pi = UV^0\) et telles que \(V^0 Y_t\) 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 \(\Pi\) et de voir si on peut rejeter des restrictions impliquées par le rang réduit de \(\Pi\) : 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), intégrées\(^2\), le test de de causalité de Granger proposé par EViews donne les résultats suivants, à droite

![Pairwise Granger Causality Tests](image)

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 si \(Z_2\) 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 \(Z_1\).

\(^2\)Ce 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**

<table>
<thead>
<tr>
<th>Lags</th>
<th>Included Observations</th>
<th>LR Test (10%)</th>
<th>LR Test (5%)</th>
<th>LR Test (1%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>0.007596</td>
<td>0.601353</td>
<td>3.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.007596</td>
<td>0.601353</td>
<td>3.76</td>
</tr>
</tbody>
</table>

| Log likelihood | -385.9191 |

<table>
<thead>
<tr>
<th>Hypothesized No. of CE(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>L.R. rejects any cointegration at 5% significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.007596, 0.601353, 3.76, 6.65</td>
</tr>
</tbody>
</table>

### Cas étudiés

- **(i)** pas de trend, pas de constante dans la relation de cointégration $\Pi Y_{t-1} = \alpha \beta Y_{t-1}$
- **(ii)** pas de trend dans la relation de cointégration $\Pi Y_{t-1} = \alpha (\beta Y_{t-1} + \mu)$
- **(iii)** trend linéaire dans la relation de cointégration $\Pi Y_{t-1} = \alpha (\beta Y_{t-1} + \mu + \gamma)$
- **(iv)** trend linéaire dans la relation de cointégration $\Pi Y_{t-1} = \alpha (\beta Y_{t-1} + \mu + \gamma)$
- **(v)** trend quadratique dans la relation de cointégration $\Pi Y_{t-1} = \alpha (\beta Y_{t-1} + \mu + \gamma) + \alpha_3 (\gamma + \theta t)$

### Conclusion

EViews propose plusieurs options pour les tests de cointégration. 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 Ratio - est défini par

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

où $\lambda_i$ est la $i$ème valeur propre (lorsqu’elle 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 cointégration ($r = 1$), la deuxième ligne teste l’hypothèse de deux relations de cointégration ($r = 2$) 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 normalisées 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:**

<table>
<thead>
<tr>
<th>Z1</th>
<th>Z2</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z1</td>
<td>0.000000</td>
<td>-0.887399</td>
</tr>
<tr>
<td>Z2</td>
<td>0.252300</td>
<td>13.86919</td>
</tr>
</tbody>
</table>

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

---

3 Les 5 cas étudiés sont les suivants :

(i) pas de trend, pas de constante dans la relation de cointégration $\Pi Y_{t-1} = \alpha \beta Y_{t-1}$

(ii) pas de trend dans la relation de cointégration $\Pi Y_{t-1} = \alpha (\beta Y_{t-1} + \mu)$

(iii) trend linéaire dans la relation de cointégration $\Pi Y_{t-1} = \alpha (\beta Y_{t-1} + \mu + \gamma)$

(iv) trend linéaire dans la relation de cointégration $\Pi Y_{t-1} = \alpha (\beta Y_{t-1} + \mu + \gamma) + \alpha_3 (\gamma + \theta t)$

(v) trend quadratique dans la relation de cointégration $\Pi Y_{t-1} = \alpha (\beta Y_{t-1} + \mu + \gamma) + \alpha_3 (\gamma + \theta t)$
de cette rupture pour les valeurs positives, avec deux tendances distinctes.

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

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

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

\[ Y_t = A_1 Y_{t-1} + ... + A_p Y_{t-p} + \varepsilon_t, \]

où \((Y_t)\) est un vecteur de processus \(I(1)\). On réécrit alors ce modèle sous la forme décrite ci-dessus, à savoir

\[ \Delta Y_t = \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \Gamma_2 \Delta Y_{t-2} + ... + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t, \]
où \( \Pi = (A_1 + \ldots + A_p) - I \) et \( \Gamma = -(A_{i+1} + \ldots + A_p) \). D'après le théorème de représentation de Granger, il existe \( \alpha \) et \( \beta \) 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 \( \beta \) 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 \( Y_t = A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + BX_t + \varepsilon_t \).

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

![Graphique PIB et Consommation](image)

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

<table>
<thead>
<tr>
<th>Eigenvalue</th>
<th>Likelihood</th>
<th>5 Percent</th>
<th>1 Percent</th>
<th>Hypothesized</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>None **</td>
</tr>
</tbody>
</table>

Unnormalized Cointegrating Coefficients:

- PIB CONSUMMATION
  - 1.000000 -1.492961

Normalized Cointegrating Coefficients: 1 Cointegrating Equation(s)

- PIB CONSUMMATION
  - 0.000000 0.100000
  - 0.000000 0.000000

Log likelihood -434.1513
Il est toutefois possible, sur la version 4.1, d’obtenir des sorties plus complètes sur ce test.

<table>
<thead>
<tr>
<th>Hypothesized No. of CE(s)</th>
<th>Eigenvalue</th>
<th>Trace Statistic</th>
<th>5 Percent Critical Value</th>
<th>1 Percent Critical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0.196115</td>
<td>13.52735</td>
<td>18.17</td>
<td>23.46</td>
</tr>
<tr>
<td>At most 1 *</td>
<td>0.120636</td>
<td>5.013688</td>
<td>3.74</td>
<td>6.40</td>
</tr>
</tbody>
</table>

**(*)** denotes rejection of the hypothesis at the 5%(1%) level

Trace test indicates no cointegration at both 5% and 1% levels

**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: CONSOMMATION PIB
Lags interval: 1 to 1

<table>
<thead>
<tr>
<th>Data Trend: None</th>
<th>Linear</th>
<th>Quadratic</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Trend</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Intercept</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Trend</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td>Trend</td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Selected (5% level) Number of Co integrating Relations by Model (columns)

<table>
<thead>
<tr>
<th>Trace</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max-Eig</td>
<td>2</td>
</tr>
</tbody>
</table>

Log Likelihood by Rank (rows) and Model (columns)

<table>
<thead>
<tr>
<th>No Trend</th>
<th>Intercept</th>
<th>Trend</th>
<th>Intercept</th>
<th>Intercept</th>
<th>Intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-407.9952</td>
<td>-407.9952</td>
<td>-395.3806</td>
<td>-395.3806</td>
<td>-395.1417</td>
</tr>
<tr>
<td>1</td>
<td>-401.9496</td>
<td>-394.8963</td>
<td>-392.3676</td>
<td>-391.0583</td>
<td>-390.8848</td>
</tr>
<tr>
<td>2</td>
<td>-399.8374</td>
<td>-392.0803</td>
<td>-392.0803</td>
<td>-388.3780</td>
<td>-388.3780</td>
</tr>
</tbody>
</table>

Akaike Information Criteria by Rank (rows) and Model (columns)

<table>
<thead>
<tr>
<th>No Trend</th>
<th>Intercept</th>
<th>Trend</th>
<th>Intercept</th>
<th>Intercept</th>
<th>Intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>20.58362*</td>
<td>20.58362*</td>
<td>20.58362*</td>
<td>20.58362*</td>
<td>20.58362*</td>
</tr>
<tr>
<td>1</td>
<td>20.67393</td>
<td>20.67393</td>
<td>20.67393</td>
<td>20.67393</td>
<td>20.67393</td>
</tr>
<tr>
<td>2</td>
<td>20.73733</td>
<td>20.73733</td>
<td>20.73733</td>
<td>20.73733</td>
<td>20.73733</td>
</tr>
</tbody>
</table>

Schwarz Criteria by Rank (rows) and Model (columns)

<table>
<thead>
<tr>
<th>No Trend</th>
<th>Intercept</th>
<th>Trend</th>
<th>Intercept</th>
<th>Intercept</th>
<th>Intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>20.83955*</td>
<td>20.83955*</td>
<td>20.83955*</td>
<td>20.83955*</td>
<td>20.83955*</td>
</tr>
<tr>
<td>1</td>
<td>20.17263</td>
<td>20.17263</td>
<td>20.17263</td>
<td>20.17263</td>
<td>20.17263</td>
</tr>
</tbody>
</table>

**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,

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

avec

\[
Y_t = \begin{bmatrix}
Y_t^1 \\
Y_t^2 \\
\vdots \\
Y_t^k \\
\end{bmatrix}, \quad A_i = \begin{bmatrix}
a_{i1}^1 & a_{i2}^1 & \ldots & a_{ik}^1 \\
a_{i1}^2 & a_{i2}^2 & \ldots & a_{ik}^2 \\
\vdots & \vdots & \ddots & \vdots \\
a_{i1}^k & a_{i2}^k & \ldots & a_{ik}^k \\
\end{bmatrix}, \quad A_0 = \begin{bmatrix}
a_{11}^1 \\
a_{12}^1 \\
\vdots \\
a_{1k}^1 \\
\end{bmatrix}, \quad \varepsilon_t = \begin{bmatrix}
\varepsilon_t^1 \\
\varepsilon_t^2 \\
\vdots \\
\varepsilon_t^k \\
\end{bmatrix}.
\]

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 \((k \times k)\), \(A(z) = 1 - A_1 z - A_2 z^2 - ... - A_p z^p\), et on apparaît polynôme caractéristique le polynôme déterminant \((\Pi - A_1 z - A_2 z^2 - ... - A_p z^p)\).

**Remarque 5** L'écriture \(\text{VAR}\) présentée ici suppose que la ième équation, définissant \(Y_i^t\) ne fasse intervenir aucun \(Y_j^t\). Ceci n'est pas forcément toujours le cas : considérons les deux processus stationnaires \((X_t)\) et \((Y_t)\) définis par les relations suivantes

\[
\left\{
\begin{array}{l}
X_t = \alpha + \beta_1 X_{t-1} + \ldots + \beta_p X_{t-p} - \delta Y_t + \gamma_1 Y_{t-1} + \ldots + \gamma_q Y_{t-q} + \epsilon_t \\
Y_t = a + b_1 Y_{t-1} + \ldots + b_r Y_{t-r} - d X_t + c_1 X_{t-1} + \ldots + c_s X_{t-s} + \eta_t,
\end{array}
\right.
\]

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

\[
Z_t = \begin{bmatrix} X_t \\ Y_t \end{bmatrix}, \quad u_t = \begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix}, \quad D = \begin{bmatrix} 1 & \delta \\ d & 1 \end{bmatrix}, \quad A = \begin{bmatrix} \alpha & \beta_1 & \gamma_1 & \ldots & \gamma_q \\ a & b_1 & c_1 & \ldots & c_s \end{bmatrix}
\]

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

\[Z_t = \Psi + \Psi_1 Z_{t-1} + \ldots + \Psi_n Z_{t-n} + v_t\] où \(v_t = B^{-1} u_t\).

On constate que les innovations \((u_t)\) sont alors fonction de l'innovation de la forme structurelle, \((\epsilon_t)\) et \((\eta_t)\), et peuvent être corrélés.

\[v_t = \left( \frac{\epsilon_t - \delta \eta_t}{1 - \delta d}, \frac{\eta_t - d \epsilon_t}{1 - \delta d} \right)\]

Si chacune des composantes sont i.i.d. les variances sont alors respectivement \([\sigma^2 + \delta \sigma^2] / [1 - \delta d]^2\) et \([\sigma^2 + \delta \sigma^2] / [1 - \delta d]^2\). 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 \(\text{VAR}(p)\) est stationnaire si le polynôme caractéristique (défini à partir du déterminant \(\det(\Pi - A_1 z - A_2 z^2 - ... - A_p z^p)\)) a ses racines à l'extérieur du cercle unité.

**Preuve.** Hamilton (1994) page 259

**Exemple 7** Le processus bivarié défini par

\[
\begin{bmatrix} Y_1^t \\ Y_2^t \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 0.7 & 0.4 \\ 0.2 & 0.3 \end{bmatrix} \begin{bmatrix} Y_1^{t-1} \\ Y_2^{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_1^t \\ \epsilon_2^t \end{bmatrix},
\]

a pour polynôme caractéristique

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

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

\[
\begin{bmatrix} Y_1^t \\ Y_2^t \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 0.2 & 0.4 \\ 0.7 & 0.3 \end{bmatrix} \begin{bmatrix} Y_1^{t-1} \\ Y_2^{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_1^t \\ \epsilon_2^t \end{bmatrix},
\]

a pour polynôme caractéristique

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

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

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

a pour polynôme caractéristique

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

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 \(Y_t = (Y_t^1, \ldots, Y_t^n)^\top \in \mathbb{R}^n\) satisfaisant la représentation VAR \((p)\) suivante : pour tout \(t\),

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

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

\[
Z_t = \begin{pmatrix}
Y_t - \mu \\
Y_{t-1} - \mu \\
\vdots \\
Y_{t-p+1} - \mu
\end{pmatrix} = A_{(n\times np)} \begin{pmatrix}
Y_t^- \\
Y_{t-1}^- \\
\vdots \\
Y_{t-p+1}^-
\end{pmatrix} \text{ et } \eta_t = \begin{pmatrix}
\varepsilon_t \\
0_n \\
\vdots \\
0_n
\end{pmatrix}.
\]

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

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

\[
= \begin{bmatrix} 1 - 0.2L + 0.4L^2 & -0.7L + 0.6L^2 \\ -0.3L + 0.1L^2 & 1 - 0.4L + 0.8L^2 \end{bmatrix}.
\]

L’espérance \(\mu\) 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

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

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 \((p)\) :

\[
Y_t = A_0 + A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + \varepsilon_t + B_1 \varepsilon_{t-1} + \ldots + B_q \varepsilon_{t-q},
\]
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 mutivariées.
- 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

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

et notons

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

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

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

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

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

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 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 moco, 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 des modèles peuvent toujours être estimés par des moco 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 $\gamma_{ij}(h) = E\left[ (Y_i^t - \bar{Y}) (Y_j^{t-h} - \bar{Y}) \right]$.

**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

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

On peut alors noter que $\mathcal{X}(h) = \mathcal{X}(-h)'$, 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 $Y_2$. 

![Graphique des séries temporelles Y1 et Y2](image)
Les observations sont les suivantes,

| 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 | -1.937 | -0.704 |
| 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 - 04 | -2.626 | -0.845 | 1995 - 12 | -3.935 | -3.727 | 1998 - 08 | -1.471 | 0.434 |
| 1993 - 05 | -1.924 | -1.356 | 1996 - 01 | -5.133 | -3.934 | 1998 - 09 | -0.363 | -0.776 |
| 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 - 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 \( \rho_{Y_2 \rightarrow Y_1}(h) = \text{corr}(Y_1^t, Y_2^{t-h}) \) et \( \rho_{Y_1 \rightarrow Y_2}(h) = \text{corr}(Y_2^t, Y_1^{t-h}) \).

<table>
<thead>
<tr>
<th>( Y_1 \rightarrow Y_2 )</th>
<th>( Y_2 \rightarrow Y_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.9088</td>
</tr>
<tr>
<td>1</td>
<td>0.8473</td>
</tr>
<tr>
<td>2</td>
<td>0.7493</td>
</tr>
<tr>
<td>3</td>
<td>0.8291</td>
</tr>
<tr>
<td>4</td>
<td>0.8354</td>
</tr>
<tr>
<td>5</td>
<td>0.8441</td>
</tr>
<tr>
<td>6</td>
<td>0.8551</td>
</tr>
<tr>
<td>7</td>
<td>0.8646</td>
</tr>
<tr>
<td>8</td>
<td>0.8746</td>
</tr>
<tr>
<td>9</td>
<td>0.8844</td>
</tr>
<tr>
<td>10</td>
<td>0.8941</td>
</tr>
<tr>
<td>11</td>
<td>0.9038</td>
</tr>
<tr>
<td>12</td>
<td>0.9135</td>
</tr>
<tr>
<td>13</td>
<td>0.9232</td>
</tr>
<tr>
<td>14</td>
<td>0.9329</td>
</tr>
<tr>
<td>15</td>
<td>0.9426</td>
</tr>
<tr>
<td>16</td>
<td>0.9522</td>
</tr>
<tr>
<td>17</td>
<td>0.9618</td>
</tr>
<tr>
<td>18</td>
<td>0.9715</td>
</tr>
<tr>
<td>19</td>
<td>0.9808</td>
</tr>
<tr>
<td>20</td>
<td>0.9901</td>
</tr>
</tbody>
</table>

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

\[
\begin{bmatrix}
Y_1^t \\
Y_2^t
\end{bmatrix}
= \begin{bmatrix}
a_1 & a_{11} & a_{12} \\
a_2 & a_{21} & a_{22}
\end{bmatrix}
\begin{bmatrix}
Y_1^{t-1} \\
Y_2^{t-1}
\end{bmatrix}
+ \begin{bmatrix}
c_1^t \\
c_2^t
\end{bmatrix},
\]

qui peut se récrire

\[
\begin{align*}
Y_1^t &= a_1 + a_{11}Y_1^{t-1} + a_{12}Y_2^{t-1} + c_1^t \\
Y_2^t &= a_2 + a_{21}Y_1^{t-1} + a_{22}Y_2^{t-1} + c_2^t,
\end{align*}
\]

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

Les résidus estimés, \( \hat{c}_1^t \) et \( \hat{c}_2^t \) sont alors estimés (représentés ci-dessus). Leur matrice de variance-covariance est alors

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

20
d'où les expressions des critères $AIC$ et $SC$,

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

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

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,

$$\begin{align*}
Y^1_t &= a_1 + a_{11} Y^1_{t-1} + a_{12} Y^2_{t-1} + \varepsilon^1_t, \\
Y^2_t &= a_2 + a_{21} Y^1_{t-1} + a_{22} Y^2_{t-1} + \varepsilon^2_t,
\end{align*}$$

qui peut se récrire

$$\begin{align*}
Y^1_t &= a_1 + a_1 Y^1_{t-1} + a_{12} Y^2_{t-1} + \varepsilon^1_t, \\
Y^2_t &= a_2 + a_{21} Y^1_{t-1} + a_{22} Y^2_{t-1} + \varepsilon^2_t.
\end{align*}$$

Pour estimer les paramètres de ce modèle, nous pouvons utiliser les mco sur chacune des équations indépendamment:
dont les sorties graphiques donnent respectivement

\[ \text{Comparaison de différents modèles } \text{VAR} (p) \]

<table>
<thead>
<tr>
<th></th>
<th>( \Sigma )</th>
<th>( \text{AIC}^{(1)} )</th>
<th>( \text{AIC}^{(2)} )</th>
<th>( \text{AIC}^{\text{tot}} )</th>
<th>( \text{SC}^{(1)} )</th>
<th>( \text{SC}^{(2)} )</th>
<th>( \text{SC}^{\text{tot}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>\text{VAR (1) avec constante}</td>
<td>0.796051</td>
<td>-0.016462</td>
<td>0.231237</td>
<td>-0.144759</td>
<td>0.063674</td>
<td>0.311373</td>
<td>-0.037911</td>
</tr>
<tr>
<td>\text{VAR (1) sans constante}</td>
<td>0.805407</td>
<td>0.003241</td>
<td>0.248250</td>
<td>-0.216407</td>
<td>0.056665</td>
<td>0.301674</td>
<td>-0.026226</td>
</tr>
<tr>
<td>\text{VAR (2) avec constante}</td>
<td>0.741204</td>
<td>0.018655</td>
<td>0.242400</td>
<td>-0.132813</td>
<td>0.152215</td>
<td>0.375960</td>
<td>0.080883</td>
</tr>
<tr>
<td>\text{VAR (2) sans constante}</td>
<td>0.749653</td>
<td>0.038084</td>
<td>0.254792</td>
<td>-0.121478</td>
<td>0.144932</td>
<td>0.361640</td>
<td>0.092217</td>
</tr>
<tr>
<td>\text{VAR (3) avec constante}</td>
<td>0.059019</td>
<td>0.250520</td>
<td></td>
<td></td>
<td>0.246003</td>
<td>0.437504</td>
<td></td>
</tr>
<tr>
<td>\text{VAR (3) sans constante}</td>
<td>0.074677</td>
<td>0.276794</td>
<td></td>
<td></td>
<td>0.233939</td>
<td>0.437246</td>
<td></td>
</tr>
<tr>
<td>\text{VAR (4) avec constante}</td>
<td>0.092828</td>
<td>0.289736</td>
<td></td>
<td></td>
<td>0.333336</td>
<td>0.530143</td>
<td></td>
</tr>
<tr>
<td>\text{VAR (4) sans constante}</td>
<td>0.111325</td>
<td>0.318514</td>
<td></td>
<td></td>
<td>0.325021</td>
<td>0.532210</td>
<td></td>
</tr>
</tbody>
</table>

La comparaison des critères \( \text{AIC} \) pousse à choisir un modèle \text{VAR (1) sans constante}. Le modèle retenu

\[
\begin{bmatrix}
Y^1_t \\
Y^2_t
\end{bmatrix} = 
\begin{bmatrix}
0.175 & 0.735 \\
0.267 & 0.394
\end{bmatrix} 
\begin{bmatrix}
Y^1_{t-1} \\
Y^2_{t-1}
\end{bmatrix} + 
\begin{bmatrix}
\varepsilon^1_t \\
\varepsilon^2_t
\end{bmatrix},
\text{où } \Sigma = 
\begin{bmatrix}
0.931916 & 0.553652 \\
0.553652 & 1.193174
\end{bmatrix}.
\]
1.3.6 Prévision à l'aide des modèles VAR

Expression de la prévision à horizon $h$ Considérons un modèle VAR (1), $Y_t = A_0 + A_1Y_{t-1} + \varepsilon_t$, estimé à partir de $T$ observations. La prévision à horizon 1 faite en $T$ est alors

$$\hat{Y}_T(1) = \hat{A}_0 + \hat{A}_1Y_T.$$ 

La prévision à horizon 2 est alors

$$\hat{Y}_T(2) = \hat{A}_0 + \hat{A}_1\hat{Y}_T(1) = \hat{A}_0 + \hat{A}_1\left[\hat{A}_0 + \hat{A}_1Y_T\right] = \hat{A}_0 + \hat{A}_1\hat{A}_0 + \hat{A}_1^2Y_T.$$ 

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

$$\hat{Y}_T(h) = \hat{A}_0 + \hat{A}_1\hat{Y}_T(h-1) = \left[I + \hat{A}_1 + \ldots + \hat{A}_1^{h-1}\right]\hat{A}_0 + \hat{A}_1^hY_T.$$ 

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

$$\Sigma_2 = \Sigma_1 + \hat{A}_1\Sigma\hat{A}_1^2 = \Sigma + \hat{A}_1\hat{A}_0\hat{A}_1,$$

et à horizon 3, elle vaut

$$\Sigma_2 = \Sigma_3 + \hat{A}_1\Sigma_1\hat{A}_1^2 = \Sigma + \hat{A}_1\Sigma_1\hat{A}_1^2.$$ 

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

$$\Sigma_h = \Sigma_{h-1} + \hat{A}_1^{h-1}\Sigma\hat{A}_1^{h-1} = \Sigma + \hat{A}_1\Sigma_1\hat{A}_1^2 + \hat{A}_1^2\Sigma\hat{A}_1^2 + \ldots + \hat{A}_1^{h-1}\Sigma\hat{A}_1^{h-1}.$$ 

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 $\Sigma_h$ : l’intervalle de confiance de la prévision au niveau $1 - \alpha/2$ est donnée par

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

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

Application sur l’exemple précédant 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

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

En utilisant cette estimation, on obtient de façon récursive

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

$$\begin{align*}
\hat{Y}_{2000-03}^{1} &= 0.175336.\hat{Y}_{2000-02}^{1} + 0.7349.\hat{Y}_{2000-02}^{2} \\
\hat{Y}_{2000-03}^{2} &= 0.266518.\hat{Y}_{2000-02}^{1} + 0.3943.\hat{Y}_{2000-02}^{2}.
\end{align*}$$ 

La variance de l’erreur de prévision est donnée par $\Sigma_h$, 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 \((\varepsilon_t)\) le processus d’erreur, et \(\Xi(h)\) sa fonction d’autocorrélation. L’hypothèse à tester est alors

\[ H_0 : \Xi(1) = \ldots = \Xi(h) = 0. \]

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

\[ Q(h) = n \sum_{i=1}^{h} \text{trace}(\Xi(i)\Xi(0)^{-1}\Xi(i)\Xi(0)^{-1}). \]

La statistique de Ljung & Box s’écrit

\[ Q'(h) = n^2 \sum_{i=1}^{h} \frac{1}{n-i} \text{trace}(\Xi(i)\Xi(0)^{-1}\Xi(i)\Xi(0)^{-1}). \]

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,

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

où on supposera \(\Phi(0) = 1\), ce qui permettra d’interpréter le vecteur \(u_t\) comme le processus d’innovation.

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

Le test peut se faire sous EVICUS (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 \(\text{ou variables} (X_t)\) 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 \(VAR(p)\) :

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

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

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

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 \(N \times (K^2p + K)\) de rang \(N\) et \(c\) est un vecteur \(\mathbb{R}^N\). Supposons que \(\hat{\beta}\), estimateur de \(\beta\), vérifie une propriété de normalité asymptotique, c’est à dire \(\sqrt{T} (\hat{\beta} - \beta) \overset{L}{\to} \mathcal{N} (0, \Gamma^{-1} \Sigma)\), alors \(\sqrt{T} (R\hat{\beta} - R\beta) \overset{L}{\to} \mathcal{N} (0, R(\Gamma^{-1} \Sigma) R')\) et on a alors

\[ T (R\beta - c)' \left[R (\Gamma^{-1} \Sigma) R'\right]^{-1} (R\beta - c) \overset{L}{\to} \chi^2 (N). \]

En remplaçant les matrices \(\Gamma\) et \(\Sigma\) par leurs estimateurs usuels

\[ \hat{\Gamma} = \frac{1}{T} ZZ' \text{ et } \hat{\Sigma} = \frac{1}{T - Kp - 1} \hat{u}\hat{u}'. \]
on obtient l’estimateur de Wald
\[
\lambda = \left( R\hat{\beta} - c \right)' R \left( (Z'Z)^{-1} \hat{\Sigma} R \right)^{-1} \left( R\hat{\beta} - c \right),
\]
qui suit asymptotiquement un chi-deux à \( N \) degrés de liberté. Toutefois, une alternative également utilisée est de considérer \( \lambda/N \) qui suit asymptotiquement une loi de Fisher à \( N - Kp - 1 \) degrés de liberté.

Dans le cas du test de causalité de Granger le coefficient \( c \) est nul, de taille \( (m \times n) \times p \), 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
\[
\begin{bmatrix}
X_t \\
Y_t
\end{bmatrix} = \begin{bmatrix}
M^X & A^X_1 & A^{XY} \\
M^Y & A^Y_1 & A^{YX}
\end{bmatrix} \begin{bmatrix}
X_{t-1} \\
Y_{t-1}
\end{bmatrix} + \begin{bmatrix}
A^{XY}_2 & A^{YX}_2 \\
A^{X2}_3 & A^{Y2}_3
\end{bmatrix} \begin{bmatrix}
X_{t-2} \\
Y_{t-2}
\end{bmatrix} + \cdots + \begin{bmatrix}
A^{XY}_p & A^{YX}_p \\
A^{Xp}_p & A^{Yp}_p
\end{bmatrix} \begin{bmatrix}
X_{t-p} \\
Y_{t-p}
\end{bmatrix} + \begin{bmatrix}
\varepsilon_t^X \\
\varepsilon_t^Y
\end{bmatrix},
\]
on a
\[
B = \begin{bmatrix}
M^X & A^X_1 & A^{XY} & M^Y & A^Y_1 & A^{YX} \\
M^X & A^X_1 & A^{XY} & M^Y & A^Y_1 & A^{YX} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
M^X & A^X_1 & A^{XY} & M^Y & A^Y_1 & A^{YX} \\
M^X & A^X_1 & A^{XY} & M^Y & A^Y_1 & A^{YX} \\
M^X & A^X_1 & A^{XY} & M^Y & A^Y_1 & A^{YX} \\
M^X & A^X_1 & A^{XY} & M^Y & A^Y_1 & A^{YX} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
M^X & A^X_1 & A^{XY} & M^Y & A^Y_1 & A^{YX}
\end{bmatrix}
\]
et donc
\[
\beta = [M^X_1 \cdots M^X_n M^Y_1 \cdots M^Y_m A^{XY}_{1,1} \cdots A^{XY}_{1,n} \cdots A^{XY}_{p,1} \cdots A^{XY}_{p,m}]' \in \mathbb{R}^{(n+m)^2 p+n+m}.
\]
La matrice \( R \) est alors la matrice remplie de 0 et de 1 permettant de tester la nullité des coefficients \( A^{XY}_1, \ldots, A^{XY}_p \), c’est à dire
\[
R = \begin{bmatrix}
\text{nullité de } A^{XY}_{1,1} & 0 & \cdots & 0 \\
\text{nullité de } A^{XY}_{1,1} & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\text{nullité de } A^{XY}_{p,m} & 0 & \cdots & 0
\end{bmatrix}
\]
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 \( \text{VAR}(p) \) : analyse des chocs

\( (Y_t) \) suit un modèle \( \text{VAR} \) d’ordre \( p \) si et seulement si
\[
Y_t = M + A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + \varepsilon_t,
\]
où \( \varepsilon_t \) est un bruit blanc de dimension \( K \), de matrice de variance-covariance \( \Sigma \).

Propriété 3 Ce processus admet une représentation \( \text{VAR}(1) \) de la forme \( Z_t = \mu + AZ_{t-1} + U_t \) où
\[
Z_t = \begin{bmatrix}
Y_2 \\
Y_{t-1} \\
\vdots \\
Y_{t-p+1}
\end{bmatrix}, \quad \mu = \begin{bmatrix}
M \\
0 \\
\vdots \\
0
\end{bmatrix}, \quad U_t = \begin{bmatrix}
\varepsilon_t \\
0 \\
\vdots \\
0
\end{bmatrix} \quad \text{et} \quad A = \begin{bmatrix}
A_1 & \cdots & A_{p-1} & A_p \\
\ddots & \ddots & \ddots & \ddots \\
\ddots & \ddots & \ddots & \ddots \\
0 & \ddots & \ddots & \ddots
\end{bmatrix}.
\]
Remarque 8 En posant $J = [I_K, 0, ..., 0]$, 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)

$$Y_t = \mu + \sum_{i=0}^{\infty} \Theta_i \varepsilon_{t-i} \quad \text{où} \quad \Theta_i = JA_i J', \quad \mu = (I_K - A_1 - ... - A_p)^{-1} M .$$  \hspace{1cm} (7)

Impact d’une impulsion Nous allons étudier ici l’effet d’une innovation d’une des variables $(Y_t)$ sur le système de $K$ variables. Pour cela, on suppose que $Y_t = \mu$ pour $t < 0$ et qu’en $t = 0$, la ième variable augmente de 1 (soit $\varepsilon_i^t = 0$ pour $j \neq i$ et $\varepsilon_i^0 = 1$). L’idée est alors d’étudier comment 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 $\delta = (1, 0, 0, ..., 0)^t$ 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 $Y_{it}$ suite à une impulsion sur $Y_t$, à la date $t = 0$, en supposant que toutes les autres variables sont constantes pour $t \leq 0$.

En utilisant l’écriture $VMA(\infty)$, on en déduit

$$Y_0 = \sum_{i=0}^{\infty} \Theta_i \varepsilon_{t-i} = \Theta_0 \delta + 0 + 0 + ... = \delta \quad (\text{car} \; \Theta_0 = I),$$

$$Y_1 = \sum_{i=0}^{\infty} \Theta_i \varepsilon_{t-i} = 0 + \Theta_1 \delta + 0 + ... = \Theta_1 \delta,$$

et plus généralement, $Y_t = \Theta_t \delta$ où la matrice $\Theta_t$ est le coefficient du retard d’ordre $h$ de l’écriture $VMA(\infty)$, c’est à dire $\Theta_t = JA_{t} J^t$. 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 cumulées. Pour cela, on considère les matrices définies par $\Psi_t = I + \Theta_1 + \Theta_2 + ... + \Theta_h$, avec la matrice limite $\Psi_\infty = I + \Theta_1 + \Theta_2 + ... + \Theta_h + ... = (I_K - A_1 - ... - A_p)^{-1}$. Un exemple sera traité dans la partie suivante.

Orthogonalisation des innovations et décomposition de la variance Dans l’analyse précédente, 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 $(\varepsilon_t)$ s’écrivait sous la forme $(0, ..., 0, 1, 0, ..., 0)^t$.

Supposons que la matrice de covariance de l’innovation soit de la forme $\mathbb{E}(\varepsilon_t \varepsilon_t^t) = \Sigma \neq I$. 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 :

$$\Sigma = \mathbb{E}(\varepsilon_t \varepsilon_t^t) = ADA' \quad \text{où} \quad A = \begin{bmatrix} 1 & 0 \\ * & 1 \end{bmatrix} \quad \text{et} \quad D = \begin{bmatrix} * & 0 \\ 0 & * \end{bmatrix}.$$

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 $(\eta_t)$ sont alors des combinaisons linéaires des innovations du modèle initial $(\varepsilon_t)$, qui sont indépendantes (ou tout du moins non corrélées) :

$$\mathbb{E}(\eta_t \eta_t^t) = A^{-1} \mathbb{E}(\varepsilon_t \varepsilon_t^t) (A^{-1})^t = A^{-1} \Sigma (A')^{-1} = A^{-1} ADA' (A')^{-1} = D.$$

Un processus bivarié $(Y_t)$ défini par un modèle $VAR(1)$ de la façon suivante

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

26
devient alors
\[
\begin{cases}
X_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 Y_{t-1} + \eta_t^X \\
Y_t = \beta_0 + \beta_1 X_{t-1} + \beta_2 Y_{t-1} + \rho \sigma_2 / \sigma_1 \eta_t^X + \eta_t^Y.
\end{cases}
\]

L'idée est alors de considérer un choc unitaire non pas sur une composante de \( (\varepsilon_t) \) mais sur la composante \( (\eta_t^Y) \) : 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 \( (X_t) \) : c'est cela que l'on appelle **décomposition de la variance**. Considérons le processus \( (X_t) \) admettant la représentation \( VAR(p) \) suivante
\[
A(L) Y_t = Y_t - A_0 Y_{t-1} - ... - A_p Y_{t-p} = M + \varepsilon_t.
\]
On suppose que le bruit blanc \( (\varepsilon_t) \) est de matrice de variance-covariance \( \Sigma \). On suppose que ce processus est stationnaire, de telle sorte qu'il puisse être représenté sous forme la \( VMA(\infty) \) suivante
\[
Y_t = \mu + \varepsilon_t + \Theta_1 \varepsilon_{t-1} + ... + \Theta_h \varepsilon_{T+h-1} + ... = \mu + \Theta(L) \varepsilon_t.
\]

L'erreur de prévision faite en \( T \) à horizon \( h \) s'écrit sous la forme
\[
X_{T+h} - \hat{X}_T(h) = X_{T+h} - EL(X_{T+h}|X_T, X_{T-1}, ..., X_1)
\]
\[
= X_{T+h} - EL(X_{T+h}|\varepsilon_T, \varepsilon_{T-1}, ..., \varepsilon_1) = \varepsilon_{T+h} + \Theta_1 \varepsilon_{T+h-1} + ... + \Theta_{h-1} \varepsilon_{T+1}.
\]

L'espérance vaut alors 0 (les bruits blancs étant, par définition, des processus centrés), et la matrice de variance-covariance est alors donnée par
\[
\mathbb{E} \left( \left( X_{T+h} - \hat{X}_T(h) \right) \left( X_{T+h} - \hat{X}_T(h) \right)' \right) = \sigma_1 \Sigma \Theta_1 + ... + \Theta_{h-1} \Sigma \Theta_{h-1}.'
\]

Cette erreur de prévision est alors exprimée en fonction de la matrice de variance-covariance \( \Sigma \) des résidus (non diagonale). Comme précédemment, considérons la transformation \( \eta_t = A^{-1} \varepsilon_t \) où \( \varepsilon_t = A \eta_t \), où \( \Sigma = ADA' \):
\[
\varepsilon_t = \begin{bmatrix}
\varepsilon_t^1 \\
\vdots \\
\varepsilon_t^d
\end{bmatrix} = [a_1, a_2, ..., a_n] \begin{bmatrix}
\eta_t^1 \\
\vdots \\
\eta_t^d
\end{bmatrix} \quad \text{où} \quad a_i \in \mathbb{R}^d.
\]

Dès lors
\[
\Sigma = \mathbb{E}(\varepsilon_t \varepsilon_t') = a_1 a_1' V(\eta_t^1) + a_2 a_2' V(\eta_t^2) + ... + a_d a_d' V(\eta_t^d).
\]

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

**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


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) différents:

<table>
<thead>
<tr>
<th>Retards</th>
<th>AIC</th>
<th>HQ</th>
<th>SC</th>
<th>FPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-24.50</td>
<td>-24.38</td>
<td>-24.41</td>
<td>2.500</td>
</tr>
<tr>
<td>3</td>
<td>-24.41</td>
<td>-24.07</td>
<td>-23.55</td>
<td>2.748</td>
</tr>
<tr>
<td>4</td>
<td>-24.36</td>
<td>-23.90</td>
<td>-23.21</td>
<td>2.910</td>
</tr>
</tbody>
</table>

Lütkepohl retenait, en utilisant le critère \textit{AIC} d’Akaike et le \textit{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

\[
\begin{bmatrix}
I_t \\
R_t \\
C_t
\end{bmatrix} = 
\begin{bmatrix}
-0.0167 \\
0.0158 \\
0.0129
\end{bmatrix} + 
\begin{bmatrix}
-0.3196 & 0.1460 & 0.9612 \\
0.0439 & -0.1527 & 0.2885 \\
-0.0024 & 0.2248 & -0.2640
\end{bmatrix} 
\begin{bmatrix}
I_{t-1} \\
R_{t-1} \\
C_{t-1}
\end{bmatrix} + 
\begin{bmatrix}
0.0500 & 0.0192 & -0.0010 \\
0.0339 & 0.3549 & -0.0222
\end{bmatrix} + 
\begin{bmatrix}
\varepsilon_t^1 \\
\varepsilon_t^2 \\
\varepsilon_t^3
\end{bmatrix},
\]

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 \textit{VMA}(\infty), permettant de calculer les fonctions d’impulsions, est la suivante

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

les fonctions d’autocovariance du processus sont alors données par

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

e les fonctions d’autocorrélation sont alors

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

\(^4\text{AIC est le critère d’Akaike, donné pour } m \text{ retards, par}
\]

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

\(^\text{où } T \text{ est le nombre d’observation, } m \text{ le nombre de retards, } K \text{ le nombre de variables, et } \Sigma_m \text{ est l’estimateur (du maximum de vraisemblance) de la matrice de variance-covariance du bruit.}
\]

\(^\text{HQ est le critère de Hannan-Quinn, défini par}
\]

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

\(^\text{SC est le critère de Schwarz défini par}
\]

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

\(^\text{FPE correspond au critère "final prediction error", défini par}
\]

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

\(^\text{(FPE est ici donnée à un facteur multiplicatif près (} \times 10^2 \text{)).}
\]

29
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

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

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

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

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

$$\lambda = \frac{1}{4} \left( R\bar{\beta} - c \right)' \left( R \left( Z'Z \right)^{-1} \bar{\Sigma} R' \right)^{-1} \left( R\bar{\beta} - c \right) = 1.59,$$

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 $\Sigma$ 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

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

La prévision optimale se fait alors, de façon récursive,

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

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

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

Les matrices de variance covariance à des ordres plus élevés se calculent à l’aide de la forme $VMA (\infty)$, 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, finalement, les prévisions suivantes :

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

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

$$\hat{\Theta}_1 = \begin{bmatrix}
-0.319 & 0.145 & 0.961 \\
0.043 & -0.152 & 0.288 \\
-0.002 & 0.224 & -0.263
\end{bmatrix} = \hat{A}_1 \quad \text{et} \quad \hat{\Theta}_2 = \begin{bmatrix}
-0.054 & 0.621 & 0.415 \\
0.028 & 0.133 & -0.088 \\
0.045 & 0.260 & 0.109
\end{bmatrix} = \hat{A}_1 \hat{\Theta}_1 + \hat{A}_2.$$
En notant $\theta_0 = I$, et $\psi_i = \theta_0 + \theta_1 + \ldots + \theta_i$, on a alors les estimations suivantes, avec, entre parenthèses, les intervalles de confiance

<table>
<thead>
<tr>
<th>$i$</th>
<th>$\theta_i$</th>
<th>$\psi_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-0.320</td>
<td>0.146</td>
</tr>
<tr>
<td></td>
<td>(0.125)</td>
<td>(0.562)</td>
</tr>
<tr>
<td></td>
<td>0.044</td>
<td>-0.153</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.143)</td>
</tr>
<tr>
<td></td>
<td>-0.002</td>
<td>0.225</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.115)</td>
</tr>
<tr>
<td></td>
<td>0.680</td>
<td>0.146</td>
</tr>
<tr>
<td></td>
<td>(0.125)</td>
<td>(0.562)</td>
</tr>
<tr>
<td></td>
<td>0.044</td>
<td>0.847</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.143)</td>
</tr>
<tr>
<td></td>
<td>-0.002</td>
<td>0.225</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.115)</td>
</tr>
<tr>
<td>2</td>
<td>-0.054</td>
<td>0.262</td>
</tr>
<tr>
<td></td>
<td>(0.129)</td>
<td>(0.546)</td>
</tr>
<tr>
<td></td>
<td>0.020</td>
<td>0.114</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.135)</td>
</tr>
<tr>
<td></td>
<td>0.045</td>
<td>0.261</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.108)</td>
</tr>
<tr>
<td></td>
<td>0.119</td>
<td>0.353</td>
</tr>
<tr>
<td></td>
<td>(0.084)</td>
<td>(0.384)</td>
</tr>
<tr>
<td></td>
<td>-0.009</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.078)</td>
</tr>
<tr>
<td></td>
<td>-0.001</td>
<td>-0.098</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.078)</td>
</tr>
<tr>
<td>3</td>
<td>0.66</td>
<td>0.961</td>
</tr>
<tr>
<td></td>
<td>0.353</td>
<td>0.120</td>
</tr>
<tr>
<td></td>
<td>0.042</td>
<td>0.091</td>
</tr>
<tr>
<td></td>
<td>0.064</td>
<td>0.103</td>
</tr>
<tr>
<td></td>
<td>0.037</td>
<td>0.176</td>
</tr>
<tr>
<td></td>
<td>0.033</td>
<td>0.156</td>
</tr>
</tbody>
</table>

avec les comportements limites suivants :

$$\lim_{i \to \infty} \hat{\theta}_i = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{et} \quad \lim_{i \to \infty} \hat{\psi}_i = \begin{bmatrix} 0.756 & 0.866 & 1.295 \\ 0.076 & 1.076 & 0.344 \\ 0.053 & 0.505 & 0.964 \end{bmatrix}.$$

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)
n’influent 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 $\Phi_1$ n’est plus une matrice $2 \times 2$). 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 $Y$) $Y_1, Y_2, \ldots, Y_n$ ne sont pas observées pour des individus différents (et peuvent alors être supposées "indépendantes") mais sont des observations temporelles $Y_1, Y_2, \ldots, Y_n$. 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

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

Le problème consiste alors à estimer la constante $\alpha$ et les coefficients $\beta_j$. 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 $X_{t-1}, X_{t-2}, \ldots$ etc, et donc les estimations par moindres carrés des coefficients $\beta_j$ devraient être assez imprécises. Sauf si c’est l’impact à long terme qui nous intéresse. Soit $V \left( \hat{\beta} \right)$ la matrice de covariance du vecteur $\hat{\beta}$ estimateur mco du vecteur $\beta = (\beta_0, \beta_1, \ldots, \beta_q)$.

Alors, si $\hat{\gamma}$ désigne la somme des $\hat{\beta}_j$, la variance de $\hat{\gamma}$

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

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, $V \left( \hat{\gamma} \right)$ peut être plus petite que la somme des $V \left( \hat{\beta}_j \right)$ ou même que chaque $V \left( \hat{\beta}_j \right)$. Si c’est le paramètre $\gamma$ qui nous intéresse plutôt que les $\beta_j$ individuels, l’approche la plus simple consiste à estimer une version reparamétrée de (9) par moindres carrés. Cette version s’écrit alors

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

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 $\beta_j$. 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 $\beta_j$ de (9) doivent apparaître dans un polynôme de degré $d$ donné. Par, si le polynôme était du second degré, nous aurions

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

Si $q > 2$, on peut noter qu’il y aura moins de paramètres $\alpha_i$ que $\beta_j$ et, par conséquent, cette relation impose $q - 2$ contraintes sur les $\beta_j$. 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 \( \beta_j \) par \( \alpha_0 + \alpha_1 j + \alpha_2 j^2 \), c'est à dire

\[
Y_t = \alpha + \alpha_0 \sum_{j=0}^{q} X_{t-j} + \alpha_0 \sum_{j=0}^{q} j X_{t-j} + \alpha_0 \sum_{j=0}^{q} j^2 X_{t-j} + \varepsilon_t, \tag{11}
\]

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 \( \beta_j \) sont simplement des contraintes linéaires. La résolution de (10) montre que

\[
-\beta_3 + 3\beta_2 - 3\beta_1 + \beta_0 = 0 \\
-\beta_4 + 3\beta_3 - 3\beta_2 + \beta_1 = 0 \\
-\beta_5 + 3\beta_4 - 3\beta_3 + \beta_2 = 0 \\
\ldots \text{ et ainsi de suite.}
\]

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

\[
R = \begin{bmatrix}
1 & -3 & 3 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 1 & -3 & 3 & -1 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & 1 & -3 & 3 & -1 & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & -3 & 3 & -1
\end{bmatrix}.
\]

Puisque les contraintes sont linéaires, on peut les tester facilement, par exemple à l’aide d’un test de Fisher : le modèle non contraint est (11), le modèle non contraint 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 \( \beta \) (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 sensée 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

\[
Y_t^* = X_t \beta^* + \varepsilon_t = (X_t^1, \ldots, X_t^q) \begin{pmatrix} \beta_1^* \\ \vdots \\ \beta_d^* \end{pmatrix} + \varepsilon_t
\]

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

\[
Y_t = Y_{t-1} - (1 - \delta) Y_{t-1} + (1 - \delta) X_t \beta^* + (1 - \delta) \varepsilon_t + \eta_t
\]

\[
= X_t \beta + \delta Y_{t-1} + \varepsilon_t,
\]

34
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 $\delta$ 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 $(\varepsilon_t)$ 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.


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

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

$$Y_t = \beta Y_{t-1} + u_t$$

$L’estimation mco$ de $\beta$ est donné par

$$\hat{\beta} = \frac{\sum_{t=2}^{n} Y_t Y_{t-1}}{\sum_{t=2}^{n} Y_{t-1}^2} = \beta \frac{\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

$$p \lim_{n \to \infty} \hat{\beta} = \beta + p \lim_{n \to \infty} \frac{\sum_{t=2}^{n} u_t Y_{t-1}}{\sum_{t=2}^{n} Y_{t-1}^2} = \beta.$$

**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

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

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 \((\varepsilon_t)\), \((X_t)\) et \((Y_t)\) convergeraient vers des valeurs de long terme \((\text{stables})\) \(X^*\) 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 \(\lambda\) peut alors être vu 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 \((\text{toutes équivalentes})\) :

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. \(\Delta Y_t = \alpha + (\beta - 1) \left(Y_{t-1} - X_{t-1}\right) + \gamma_0 \Delta X_t + (\gamma_0 + \gamma_1 + \beta_1 - 1) X_{t-1} + \varepsilon_t,\)
5. \(\Delta Y_t = \alpha + (\beta - 1) \left(Y_{t-1} - \lambda X_{t-1}\right) + \gamma_0 \Delta X_t + \varepsilon_t,\)

où \(\Delta\) est l’opérateur des différences premières, \(\Delta = (1-L)\). Si l’on s’intéresse à la somme des \(\gamma_i\), les estimations et les écarts types s’obtiennent directement à partir de l’estimation par 

\[
\text{mco de (2) ou (3), et si l’on s’intéresse à \(\lambda\), 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 \textbf{forme à correction d’erreur.} Le paramètre \(\lambda\) 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 \(X_{t-1}\) 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 \((\beta - 1) \left(Y_{t-1} - \lambda X_{t-1}\right)\) qui apparaît dans (5) terme de correction d’erreur, et le modèle est alors appelé \textbf{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 à 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 \(VARMAX\) \((\text{Vector AutoRegressive and Moving-Average processes with eXogenous regressors})\). La synthèse de cette procédure est la suivante,

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

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

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 

\[
\begin{pmatrix}
X_t \\
Y_t
\end{pmatrix} = 
\begin{pmatrix}
1.2 & 0.5 \\
0.6 & 0.3
\end{pmatrix}
\begin{pmatrix}
X_{t-1} \\
Y_{t-1}
\end{pmatrix} + 
\begin{pmatrix}
\varepsilon_t \\
\eta_t
\end{pmatrix}
\]

où 

\[
\Sigma = V(u_t) = 
\begin{pmatrix}
1.0 & 0.5 \\
0.5 & 1.25
\end{pmatrix}
\]
que nous simulons sur 100 observations,

\[ \text{PROC VARMAX data=donnes;} \]
\[ \text{MODEL X Y / p=1 noint lagmax=3; run ;} \]

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

\[
\begin{array}{|c|c|c|}
\hline
\text{Variable} & \text{Type} & \text{NoMissN} \\
\hline
\text{y1} & \text{DEP} & 100 \\
\text{y2} & \text{DEP} & 100 \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
\text{Variable} & \text{Type} & \text{NoMissN} & \text{Mean} & \text{StdDev} & \text{Min} & \text{Max} \\
\hline
\text{y1} & \text{DEP} & 100 & -0.17054 & 2.05299 & -4.80463 & 3.94386 \\
\text{y2} & \text{DEP} & 100 & -0.15409 & 2.02123 & -4.63446 & 3.69658 \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|}
\hline
\text{Type of Model} & \text{Estimation Method} & \text{Least Squares Estimation} \\
\hline
\text{VAR(1)} & \text{Least Squares Estimation} \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|}
\hline
\text{AR Coefficient Estimates} & \text{Lag Variable} & \text{y1} & \text{y2} \\
\hline
\text{y1} & \text{y1} & 1.20790 & -0.52848 \\
\text{y2} & \text{y2} & 0.58047 & 0.29561 \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
\text{Equation} & \text{Parameter} & \text{Estimate} & \text{Std Error} & \text{t Ratio} & \text{Prob>|T|} & \text{Variable} \\
\hline
\text{y1} & \text{AR(1,1)} & 1.20790 & 0.06766 & 17.85 & 0.0001 & \text{y1(t-1)} \\
\text{y1} & \text{AR(1,2)} & -0.52848 & 0.06925 & -7.63 & 0.0001 & \text{y2(t-1)} \\
\text{y2} & \text{AR(1,1)} & 0.58047 & 0.08304 & 6.99 & 0.0001 & \text{y1(t-1)} \\
\text{y2} & \text{AR(1,2)} & 0.29561 & 0.08500 & 3.48 & 0.0008 & \text{y2(t-1)} \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|}
\hline
\text{Covariance Matrix for the Innovation} & \text{Variable} & \text{y1} & \text{y2} \\
\hline
\text{y1} & 0.85478 & 0.60495 \\
\text{y2} & 0.60495 & 1.28770 \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|}
\hline
\text{Information Criteria} & \text{Variable} & \text{y1} & \text{y2} \\
\hline
\text{AICC(Corrected AIC)} & 0.85478 & 0.60495 \\
\text{HQC(Hannan-Quinn Criterion)} & 0.60495 & 1.28770 \\
\text{AIC(Akaike Information Criterion)} & 0.85478 & 0.60495 \\
\text{SBC(Schwarz Bayesian Criterion)} & 0.60495 & 1.28770 \\
\text{FPEC(Final Prediction Error Criterion)} & 0.60495 & 1.28770 \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|}
\hline
\text{Residual Cross-Covariance Matrices} & \text{Lag Variable} & \text{y1} & \text{y2} \\
\hline
\text{y1} & 0.09129 & 0.08060 \\
\text{y2} & 0.13094 & 0.03236 \\
\hline
\end{array}
\]

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

\[ \hat{\Phi} = \begin{bmatrix} 1.20790 & -0.52848 \\ 0.58047 & 0.29561 \end{bmatrix} \]

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,

\[ \hat{\Sigma} = \begin{bmatrix} 0.85478 & 0.60495 \\ 0.60495 & 1.28770 \end{bmatrix} \]

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

1.6.2 Étimation 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 modèle 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:

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

Considérons alors le programme $SAS$ suivant:

```sas
PROC VARMAX data=donnees;
MODEL X Y / p=2 noint dftest cointest=(johansen);
run ;
```

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 dît 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 $\alpha$ et $\beta$ : $(\alpha = 1.07837 -2.19184 \beta)$ correspondant à la première colonne de la matrice $\beta$ et $\hat{\alpha} = (-0.04867, 0.09518)$ correspondant à la seconde ligne de la matrice $\alpha$. On retrouve également une estimation des matrices autorégressives,

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

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

$$
\Sigma = \begin{pmatrix} 98.72 & 4.432 \\ 4.432 & 113.02 \end{pmatrix}
$$

relativement proche de $\Sigma = \begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix}$.

On obtient alors des sorties relativement proches de celles décrites dans le paragraphe précédent.
Estimating the VECM(2) model with one cointegrating relationship

The VARMAX Procedure
Type of Model: VECM(2) estimation Method: Least Squares Estimation
Cointegrated Rank: 1
Estimation Method: Maximum Likelihood Estimation
Type of Model: VECM(2)

Univariate Model Diagnostic Checks

| Variable | R-square | StdDev | F Value | Prob>|F |
|----------|----------|--------|---------|------|
| y1       | -0.7432  | 0.1670 | 482.78  | <.0001 |
| y2       | 0.4951   | 0.2412 | 113.02  | <.0001 |

Model Parameter Estimates

| Equation Parameter | Estimate | Std Error | T Ratio | Prob>|T |
|--------------------|----------|-----------|---------|------|
| y(t)               |          |           |         |      |
| y1                 | 0.9391   | 9.7343    | 482.78  | <.0001 |
| y2                 | -0.5715  | -0.5715   | <.0001  |      |
| y1(t-1)            | 0.1067   | 0.0514    | 2.07    | 0.0409 |
| y2(t-1)            | 0.5204   | 0.0361    | 14.41   | <.0001 |
| y1(t-2)            | -0.2080  | 0.0338    | -6.16   | <.0001 |
| y2(t-2)            | 0.7492   | 0.0516    | 14.52   | <.0001 |

Residual Cross-Correlation Matrix

<table>
<thead>
<tr>
<th>Lag</th>
<th>y1</th>
<th>y2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>96.8550</td>
<td>3.2930</td>
</tr>
<tr>
<td>1</td>
<td>3.2930</td>
<td>96.8550</td>
</tr>
<tr>
<td>2</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>3</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>4</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>5</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>6</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>7</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>8</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>9</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>10</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>11</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>12</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Univariate Model Diagnostic Checks

| Variable | R-square | StdDev | F Value | Prob>|F |
|----------|----------|--------|---------|------|
| y1       | 0.9391   | 9.7343 | <.0001  |      |
| y2       | -0.5715  | -0.5715| <.0001  |      |

The VARMAX Procedure
Type of Model: VECM(2) estimation Method: Least Squares Estimation
Cointegrated Rank: 1
Estimation Method: Maximum Likelihood Estimation
Type of Model: VECM(2)

Univariate Model Diagnostic Checks

| Variable | R-square | StdDev | F Value | Prob>|F |
|----------|----------|--------|---------|------|
| y1       | 0.9391   | 9.7343 | <.0001  |      |
| y2       | -0.5715  | -0.5715| <.0001  |      |

Model Parameter Estimates

| Equation Parameter | Estimate | Std Error | T Ratio | Prob>|T |
|--------------------|----------|-----------|---------|------|
| y(t)               |          |           |         |      |
| y1                 | 0.9391   | 9.7343    | <.0001  |      |
| y2                 | -0.5715  | -0.5715   | <.0001  |      |
| y1(t-1)            | 0.1067   | 0.0514    | 2.07    | 0.0409 |
| y2(t-1)            | 0.5204   | 0.0361    | 14.41   | <.0001 |
| y1(t-2)            | -0.2080  | 0.0338    | -6.16   | <.0001 |
| y2(t-2)            | 0.7492   | 0.0516    | 14.52   | <.0001 |

Residual Cross-Correlation Matrix

<table>
<thead>
<tr>
<th>Lag</th>
<th>y1</th>
<th>y2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>96.8550</td>
<td>3.2930</td>
</tr>
<tr>
<td>1</td>
<td>3.2930</td>
<td>96.8550</td>
</tr>
<tr>
<td>2</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>3</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>4</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>5</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>6</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>7</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>8</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>9</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>10</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>11</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>12</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Univariate Model Diagnostic Checks

| Variable | R-square | StdDev | F Value | Prob>|F |
|----------|----------|--------|---------|------|
| y1       | 0.9391   | 9.7343 | <.0001  |      |
| y2       | -0.5715  | -0.5715| <.0001  |      |
### Univariate Model Diagnostic Checks

<table>
<thead>
<tr>
<th>Variable</th>
<th>DW(1)</th>
<th>ChiSq F Value</th>
<th>Prob&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>1.93</td>
<td>2.14</td>
<td>0.3437</td>
</tr>
<tr>
<td>y2</td>
<td>1.83</td>
<td>0.46</td>
<td>0.7951</td>
</tr>
</tbody>
</table>

Estimating the VECM(2) model with one cointegrating relationship

### VARMAX Procedure

<table>
<thead>
<tr>
<th>Variable</th>
<th>AR1</th>
<th>AR1-2</th>
<th>AR1-3</th>
<th>AR1-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>F</td>
<td>Value</td>
<td>Prob&gt;F</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
<td>0.8541</td>
<td>0.69</td>
<td>0.69</td>
</tr>
<tr>
<td>y2</td>
<td>0.44</td>
<td>0.5068</td>
<td>3.11</td>
<td>0.0491</td>
</tr>
</tbody>
</table>

### Infinite Order AR Representation

Lag Variable:

<table>
<thead>
<tr>
<th>y1</th>
<th>y2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.21013</td>
</tr>
<tr>
<td>2</td>
<td>0.74332</td>
</tr>
</tbody>
</table>

### 1.7 Conseils bibliographiques

2 Les modèles ARCH - Autorégressifs Conditionellement Hétéroscédastiques

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

![Histogramme de la série CAC40](image1)

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.

![Diagramme de la série ARMA](image2)

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 zones 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 : \(\mathbb{E}(Y_t) = \mu\) pour tout \(t \in \mathbb{Z}\),
- les autocovariances ne dépendent que de la différence entre les observations : \(\text{cov}(Y_t, Y_s) = \gamma(|t - s|)\).

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 \(n, t_1, \ldots, t_n\) et \(h\), on a l’égalité entre les lois jointes

\[
\mathcal{L}(Y_{t_1}, \ldots, Y_{t_n}) = \mathcal{L}(Y_{t_1+h}, \ldots, Y_{t_n+h})
\]

**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

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

### 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

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

où \((\varepsilon_t)\) est un bruit blanc gaussien, centré, de variance \(\sigma^2\), soit \(\varepsilon_t \sim N(0, \sigma^2)\).

Ceci peut aussi s’écrire

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

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-corrélées, mais ne sont pas indépendantes.

2.2.1 Processus ARCH \((1)\)

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

Dans le cas où \((\varepsilon_t)\) est un bruit blanc fort \((\varepsilon_t \ i.i.d.)\) gaussien, alors \((\varepsilon_t)\) peut être vu comme une différence de martingale. Pour que le processus \((X_t)\) soit stationnaire au second ordre, la variance marginale de \((X_t)\) soit 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 $h \geq 1$,
\[ \mathbb{E} \left( X_t | X_{t-h} \right) = 0, \]
puisque $\mathbb{E} \left( X_t | X_{t-h} \right) = \mathbb{E} \left( \mathbb{E} \left( X_t | X_{t-1} \right) | X_{t-h} \right) = \mathbb{E} \left( 0 | X_{t-h} \right) = 0$. De plus,
\[ V \left( X_t | X_{t-h} \right) = \alpha_0 \frac{1 - \alpha^h}{1 - \alpha^2} + \alpha_1 X^2_{t-h} \text{ et } V \left( X_t \right) = \frac{\alpha_0}{1 - \alpha^2}. \] (15)
Aussi,
\[ V \left( X_t | X_{t-h} \right) - V \left( X_t \right) = \alpha^h \left[ X^2_{t-h} - V \left( X_{t-h} \right) \right]. \]

Comme nous l’avions vu en introduction, les modèles $ARCH$ permettent d’avoir des processus avec des queues de distribution plus épaisses que les processus $ARMA$. La kurtosis conditionnelle est obtenue à l’aide de la relation
\[ \mathbb{E} \left( X^4_t | X_{t-1} \right) = 3 \left[ \alpha_0 + \alpha_1 X^2_{t-1} \right]^2, \]
et la kurtosis à la date $t$ est
\[ \mathbb{E} \left( X^4_t \right) = 3 \left[ \alpha^2_0 + \frac{2 \alpha_1 \alpha^2_0}{1 - \alpha^2} + \alpha^2_2 \mathbb{E} \left( X^4_{t-1} \right) \right] = \frac{3 \alpha^2_0 (1 + \alpha^2)}{(1 - 3 \alpha^2) (1 - \alpha^2)}. \]
La condition d’existence du moment d’ordre 4 est $3 \alpha^2 < 1$, et on en déduit alors l’expression de la kurtosis $\kappa = \frac{\mathbb{E} \left( X^4_t \right)}{\mathbb{E} \left( X^2_t \right)^2} = \frac{3 - \alpha^2}{1 - 3 \alpha^2} > 3 \text{ (cas de la loi normale)}$.

Les queues de la distribution marginale d’un processus $ARCH$ (1) sont donc plus épaisses 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
\[ X_t = \varepsilon_t h_t \text{ où } h^2_t = \alpha_0 + \alpha_1 X^2_{t-1} \text{ et } \varepsilon_t \sim N\left(0, \sigma^2\right), \]
et les modèles $ARCH$ ($p$) font intervenir plusieurs retards,
\[ X_t = \varepsilon_t h_t \text{ où } h^2_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X^2_{t-i} \text{ et } \varepsilon_t \sim N\left(0, \sigma^2\right). \]
La volatilité de la date $t$ est alors fonction des carrés des écarts à la moyenne observés dans le passé proche. Si les coefficients $\alpha_i$ sont tous positifs (et assez grands), il y a une persistance des niveaux de volatilité : on observe des périodes de forte volatilité suivies de période de faible volatilité. On peut alors écrire
\[ \mathbb{E} \left( X_t | X_{t-1} \right) = 0, \text{ et } \mathbb{E} \left( X^2_t | X_{t-1} \right) = h^2_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X^2_{t-i}. \]

### 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,
\[ X_t = \varepsilon_t h_t \text{ où } h^2_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X^2_{t-i} + \sum_{j=1}^{q} \beta_j h^2_{t-j} \text{ et } \varepsilon_t \sim N\left(0, \sigma^2\right). \] (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 $(h^2_t)$. On a alors
\[ \mathbb{E} \left( X_t | X_{t-1} \right) = 0 \text{ et } \mathbb{E} \left( X^2_t | X_{t-1} \right) = h^2_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X^2_{t-i} + \sum_{j=1}^{q} \beta_j h^2_{t-j}. \]
Propriété 5 Si le processus GARCH \((p, q)\) est stationnaire au second ordre, alors nécessairement

\[
\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1,
\]

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 modèle 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 admisit 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

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

### 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 \(\varepsilon_t\) 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 \(\varepsilon_t\) n’est plus un bruit blanc, en particulier lorsque \(\varepsilon_t\) est un processus AR\((1)\), \(\varepsilon_t = \theta \varepsilon_{t-1} + \eta_t\), et lorsque \(\varepsilon_t\) 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

\[
Y_t = \mu + \phi Y_{t-1} + \varepsilon_t \quad \text{où } |\phi| < 1 \text{ et } \varepsilon_t \text{ satisfait}
\]

\[
\begin{align*}
(1) \quad & \mathbb{E}(\varepsilon_t | Y_{t-1}) = 0 \\
(2) \quad & V(\varepsilon_t | Y_{t-1}) = c + \alpha \varepsilon_{t-1}^2.
\end{align*}
\]

L’espérance conditionnelle de \((Y_t)\) est alors

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

et sa variance conditionnelle, en utilisant (15)

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

qui n’est pas constante (puisqu’elle dépend de \(\varepsilon_{t-h}\)).

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

#### Propriété 6
1. Le processus \(\varepsilon_t\) est orthogonal aux valeurs passées, quel que soit le retard : \(\mathbb{E}(\varepsilon_t | Y_{t-h}) = 0\) pour tout \(h\).
2. La propriété d’orthogonalité implique que les corrélations conditionnelles sont nulles : \(\text{cov}(\varepsilon_t, \varepsilon_{t+k} | Y_{t-h}) = 0\).
Preuve. Cette seconde propriété s'obtient de la façon suivante

$$\text{cov} (\varepsilon_t, \varepsilon_{t+k} | \varepsilon_{t-h}) = \mathbb{E} (\varepsilon_t \varepsilon_{t+k} | \varepsilon_{t-h}) - \mathbb{E} (\varepsilon_t | \varepsilon_{t-h}) \mathbb{E} (\varepsilon_{t+k} | \varepsilon_{t-h}) = \mathbb{E} (\varepsilon_t \varepsilon_{t+k} | \varepsilon_{t-h})$$

$$\text{cov} (\varepsilon_t, \varepsilon_{t+k} | \varepsilon_{t-h}) = \mathbb{E} (\varepsilon_{t+k} | \varepsilon_{t-h})$$

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 $\varepsilon_t$ n’est pas constante, la variance (non conditionnelle) est constante :

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

et donc

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

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

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 $\varepsilon_t$ et $\varepsilon_{t-h}$. L’idée peut alors être de tester le caractère ARCH des résidus.
obtenus, pour expliquer cette forte kurtosis.

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

\[ Y_t = 0.79404 Y_{t-1} + \varepsilon_t \quad \text{où} \quad \varepsilon_t = \eta_t \frac{1.335464}{(0.116)} + 0.42691 \frac{\varepsilon_t^2}{2} \]

et où \( \eta_t \) 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

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

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

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

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

Nous allons tester ici l’hypothèse \( H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0 \), 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 \( \hat{\varepsilon}_{t-1}^2, \ldots, \hat{\varepsilon}_{t-p}^2 \):

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

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

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

(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

\[ \hat{\ell}_{\phi=1} = \hat{\phi} - 1 \quad \text{sous} \quad \hat{\phi} = 1 \quad \text{et} \quad \hat{\phi} = \frac{\sum_{t=1}^{T} \hat{\phi} X_{t-1} \hat{\eta}_t}{\sum_{t=1}^{T} \hat{\varepsilon}_{t-1}^2}, \]

satisfait à

\[ \hat{\phi} = 1 \quad \text{sous} \quad \hat{\phi} = 1 \quad \text{et} \quad \hat{\phi} = \frac{\sum_{t=1}^{T} \hat{\phi} X_{t-1} \hat{\eta}_t}{\sum_{t=1}^{T} \hat{\varepsilon}_{t-1}^2}, \]

Dans 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 $\hat{\sigma}(\hat{\phi})$ écart type (par moindre carrés) de l’estimateur de $\phi$, et sa distribution est donnée par

$$\hat{\phi}_{\phi=1} = \frac{\hat{\phi} - 1}{\hat{\sigma}(\hat{\phi})} \xi - \int_0^1 W_t dW_t \neq N(0,1)$$

où $(W_t)$ 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 résidus.$^6$

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

$$\mathcal{L}(t) = \text{constante} - \frac{1}{2} \sum_{t=1}^T \frac{\varepsilon_t^2}{h_t},$$

et la log-vraisemblance totale du modèle s’écrit

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

La méthode pour estimer les paramètres est alors la suivante

1. Calcul des résidus/erreurs $\varepsilon_t^2$, et on effectue la régression linéaire de $\varepsilon_t^2$ sur son passé $\varepsilon_{t-1}^2, \ldots, \varepsilon_{t-p}^2$ :

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + u_t.$$

2. 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 $\hat{\alpha} = (\hat{\varepsilon}^2 - \varepsilon^2) \hat{\varepsilon}^2 h$ et

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

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 $\hat{\sigma}$, 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 con…ance 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, si $l_t(y, \theta)$ désigne la vraisemblance de $Y_t$, conditionnelle au passé, la vraisemblance de $Y_1, \ldots, Y_T$, conditionnelle à $Y_0$ est alors

$$L(y, \theta) = \prod_{t=1}^T l_t(y, \theta)$$

et on pose $\hat{\theta} = \arg \max \log L(y, \theta)$.

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

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

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

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

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

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 \( \sigma^2 \), sont facilement estimables, une estimation de la variance des autocorrélations peut être obtenue comme

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

et l’intervalle de confiance corrigé des effets ARCH devient, à 95\%, \( \pm 1.96/\sqrt{T} \). La statistique de Box-Pierce corrigée est alors donnée par

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

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 \( ARCH \) dont les erreurs suivent un bruit blanc ou un modèle \( ARMA \) dont les erreurs suivent un bruit blanc. 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

\[
\hat{X}_t (1) = EL (X_{t+1} | X_t, X_{t-1}, \ldots, \varepsilon_t, \varepsilon_{t-1}, \ldots) = \mu + \rho X_t + EL (\varepsilon_{t+1} | \varepsilon_t, \varepsilon_{t-1}, \ldots).
\]

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

\[
\hat{X}_t (h) = \mu + \rho \hat{X}_t (h-1) = \frac{1 - \rho^h}{1 - \rho} \mu + \rho^h X_t,
\]

dont l’erreur de prévision est \( e_h = X_{t+h} - \hat{X}_t (h) = \varepsilon_{t+h} + \rho \varepsilon_{t+h-1} + \ldots + \rho^{h-1} \varepsilon_{t+1} \). On peut noter que \( \hat{X}_t (h) \to \mu / (1 - \rho) \) quand \( h \to \infty \).

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

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

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

\[
\left[ \hat{X}_t (1) - 1.96\sigma, \hat{X}_t (1) + 1.96\sigma \right],
\]

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

\[
\left[ \hat{X}_t (h) - 1.96\sigma \sqrt{1 + \rho^2 + \ldots + \rho^{2(h-1)}}, \hat{X}_t (h) + 1.96\sigma \sqrt{1 + \rho^2 + \ldots + \rho^{2(h-1)}} \right].
\]

Modèle \( AR(1) - ARCH(1) \) Supposons que \( (\varepsilon_t) \) 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 \( (\eta_t) \) est un bruit blanc que l’on supposera de variance unitaire, indépendant du passé de \( (\varepsilon_t) \), c’est à dire \( \eta_t \) indépendant de \( \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots \). On notera tout d’abord que \( \mathbb{E} (\varepsilon_t \varepsilon_s) = 0 \) pour tout \( t \), que la corrélation entre deux dates est donnée par la relation

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

\[50\]
et que la variance à la date $t$ vérifie
\[ \sigma_t^2 = \mathbb{E} \left( \varepsilon_t^2 \right) = \mathbb{E} \left( \eta_t^2 \right) = \mathbb{E} \left( \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \right) = \alpha_0 + \alpha_1 \mathbb{E} \left( \varepsilon_{t-1}^2 \right) = \alpha_0 + \alpha_1 \sigma_{t-1}^2, \]
c'est à dire $\sigma_t^2 = \alpha_0/\left(1 - \alpha_1\right)$. Pour obtenir un intervalle de confiance pour $X_{t+h}$, rajoutons une hypothèse supplémentaire de normalité du bruit ($\eta_t$). Dans ce cas, l'intervalle de confiance, à 95%, pour la valeur à la date $t+1$ est donné par
\[ \left[ \bar{X}_t(1) - 1.96\sigma \sqrt{\alpha_0 + \alpha_1 \sigma_t^2}, \bar{X}_t(1) + 1.96\sigma \sqrt{\alpha_0 + \alpha_1 \sigma_t^2} \right]. \]

Aussi, dans le cas d’un modèle $ARCH$, la taille de l’intervalle de confiance dépend de la valeur $\sigma_t^2$.

**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 $(\varepsilon_t)$ suit un modèle $GARCH (1, 1)$,
\[
\begin{align*}
X_t &= \rho X_{t-1} + \varepsilon_t \\
\varepsilon_t &= \eta_t \sqrt{h_t} \\
h_t &= \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1} \\
\sigma_t^2 &= \omega + \alpha \sigma_{t-1}^2 + \beta \sigma_{t-1}^2,
\end{align*}
\]
où $\omega > 0$, $\alpha, \beta \geq 0$ et $\alpha + \beta \leq 1$ et $|\rho| < 1$. Alors, pour tout $h \geq 0$,
\[ X_{t+h} = \varepsilon_{t+h} + \rho \varepsilon_{t+h-1} + \ldots + \rho^h \varepsilon_t + \rho^{h+1} X_{t-1}, \]
de telle sorte que \( \mathbb{E} \left( X_{t+h} | X_{t-1} \right) = \rho^{h+1} X_{t-1} \) (ou même, plus généralement, \( \mathbb{E} \left( X_{t+h} | X_{t-1} \right) = \rho^{h+1} X_{t-1} \)). De plus,
\[ V \left( X_{t+h} | X_{t-1} \right) = \sum_{i=0}^{h} \rho^{2(h-i)} V \left( \varepsilon_{t+i} | \varepsilon_{t-1} \right), \]
où \( V \left( \varepsilon_{t+i} | \varepsilon_{t-1} \right) = \sigma_i^2 \) et pour $i \geq 1$,
\[
\begin{align*}
V \left( \varepsilon_{t+i} | \varepsilon_{t-1} \right) &= \mathbb{E} \left( \sigma_i^2 | \varepsilon_{t-1} \right) = \omega + (\alpha + \beta) \mathbb{E} \left( \sigma_{t+i-1}^2 | \varepsilon_{t-1} \right) \\
&= \left(1 + \ldots + (\alpha + \beta)^i \right) + (\alpha + \beta)^i \sigma_i^2 = \frac{1 - (\alpha + \beta)^i}{1 - (\alpha + \beta)} + (\alpha + \beta)^i \sigma_i^2,
\end{align*}
\]
de telle sorte que
\[
V \left( X_{t+h} | X_{t-1} \right) = \omega \left(1 - \rho^{2(h+1)}\right) + \left[ \sigma_i^2 - \frac{\omega}{1 - (\alpha + \beta)} \right] \frac{\rho^{2(h+1)} - (\alpha + \beta)^{h+1}}{\rho^2 - (\alpha + \beta)} \text{ si } \rho^2 \neq \alpha + \beta = \omega \left(1 - \rho^{2(h+1)}\right) + \left[ \sigma_i^2 - \frac{\omega}{1 - (\alpha + \beta)} \right] (h+1) \rho^{2h} \text{ si } \rho^2 = \alpha + \beta.
\]
On peut d’ailleurs noter que si $h \to \infty$, on retrouve la variance non conditionnelle de $X_t$, i.e.
\[
\lim_{h \to \infty} V \left( X_{t+h} | X_{t-1} \right) = V \left( X_t \right) = V \left( \varepsilon_t \right) \frac{1}{1 - \rho^2}.
\]

**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 $\eta_t$ gaussien, $\eta_t \sim N(0, 1)$)

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

**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 CAC 40, 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.

![Graphique montrant les séries ARCH et NASDAQ](image)

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. Queues de distribution épaisses : 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 $\left(-x^2/2\right)$. On parle alors de distribution leptokurtique.

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

5. Queues épaisses 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 épaisses (même si cela est moins important que la distribution non conditionnelle, i.e. la distribution des rendements).


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’independance 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.

$$dX_t = \mu(X_t,t) \, dt + \sigma(X_t,t) \, dW_t$$  (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 Browien géométrique, i.e. $\mu$ et $\sigma$ 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 $t = 1, ..., T$ (en supposant $(X_t)$ observé à la date 0), soit $X_t = X_{t-1} \exp(Y_t)$. Nous noterons $\Delta$ le temps entre les deux observations, exprimé en année$^8$. Puisque $(X_t)$ suit un Browien 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 $\sigma^2_Y = \sigma^2 \Delta, \overline{Y}$ et $S_Y$ les estimateurs empiriques de $\mu_Y$ et de $\sigma_Y$, i.e.

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

(qui sont alors des estimateurs consistants de $\mu_Y$ et de $\sigma_Y$). On peut alors estimer $\mu$ et $\sigma$ par

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

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 Browien géométrique, i.e. $dX_t = \mu dt + \sigma dW_t$, alors, en utilisant le fait que $\overline{Y} = 0.002276$ et $S_Y = 0.01915$ (cf statistiques ci-dessous), les paramètres $\mu$ et $\sigma$ peut estimés par

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

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 normalisé n’est pas valide.

$^8$On 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.  
\[ dS_t = \mu S_t + \sigma S_t dB_t \]
ou \( B_t \) est un mouvement brownien standard, et où \( \mu \) et \( \sigma \) sont des constantes. Toutefois, cette hypothèse de constance de \( \sigma \) s’est vite avérée, en pratique, inexacte\(^9\).

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,

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

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 \( \eta_t \) est un processus i.i.d. et \( (\sigma_t) \) une suite de variables aléatoires telles que
- \( \sigma_t \) soit mesurable par rapport à la filtration engendrée par le passé de \( \varepsilon_t \) - et éventuellement le présent - et par le passé d’un processus sous-jacent inobservable \( (\eta_t) \),
- \( \sigma_t \) et \( \eta_t \) soient des variables indépendantes,
- \( \sigma_t > 0 \) pour tout \( t \).

La variable aléatoire \( \sigma_t \) est appelée volatilité de \( \varepsilon_t \). 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 \( h > 0 \) : \( \varepsilon_t \) est alors un bruit blanc, au sens faible. Mais pas au sens fort. Plusieurs types de modèles existent alors suivant la spécification retenue pour \( (\sigma_t) \):

1. les processus conditionnellement hétéroscédastiques, de type GARCH pour lesquels la filtration \( F_t \) est engendrée par le passé de \( \varepsilon_t \). La volatilité est alors une fonction déterministe du passé de \( \varepsilon_t \). Les modèles GARCH standard sont caractérisés par une volatilité fonction affine des valeurs passées de \( \varepsilon^2_t \).
2. les processus à volatilité stochastique, pour lesquels la filtration \( F_t \) est engendrée par le passé de \( \varepsilon_t \). 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. \( \log(\sigma_t) = \omega + \rho \log(\sigma_{t-1}) + \nu_t \).
3. les processus à changement de régime, pour lesquels on suppose alors généralement que le processus sous-jacent \( (\nu_t) \) 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 \( i \), \( i = 0, 1, ..., n \), et \( u_i \) le log-rendement associé, \( u_i = \log(S_i) - \log(S_{i-1}) \) pour \( i = 1, ..., n \), on suppose que l’on peut écrire la volatilité \( \sigma^2_n \) sous la forme

\[ \sigma^2_n = \alpha_1 u^2_{n-1} + \alpha_2 u^2_{n-2} + ... + \alpha_{n-1} u^2_1 = \sum_{i=1}^{n-1} \alpha_i u^2_{n-i}, \]

où les variables \( \alpha_i \) correspondent aux poids donnés à la \( n \) ième 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, \( \sigma^2_n = \alpha_0 V + \alpha_1 u^2_{n-1} + \alpha_2 u^2_{n-2} + ... + \alpha_{n-1} u^2_1 \) où \( V \) correspond à une moyenne, à long terme de la volatilité.

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

\[ \sigma^2_n = \lambda \sigma^2_{n-1} + (1 - \lambda) u^2_{n-1} = (1 - \lambda) \sum_{i=1}^{n-1} \lambda^{i-1} u^2_{n-i} + \lambda^n \sigma^2_0. \]

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

\[ \sigma^2_n = \kappa + \beta \sigma^2_{n-1} + \alpha u^2_{n-1} \quad \text{où} \quad \kappa = \gamma V. \]

Dans le cas particulier où \( \kappa = 0 \), on retrouve le modèle EWMA.

\(^9\)Il 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’écarter 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 vette 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 \( \varepsilon_t \) 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 \( V(Y_t) = \sigma^2 / \left[ 1 - \lambda^2 \sigma^2 \right] \). Il est alors possible de montrer que \( Y_t \) s’écrit aussi

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

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

![Graphique de la série BILINEAIRE](image)

<table>
<thead>
<tr>
<th>Statistique</th>
<th>Valeur</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moyenne</td>
<td>0.049644</td>
</tr>
<tr>
<td>Médiane</td>
<td>0.075453</td>
</tr>
<tr>
<td>Maximum</td>
<td>3.793614</td>
</tr>
<tr>
<td>Minimum</td>
<td>-4.864823</td>
</tr>
<tr>
<td>Écart-type</td>
<td>1.221865</td>
</tr>
<tr>
<td>Asymétrie</td>
<td>-0.148632</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3.492173</td>
</tr>
<tr>
<td>Probabilité</td>
<td>0.000033</td>
</tr>
</tbody>
</table>

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 :

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

\[
Y_t = \begin{cases} 
\phi_1^{(1)} X_{t-1} + \phi_2^{(1)} X_{t-2} + \ldots + \phi_p^{(1)} X_{t-p} + \varepsilon_t & \text{si } X_{t-1} \leq s \\
\phi_1^{(2)} X_{t-1} + \phi_2^{(2)} X_{t-2} + \ldots + \phi_p^{(2)} X_{t-p} + \eta_t & \text{si } X_{t-1} > s,
\end{cases}
\]

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 \((\varepsilon_t)\) et \((\eta_t)\) 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 épaisse (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

\[
Y_t = \begin{cases} 
\alpha_1 + \phi_1 X_t + \varepsilon_t & \text{si } X_t \leq s \\
\alpha_2 + \phi_2 X_t + \eta_t & \text{si } X_t > s,
\end{cases}
\]

\[Y_t = (\alpha_1 + \phi_1 X_t) \mathbb{1}_{X_t \leq s} + (\alpha_2 + \phi_2 X_t) \mathbb{1}_{X_t > s} + u_t,\]

où \((u_t)\) est une séquence de bruits indépendants, dont la variance est de la forme \(V(u_t) = \sigma^2_m \mathbb{1}_{X_t \leq s} + \sigma^2_n \mathbb{1}_{X_t > s}\). 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 alternative fournie par le modèle à seuil. Plus généralement, on test \(H_0 : \alpha_1 = \alpha_2\) et \(\phi_i^{(1)} = \phi_i^{(2)}\) pour \(i = 1, \ldots, p\).
Tsay (1989) propose la statistique suivante, où \( m \) est le nombre d’états, et 
\[
Y_t = \phi_0^{(i)} + \phi_1^{(i)} Y_{t-1} + \phi_2^{(i)} Y_{t-2} + \ldots + \phi_p^{(i)} Y_{t-p} + \varepsilon_t^{(i)} \quad \text{si} \quad X_t \in \text{état } i,
\]
\[
Q(m) = \left( \frac{1}{m+1} \sum \varepsilon_t^2 - \sum \hat{a}_i^2 \right) \left( \frac{1}{n-b-2m-1} \sum \hat{a}_i^2 \right)^{-1},
\]
où \( b \) désigne le nombre d’observation utilisées pour initialiser les estimations récursives, i.e. \( b = n/10 + p \) (comme le conseillé Tsay), \( \hat{a}_i \) correspondent aux résidus \( \varepsilon_t \) standardisés, et \( \hat{a}_i \) les résidus de la régression de \( \hat{3}_{(i)} \) sur \( (1, X'_t) \).

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
\[
(\alpha_i, \phi_i) = \arg \min_{t, X_t \in \text{état } i} \text{SCR}_i \quad \text{où} \quad \text{SCR}_i = \sum_{t, X_t \in \text{état } i} (Y_t - \alpha_i - \phi_i X_t)^2,
\]
et l’estimateur de la matrice de variance covariance des coefficients estimés est alors donnée par
\[
\hat{\sigma}_i^2 = \frac{\text{SCR}_i}{n_i - (m + 1)} \quad \text{où} \quad m \text{ est le nombre d’état (ici 2)},
\]
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 \( \phi_i^{(i)} \) pour \( i = 1, 2 \). On évalue alors la variance résiduelle du modèle à seuil, conditionnellement à cette valeur du seuil : 
\[
\hat{\sigma}_T^2 (s) = \frac{1}{T} \sum_{t=1}^T (\hat{3}_t (s))^2 \quad \text{où} \quad 
\hat{3}_t (s) = Y_t - [(\alpha_1 + \phi_1 X_t) 1_{X_t \leq s} + (\alpha_2 + \phi_2 X_t) 1_{X_t > s}].
\]

Ces estimations sont alors faites pour différents seuils \( s \), et on choisit alors le seuil qui minimise la variance résiduelle,
\[
\hat{s} = \arg \min_s \hat{\sigma}_T^2 (s).
\]

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
\[
Y_t = X'i \beta + U_t.
\]
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 \( (\varepsilon_t) \) est i.i.d. Comme nous l’avons vu dans ce chapitre, il est possible que le bruit \( (U_t) \) soit conditionnellement hétéroscédastique, par exemple, modélisable par un processus ARCH (1) défini sous la forme
\[
U_t = \eta_t \sqrt{h_t} \quad \text{où} \quad h_t = \mu + \alpha U_{t-1}^2.
\]
Ce processus peut se généraliser sous la forme GARCH (1,1) suivante
\[
U_t = \eta_t \sqrt{h_t} \quad \text{où} \quad h_t = \mu + \alpha U_{t-1}^2 + \delta h_{t-1}.
\]

Parmi les autres modèles possible, on notera les modèles ARCH - M (ARCH en moyenne, ou ARCH in Mean), c’est-à-dire que le modèle linéaire s’écrira ici
\[
Y_t = X'i \beta + \gamma (h_t) + U_t \quad \text{où} \quad U_t = \eta_t \sqrt{h_t} \quad \text{avec} \quad h_t = \mu + \alpha U_{t-1}^2 + \delta h_{t-1}.
\]
Ce type de modèle a été introduit en 1987 par Engle, Lilien et Robbins. Parmi les modèles asymétriques, on notera les modèles \textit{TARCH} ou \textit{EGARCH}. Les modèles \textit{TARCH} (\textit{ARCH} à seuil, ou \textit{Threshold ARCH}) ont été introduits par Zakoian en 1994. Par exemple, le modèle \textit{TARCH} (1, 1) s’écrit

\[ U_t = \eta_t \sqrt{h_t} \text{ ou } h_t = \mu + \alpha U_{t-1}^2 + \phi U_{t-1}^2 I_{U_{t-1} < 0} + b h_{t-1}. \]

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

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

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

\[ \begin{array}{cccc}
\text{Variable} & \text{Coefficient} & \text{Std. Error} & \text{T-Statistic} & \text{Prob.} \\
\hline
C & 0.000863 & 0.000231 & 3.738457 & 0.0002 \\
AR(1) & 0.105134 & 0.018791 & 5.594793 & 0.0000 \\
\hline
\end{array} 
\]

\[ \begin{array}{cccc}
\text{Variable} & \text{Coefficient} & \text{Std. Error} & \text{T-Statistic} & \text{Prob.} \\
\hline
C & 0.000862 & 0.000228 & 3.782551 & 0.0002 \\
MA(1) & 0.106483 & 0.018791 & 5.666722 & 0.0000 \\
\hline
\end{array} 
\]

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

\[ \begin{array}{cccc}
\text{Variable} & \text{Coefficient} & \text{Std. Error} & \text{T-Statistic} & \text{Prob.} \\
\hline
\text{R-squared} & 0.0105 & Mean dependent var & 0.086119 \\
\text{Adjusted R-squared} & 0.0108 & S.D. dependent var & 1.097336 \\
\text{S.E. of regression} & 0.0105 & Akiakawa info criterion & 3.013434 \\
\text{Sum squared resid} & 3338.542 & Schwartz criterion & 3.017668 \\
\text{Log likelihood} & -4224.341 & F-statistic & 31.30171 \\
\text{Durbin-Watson stat} & 1.997377 & Prob(F-statistic) & 0.000000 \\
\hline
\end{array} 
\]

\[ \begin{array}{cccc}
\text{Variable} & \text{Coefficient} & \text{Std. Error} & \text{T-Statistic} & \text{Prob.} \\
\hline
\text{R-squared} & 0.0111 & Mean dependent var & 0.086119 \\
\text{Adjusted R-squared} & 0.0112 & S.D. dependent var & 1.097336 \\
\text{S.E. of regression} & 0.0105 & Akiakawa info criterion & 3.013434 \\
\text{Sum squared resid} & 3338.198 & Schwartz criterion & 3.017565 \\
\text{Log likelihood} & -4224.196 & F-statistic & 31.69207 \\
\text{Durbin-Watson stat} & 2.000222 & Prob(F-statistic) & 0.000000 \\
\hline
\end{array} 
\]

montrant que l’hypothèse de bruit blanc est rejetée, mais surtout, les graphiques ci-dessous 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 \textit{MA} (1) – \textit{GARCH} (1, 1) (c’est à dire un modèle \textit{MA} (1)
dont les erreurs suivent un modèle GARCH (1, 1) et ARCH – M, respectivement à gauche et à droite.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>T-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.084907</td>
<td>0.017700</td>
<td>4.797124</td>
<td>0.0000</td>
</tr>
<tr>
<td>MA(1)</td>
<td>0.17620</td>
<td>0.020952</td>
<td>8.19083</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

VARIANCE EQUATION

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>T-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.002788</td>
<td>0.000913</td>
<td>4.797124</td>
<td>0.0000</td>
</tr>
<tr>
<td>ARCH(1)</td>
<td>0.12770</td>
<td>0.020448</td>
<td>5.95503</td>
<td>0.0000</td>
</tr>
<tr>
<td>GARCH(1)</td>
<td>0.857095</td>
<td>0.021526</td>
<td>39.81666</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Inverted AR Roots

-0.17

Durbin-Watson stat 2.129878 Prob(F-statistic) 0.000000
Log likelihood -3775.938 F-statistic 5.008069
S.E. of regression 1.094213 Akaike info criterion 2.695856
R-squared 0.007104 Mean dependent var 0.086119
GARCH(1) 0.857095 0.021526 39.81666 0.0000
ARCH(1) 0.121770 0.020448 5.955103 0.0000
MA(1) 0.171620 0.020952 8.190983 0.0000
C 0.084907 0.017700 4.797124 0.0000

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 (εt) bruit blanc gaussien, c’est à dire i.i.d. dont la loi est N(0, σ²). Les hypothèses sont alors

\[ (H_0) : X_t = \mu + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t \] contre \[ (H_1) : X_t = \mu + \sum_{i=1}^{p} \phi_i X_{t-i} + \left( \varepsilon_t - \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 \right), \]
On estime le modèle \((H_0)\) par moindres carrés, et on calcule les résidus \(\widehat{\varepsilon}_t\) obtenus. On estime alors par les moindres carrés la régression
\[
\widehat{\varepsilon}_t^2 = \alpha_0 + \sum_{j=1}^q \alpha_j \widehat{\varepsilon}_{t-j}^2 + \eta_t \text{ avec } R^2 = 1 - \frac{\sum \eta_t^2}{\sum (\widehat{\varepsilon}_t - \overline{\varepsilon})^2}.
\]
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 \((\varepsilon_t)\) bruit blanc gaussien, c’est à dire i.i.d. dont la loi est \(N(0, \sigma^2)\). Les hypothèses sont alors
\[
(H_0) : X_t = \mu + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t \text{ contre } (H_1) : X_t = \mu + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t + \sum_{i=1}^Q \theta_{i,j} X_{t-i} \varepsilon_{t-j}.
\]
Là encore, on estime le modèle \((H_0)\) par moindres carrés, ainsi que \(\widehat{\sigma}^2\), estimateur de \(\sigma^2\). La statistique du test du multiplicateur de Lagrange est alors asymptotiquement équivalente à \(\chi^2\) degré de libertés.

\(AR(p)\) contre \(EXPAR(p)\) On considère un processus \((\varepsilon_t)\) bruit blanc gaussien, c’est à dire i.i.d. dont la loi est \(N(0, \sigma^2)\). Les hypothèses sont alors
\[
(H_0) : X_t = \mu + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t \text{ contre } (H_1) : X_t = \mu + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t + \sum_{i=1}^p \theta_i \exp\left(-\gamma X_{t-1}^2\right) X_{t-i} + \varepsilon_t.
\]
De façon équivalente l’hypothèse alternative peut s’écrire
\[
(H_1^\prime) : X_t = \mu + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=1}^p \theta_i X_{t-i}^2 X_{t-i-1} + \varepsilon_t.
\]
Là encore, la statistique du test du multiplicateur de Lagrange est
\[
LM_2 = \widehat{\sigma}^{-2} \left[ \sum \tilde{z}_{i,t,1} \tilde{\varepsilon}_t \right]^2 \left[ M_{11} - \bar{M}_{10} \bar{M}_{01}^{-1} \bar{M}_{01} \right] \left[ \sum \tilde{z}_{i,t,1} \tilde{\varepsilon}_t \right],
\]
 où
\[
\tilde{z}_{0,t} = (-1, X_{t-1}, \ldots, X_{t-p})^\prime
\]
\[
\tilde{z}_{1,t} = (-\widehat{\varepsilon}_{t}, \ldots, X_{t-1}, \ldots, \widehat{\varepsilon}_{t-1} X_{t-1}, \ldots, \widehat{\varepsilon}_{t-2} X_{t-1}, \ldots, \widehat{\varepsilon}_{t-Q} X_{t-1}, \ldots, \widehat{\varepsilon}_{t-Q} X_{t-p})^\prime
\]
et
\[
\bar{M}_{i} = \sum \tilde{z}_{i,t,1} \tilde{\varepsilon}_t
\]
\[
\bar{M}_{01} = \bar{M}_{10} = \sum \tilde{z}_{i,t,1} \tilde{\varepsilon}_t^2
\]
Il est possible de montrer que sous \((H_0)\), \(LM_2\) est asymptotiquement distribuée comme un chi-deux à \(pQ\) 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 reclassifier 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 connait le délai \( d \), et que l’on dispose des observations \( X_{t-d+1}, \ldots, X_0, \ldots, X_T \). On ordonne les observations \( X_1, \ldots, X_n \) selon les valeurs de \( X_{1-d}, \ldots, X_{-d} \). Soient \( X^T_1, \ldots, X^T_T \) les observations ainsi réordonnées. Pour tout \( j \), on considère le modèle linéaire suivant

\[
X^T_j = \mu_j + \sum_{k=1}^p \phi_k X_{(j-k)} + \epsilon^T_j \quad \text{pour tout } i \leq j.
\]

L’hypothèse \( (H_0) \) est alors \( \mu_j = \mu \), \( \phi_k = \phi_k \) 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 \( N \to \infty \), on pose alors \( C_{n,N}(r) \to C_{m}(r) = \alpha^{\nu_m} \). L’exposant \( \nu_m \) ainsi défini sera la dimension de corrélation dans l’immersion \( m \), dont l’estimateur naturel est

\[
\hat{\nu}_m = \lim_{r \to 0} \frac{d \log(C_m(r))}{d \log r}.
\]

A partir de cette notion, il est aisé 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 \( X^m_t = (X_t, X_{t+1}, \ldots, X_{t+m-1}) \), et \( T_m = n - m + 1 \) le nombre de vecteurs \( m \)-historiques pour la série \( (X_t) \) avec \( n \) observations. On désigne par \( I_\varepsilon(.) \) la fonction indicatrice de \( [0, \varepsilon] \), où \( \varepsilon \) est une constante positive préalablement choisie. On pose alors

\[
C_{n,m}(\varepsilon) = \frac{2}{T_m(T_m-1)} \sum_{t<s} I_\varepsilon(\|X^m_t - X^m_s\|_\infty),
\]

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

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

**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 \( C_{1,1}(\varepsilon) \) mesure toujours la proportion de points de la série distants de moins de \( \varepsilon \) de moins de \( \varepsilon \) de leur valeur à la date précédente. Dans le cas d’une série indépendante, on voit que l’on considère une surface élastique, et que l’on double la longueur des côtés \( (C_{1,1} \text{ est alors multiplié par } 2) \), 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 élastique voit ses côtés doublés : \( C_{n,1}(\varepsilon) \) est bien multiplié par 2 mais \( C_{n,2}(\varepsilon) \) n’est pas multiplié par 4.

61
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

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

et

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

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 $SCR_1 = \sum \eta_t^2$. La statistique de test est

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

qui suit sous $(H_0)$ ($hypothèse$ $de$ $modèle$ $AR(p)$) un 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

```sas
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 :

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

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 $(\eta_t)$ est normalement distribué (centré réduit)
La déclaration **MODE**: Pour les modèles **GARCH** :
- **GARCH**(option-list) : spécifie la forme du modèle **GARCH**, par exemple, pour un modèle **GARCH** (1,1) l'instruction sera

  \[
  \text{MODEL } y=x_1 \times x_2 \text{ / GARCH=(q=1, p=1);} 
  \]

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**=(obs,...,obs) : 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 =**(PHILIPS) : met en place le test de Phillips-Perron, test de stationnarité du modèle

La déclaration **OUTPUT** 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 hétéroscédastique (spécifié dans l'instruction **HETERO** ou la valeur de la variance de l'erreur conditionnelle \( h_t \) de la déclaration **GARCH= option** dans l'instruction **MODE**)

Parmi les tests possible, 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**.

Étimation 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,

![IBM Stock Returns (daily)](image)

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

**ARCH** (2) : \( Y_t | Y_{t-1}, Y_{t-2}, \ldots \sim \mathcal{N}(0, h_t) \) 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;
```

63
et la sortie SAS (estimation des paramètres et tests statistiques) est alors

Autoreg Procedure

Dependent Variable = R

Ordinary Least Squares Estimates

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
<td>0.032143</td>
<td>DFE</td>
<td>254</td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td>0.000127</td>
<td>Root MSE</td>
<td>0.011340</td>
<td></td>
</tr>
<tr>
<td>SBC</td>
<td>-1558.8</td>
<td>AIC</td>
<td>-1558.8</td>
<td></td>
</tr>
<tr>
<td>Reg Rsq</td>
<td>0.0000</td>
<td>Total Reg</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>Durbin-Watson</td>
<td>2.1377</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: No intercept term is used. R-squares are redefined.

GARCH Estimates

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
<td>0.032143</td>
<td>OBS</td>
<td>254</td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td>0.000127</td>
<td>UVAR</td>
<td>0.000126</td>
<td></td>
</tr>
<tr>
<td>Log L</td>
<td>781.0174</td>
<td>Total Reg</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>SBC</td>
<td>-1545.42</td>
<td>AIC</td>
<td>-1556.03</td>
<td></td>
</tr>
<tr>
<td>Normality Test</td>
<td>105.8567</td>
<td>Prob&gt;Chi-Sq</td>
<td>0.0001</td>
<td></td>
</tr>
</tbody>
</table>

Variable | DF | B Value | Std Error | t Ratio | Approx Prob |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ARCH0</td>
<td>1</td>
<td>0.000112</td>
<td>7.561E-6</td>
<td>14.851</td>
<td>0.0001</td>
</tr>
<tr>
<td>ARCH1</td>
<td>1</td>
<td>0.041348</td>
<td>0.0511</td>
<td>0.810</td>
<td>0.4181</td>
</tr>
<tr>
<td>ARCH2</td>
<td>1</td>
<td>0.069749</td>
<td>0.0432</td>
<td>1.616</td>
<td>0.1062</td>
</tr>
</tbody>
</table>

C'est à dire que \( \hat{\sigma}, \hat{\alpha_1} \) et \( \hat{\alpha_2} \) valent respectivement 0.00011, 0.04136 et 0.06976. Les coefficients donnés sont, pour l'estimation par moindres 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
  \[ R^2_{\text{REG}} = 1 - \frac{TSSE}{TSST} \]
  où \( TSST \) est la somme des carrés ...
- **DW**: statistique de Durbin-Watson
- **DFE**: (Error Degrees of Freedom) le nombre d'observations moins le nombre de paramètres
- **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
  \[ R^2_{\text{TOT}} = 1 - \frac{SSE}{SST} \]

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

IBM Stock Returns (daily)

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 équipondération, ±10 jours (soit 21 jours)), cette série n’est pas stationnaire

Pour stationniser la série, considérons la série différenciée, correspondant à la variation de l’indice CAC40 sur une journée (en points)
Cette transformation semble stationniser 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) :

<table>
<thead>
<tr>
<th>Test</th>
<th>Valeur de la statistique</th>
<th>p-value</th>
<th>Linéaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>LM (AR contre ARCH)</td>
<td>206.6051</td>
<td>0.00000</td>
<td>non</td>
</tr>
<tr>
<td>LM (AR contre BL)</td>
<td>9.236277</td>
<td>0.00237</td>
<td>non</td>
</tr>
<tr>
<td>LM (AR contre EXPAR)</td>
<td>39.06230</td>
<td>0.00000</td>
<td>non</td>
</tr>
<tr>
<td>RESET</td>
<td>8.876685</td>
<td>0.00293</td>
<td>non</td>
</tr>
</tbody>
</table>

L’étude des autocorrélogrammes de la série des variations et de la série au carré semblent confirmer ce phénomène : si les autocorrélogrammes de la série n’indiquent pas de corrélations importantes (laissant à penser que la série pourrait être un bruit blanc), des autocorrélations 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
  
  $$X_t | X_{t-1} \sim N \left( 0, \sigma_t^2 \right) \text{ où } \sigma_t^2 = 398.6559 + 0.1488 X_{t-2}^2.$$ 

  Considérons la série des résidus standardisés (la série des "résidus" coïncidant avec la série des variations elle-même).

  - Modèle GARCH (2, 1)

  [A INSÉRER]

  - Modèle ARMA

  [A INSÉRER]

2.8.2 Modélisation des rendements du cours Intel : modèle ARCH

[A INSÉRER]
### 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),

![Graphique des rendements mensuels de l’indice SP500](image)

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) \((\text{voir un ARMA} (3, 3))\)

![](image)

**Tableau des résultats de régression**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.006161</td>
<td>0.001955</td>
<td>3.151074</td>
<td>0.0017</td>
</tr>
<tr>
<td>MA(1)</td>
<td>0.094167</td>
<td>0.035150</td>
<td>2.678963</td>
<td>0.0075</td>
</tr>
<tr>
<td>MA(3)</td>
<td>-0.141599</td>
<td>0.035173</td>
<td>-4.025809</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

**Résultats de la régression**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.006262</td>
<td>0.001983</td>
<td>3.158519</td>
<td>0.0016</td>
</tr>
<tr>
<td>MA(1)</td>
<td>0.086019</td>
<td>0.035303</td>
<td>2.436595</td>
<td>0.0150</td>
</tr>
<tr>
<td>MA(3)</td>
<td>-0.125241</td>
<td>0.035302</td>
<td>-3.547748</td>
<td>0.0004</td>
</tr>
</tbody>
</table>

\[
X_t = 0.006 + \varepsilon_t + 0.094\varepsilon_{t-1} - 0.141\varepsilon_{t-3}
\]

Toutefois, si l’on regarde par exemple l’autocorrélogramme de la série \((\eta_t^2)\), ci-dessous à droite, on peut constater que \((\eta_t)\) n’est pas un bruit blanc.

**Tableau des résultats de régression**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>z-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>7.97E-05</td>
<td>2.42E-05</td>
<td>3.289444</td>
<td>0.0010</td>
</tr>
<tr>
<td>ARCH(1)</td>
<td>0.110723</td>
<td>0.039449</td>
<td>3.061283</td>
<td>0.0000</td>
</tr>
<tr>
<td>GARCH(1)</td>
<td>0.862719</td>
<td>0.020630</td>
<td>42.02175</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

**Résultats de la régression**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>z-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>7.06E-05</td>
<td>2.42E-05</td>
<td>3.289444</td>
<td>0.0010</td>
</tr>
<tr>
<td>ARCH(1)</td>
<td>0.110723</td>
<td>0.039449</td>
<td>3.061283</td>
<td>0.0000</td>
</tr>
<tr>
<td>GARCH(1)</td>
<td>0.862719</td>
<td>0.020630</td>
<td>42.02175</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

**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.

![Graphique des rendements journaliers de l'action IBM](image)

[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 trimestrielles\(^{10}\) aux États-Unis, de 1948 à 1993,

![Graphique du taux de chômage](image)

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.

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.

Cette présence de racine unité correspond au phénomène économique appelé "présence d’effets d’hystérésis du taux de chômage". Le test de Dickey Fuller (avec constante et tendance) rejette en revanche l’hypothèse de présence de racine unité.

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  
Convergence achieved after 3 iterations

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>0.70459</td>
<td>0.07456</td>
<td>9.60065</td>
<td>0.0000</td>
</tr>
<tr>
<td>AR(2)</td>
<td>0.25063</td>
<td>0.074075</td>
<td>-3.3889</td>
<td>0.0013</td>
</tr>
</tbody>
</table>

R-squared 0.437927  
Adjusted R-squared 0.434601  
S.E. of regression 0.330531  
Akaike info criterion 0.635393  
Sum squared resid 18.25335  
Schwarz criterion 0.672138  
Log likelihood -52.32612  
Durbin-Watson stat 2.07329  

Inverted AR Roots 0.40 + 0.30i  
0.40 - 0.30i

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: 07:51  
Sample(adjusted): 1949:1 1991:1  
Included observations: 169 after adjusting endpoints  
Convergence achieved after 3 iterations

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>0.735183</td>
<td>0.075316</td>
<td>9.761291</td>
<td>0.0000</td>
</tr>
<tr>
<td>AR(2)</td>
<td>-0.171892</td>
<td>0.078142</td>
<td>-2.199747</td>
<td>0.0292</td>
</tr>
<tr>
<td>AR(4)</td>
<td>-0.161908</td>
<td>0.060665</td>
<td>-2.668864</td>
<td>0.0084</td>
</tr>
</tbody>
</table>

R-squared 0.469192  
Adjusted R-squared 0.462797  
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 2.033721

Inverted AR Roots 0.64 - 0.45i  
0.64 + 0.45i  
-0.27 + 0.43i  
-0.27 - 0.43i

L’hypothèse de bruit blanc des résidus semble ici validée. 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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>5.801028</td>
<td>0.420081</td>
<td>13.80931</td>
<td>0.000</td>
</tr>
<tr>
<td>AR(1)</td>
<td>1.602622</td>
<td>0.577562</td>
<td>27.79812</td>
<td>0.000</td>
</tr>
<tr>
<td>AR(2)</td>
<td>-0.662140</td>
<td>0.574645</td>
<td>-11.52241</td>
<td>0.000</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.960940</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.960502</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.E. of regression</td>
<td>0.145521</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adj. R-squared</td>
<td>0.960964</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.960940</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean dependent var</td>
<td>0.099608</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. Error</td>
<td>0.145521</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-Statistic</td>
<td>10.09410</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Durbin-Watson stat 1.761995
Log likelihood -50.03334
Sum squared resid 18.01853
S.E. of regression 0.326525
Adjusted R-squared 0.960502
R-squared 0.960964
Mean dependent var 0.099608
Std. Error 0.145521

Inverted AR Roots: 0.80 -0.14i 0.80+0.14i

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)

[A INSERER]

2.9 Conseils bibliographiques

2.10 Compléments sur les séries temporelles en finance


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 $\Sigma_1, \Sigma_2, ..., \Sigma_N$ la suite des cotations, et $\nu_1, \nu_2, ..., \nu_N$, la suite correspondant au volume échangé à chaque cotation. Entre la date 0 et la date $t$ il est possible d’écrire $S_t = \Sigma_0 + (\Sigma_1 - \Sigma_0) + (\Sigma_2 - \Sigma_1) + ... + (\Sigma_N - \Sigma_{N-1})$, alors que le volume final échangé est $V_t = \nu_1 + \nu_2 + ... + \nu_N$. 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 $(\nu_t)$ 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 $\tau$, 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 » ? Supposons 11 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 prédiction devient vite simplement la moyenne du siècle passé, du XVe siècle. Elle ne rend pas bien compte du phénomène réellement observé au XVIe siècle. Il faut donc introduire une mémoire longue, même très longue, mais à la condition d’être capable de le 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’Égypte, 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 VIIe au XIe siècle, 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 XVe et du XVIIe 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 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 économétrie 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’)

12Ces modèles autosimilaires découverts par Paul Icvy ont un caractère fractal. Ils ont été popularisés par Benoît 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 $h \to \infty$ (décroissance exponentielle de $\rho(h)$). En fait, il est possible de montrer que, pour un processus ARMA stationnaire, sa fonction d’autocorrélation est bornée,

$$|\rho(h)| \leq Cr^h \text{ pour } h = 1, 2, ...$$

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

$$\rho(h) \sim C h^{2d-1} \text{ quand } h \to \infty$$

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, où $\beta$ est une constante non nulle et $\tau$ est une certaine constante.

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

$$\lim_{h \to \infty} \frac{\rho(h)}{Ch^{2d-1}} = 1.$$  

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 lente 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-persistance. 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 Fractional 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 informationel) si les cours des actifs cotés 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’autocorrelation, 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 \( \beta \) telle que \( 0 < \beta < 1 \) et une constante \( C > 0 \) telles que

\[
\lim_{\omega \to 0} \frac{f(\omega)}{C \omega^{-\beta}} = 1,
\]

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’un 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 \( \tau \), que la distribution de \( R(\tau)/\sqrt{\tau} \) est indépendante du temps \( \tau \) : en terme financier, l’ordre de grandeur de la rentabilité d’un titre financier pendant une période \( \tau \) est proportionnelle à la racine carrée de cette période. Une des application 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 repose 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

\[ X_t = Y_t - Y_{t-1} \quad \text{pour} \quad t = 1, 2, \ldots \]

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 écrites 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 \(X_0 = 0\) pour \(t < 0\) et \(\varepsilon_t\) un bruit blanc pour \(t \geq 0\), nul sinon, de variance finie \(\sigma^2\), 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
\[(1 - L)^d = 1 + \sum_{j=1}^{\infty} \frac{d(d-1)\ldots (d-j+1)}{j!} \, (-1)^j L^j.
\]

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
\[h_j \sim \frac{\Theta (1)}{\Phi (1) \Gamma (d)} \, j^{d-1} \quad \text{quand} \ j \to \infty.
\]

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 \(d < 1/2\). \((\varepsilon_t)\) est ici un bruit blanc, et nous supposerons, dans un premier temps, qu’il est de variance finie \(\sigma^2\).

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 \((\varepsilon_t)\) est un bruit blanc gaussien de variance \(\sigma^2\), peuvent être vus comme des discrétisations du mouvement brownien. De façon analogue, dans le cas où le bruit \((\varepsilon_t)\) 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 \(t\), \(B_H (0) = 0\), et \(\mathbb{E} (B_H (t) - B_H (s)) = 0\) pour tout \(s, t\), et
\[\text{cov} (B_H (t), B_H (s)) = \frac{\sigma^2}{2} \left[ |t|^{2H} - |t - s|^{2H} + |s|^{2H} \right],
\]
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
\[
\gamma(h) = \sigma^2 \frac{\Gamma(h + d)}{\Gamma(h + 1 - d) \Gamma(d) \Gamma(1 - d)} \quad \text{quand } t \to \infty,
\]
où \( \sigma^2 \) 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
\[
\gamma(h) \sim \sigma^2 h^{2d-1} \frac{\Gamma(1 - 2d)}{\Gamma(d) \Gamma(1 - d)} \quad \text{ou} \quad \rho(h) \sim \gamma_d h^{2d-1} \quad \text{quand } h \to \infty.
\]
A titre de comparaison, pour les processus \(\text{ARMA}\), 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 \(\text{FARIMA} \) Gaussien, c’est à dire que le bruit \( (\varepsilon_t) \) 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 \( \mathbb{P}(\varepsilon > x) \sim C x^{-\alpha} \)) de paramètre \( \alpha < 2 \), ou les distributions \( \alpha \)-stables (c’est à dire que leur fonction caractéristique est de la forme \( \phi(t) = \mathbb{E}(e^{it\varepsilon}) = \exp(-\sigma^\alpha |t|^\alpha) \)) où \( \sigma \) est un paramètre d’échelle) de paramètre \( 1 < \alpha \leq 2 \) (le cas \( \alpha = 2 \) correspondant au cas gaussien)\(^{13}\).

Les processus \(\text{ARFIMA}(0, d, 0)\) peuvent s’écrire sous la forme
\[
X_t = \sum_{k=0}^{+\infty} \frac{\Gamma(k + d)}{\Gamma(d) \Gamma(k + 1)} (-1)^k \varepsilon_{t-k} \quad \text{pour } d < 1/2,
\]
dont la densité spectrale est donnée par
\[
f_X(\omega) = \frac{\sigma^2}{\left|2 \sin \left( \pi \omega \right) \right|^{2d}} \approx \frac{\sigma^2}{\left|2 \pi \omega \right|^{2d}}.
\]
Le graphique ci-dessous présente les trajectoires de 3 processus \(\text{FARIMA} \), avec \( d = 0.9, 0.7 \) et 0.5,

**Remarque 20** La densité spectrale de ces processus \(\text{ARIMA}(0, d, 0)\) vérifie la relation suivante
\[
f(\omega) = \left(2 \sin \frac{\omega}{2}\right)^{-2d} \quad \text{pour } 0 < \omega \pi,
\]
de telle sorte que le comportement en 0 est donné par
\[
\text{également } f(\omega) = \omega^{-2d} \quad \text{quand } \omega \to 0.
\]
\(^{13}\)De plus amples informations peuvent être trouvées dans Samorodnitsky et Taqqu (1994) ‘Stable non-gaussian processes : stochastic models with infinite variance’

78
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 \( \gamma_k^X \) 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 \( \gamma_k^Y \) et \( \gamma_k^Z \) 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 \( \varepsilon_t \) 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

![Graphiques ARFIMA](image)

avec les autocorrélogrammes \((d = 0.3, d = 0.5, d = 0.9)\) ci-dessous,

![Autocorrélogrammes](image)

Les cas limites \( d \to 0 \) et \( d \to 1 \) 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 retiendrons 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

\[
\log |\hat{\rho}(h)| \sim \hat{\alpha} + \hat{\beta} \log h \text{ alors } \hat{d} = \frac{1}{2} \left( \hat{\beta} + 1 \right) .
\]
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 \(j = 1, \ldots, (T-1)/2\). 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

\[
\begin{align*}
\alpha &= \log f^*(0) - \xi \\
\beta &= d \\
Z_j &= -2 \log |2\sin(\lambda_j/2)| \\
\varepsilon_j &= \log(I_T(\lambda_j)) - \log f^*(\lambda_j) + \xi
\end{align*}
\]

où \(\xi\) est la constante d’Euler.

Pour \(0 \leq s < t < T\), on utilise l’estimateur

\[
\tilde{d}_R(s,t) = \frac{\sum_{j=s+1}^{t} (Z_j - \bar{Z}) \log(I_T(\lambda_j))}{\sum_{j=s+1}^{t} (Z_j - \bar{Z})^2} \quad \text{où} \quad \bar{Z} \text{ est la moyenne des } Z_j.
\]
### 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 $(\varepsilon_t)$ est un bruit blanc de variance finie $\sigma^2$. 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

$$L_W(X, \theta) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \left( \log f_X(\lambda, \theta) + \frac{I_T(\lambda)}{f_X(\lambda, \theta)} \right) d\lambda,$$

où $f_X$ et $I_T$ sont respectivement la densité spectrale et le périodogramme du processus. En notant $\theta^* = (1, d)'$ et en choisissant $\sigma^2$ (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, $L_W(X, \theta)$ s’écrit simplement sous la forme

$$L^*_W(X, \theta) = L_W(X, \theta^*) = \log \sigma^2 + \frac{1}{2\pi\sigma^2} \int_{-\pi}^{+\pi} \frac{I_T(\lambda)}{f_X(\lambda, \theta^*)} d\lambda.$$

L’estimateur de Whittle $\hat{\theta} = (\hat{\sigma^2}, \hat{d})'$ est alors obtenu en minimisant la fonction $L_W(X, \theta)$ par rapport à $\theta^*$. Puisque l’échantillon est de taille finie, cela revient à minimiser par rapport à $\theta^*$ la fonction $Q_T(\theta^*)$ définie par

$$Q_T(\theta^*) = \frac{1}{(T-1)/2} \sum_{i=1}^{(T-1)/2} \frac{I_T(\lambda_i)}{f_X(\lambda_i, \theta^*)} + \hat{\sigma^2}.$$

### 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 dès 1951, suit à un article de Hurst sur le stockag 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 communications, Halett et Raftery (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 hétérosédasticité, 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 $ARCH$ décroit 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.

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 $(R_t)$.


$\gamma_{25} = 0.074$, $\gamma_{38} = 0.050$, $\gamma_{69} = 0.053$, $\gamma_{86} = 0.062$...etc Un grand nombre d’autocorrélations sont encore significatives ($\gamma_{179}$, $\gamma_{183}$, $\gamma_{185}$ par exemple sont encore significativement non nulles).
suitants sont obtenus, par la méthode de Whittle,

<table>
<thead>
<tr>
<th></th>
<th>SP500</th>
<th>CAC40</th>
<th>DAX30</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{d}_W )</td>
<td>0.149</td>
<td>0.126</td>
<td>0.178</td>
</tr>
<tr>
<td>( \mathcal{L}_W )</td>
<td>-14.13</td>
<td>-13.71</td>
<td>-13.59</td>
</tr>
</tbody>
</table>

où \( \mathcal{L}_W \) désigne la valeur de la vraisemblance de Whittle, définie dans la partie précédente. Les coefficients estimés sont significativement positifs, ce qui confirme l’hypothèse de mémoire longue persistante.

**Étude du carré du rendement** \( R^2_t \). 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

<table>
<thead>
<tr>
<th></th>
<th>SP500</th>
<th>CAC40</th>
<th>DAX30</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{d}_W )</td>
<td>0.170</td>
<td>0.154</td>
<td>0.162</td>
</tr>
<tr>
<td>( \mathcal{L}_W )</td>
<td>-20.95</td>
<td>-20.21</td>
<td>-19.76</td>
</tr>
</tbody>
</table>

**Étude 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 \( R^2_t \)

<table>
<thead>
<tr>
<th></th>
<th>87/89</th>
<th>90/94</th>
<th>95/00</th>
<th>87/00</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{d}_W )</td>
<td>0.202</td>
<td>0.118</td>
<td>0.126</td>
<td>0.150</td>
</tr>
<tr>
<td>( \mathcal{L}_W )</td>
<td>-13.27</td>
<td>-13.84</td>
<td>-13.71</td>
<td>-13.59</td>
</tr>
</tbody>
</table>

suivi du comportement de \( R^2_t \)

<table>
<thead>
<tr>
<th></th>
<th>87/89</th>
<th>90/94</th>
<th>95/00</th>
<th>87/00</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{d}_W )</td>
<td>0.182</td>
<td>0.134</td>
<td>0.166</td>
<td>0.170</td>
</tr>
<tr>
<td>( \mathcal{L}_W )</td>
<td>-18.59</td>
<td>-20.31</td>
<td>-20.31</td>
<td>-19.66</td>
</tr>
</tbody>
</table>

**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 marché 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 marché 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 s’écarte 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 cet 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’hétéroscédasticité 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 COMPLETER]

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’innodation, la quantité \(R(d)\) est le volume du réservoir. Il est également possible de définir \(S(t)\) de telle sorte que

\[S^2(t) = \frac{1}{d} \sum_{i=1}^{d} X^2_i - \frac{1}{d^2} \left( \sum_{i=1}^{d} X_i \right)^2\]

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 Lawrance et Kottegoda (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 États-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.

![Graphique montre les fonctions d'autocorrélation.](image)

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

\[(1 - \phi L^{12})(1 - L)^d (X_t - [\hat{c} + \hat{\delta}_{1980-07}]) = (1 - \theta L^{12}) \varepsilon_t,\]


<table>
<thead>
<tr>
<th>Méthode</th>
<th>EML</th>
<th>MPL</th>
<th>NLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d)</td>
<td>0.400</td>
<td>0.411</td>
<td>0.409</td>
</tr>
<tr>
<td>(\hat{\phi})</td>
<td>0.769 (0.101)</td>
<td>0.783 (0.091)</td>
<td>0.620 (0.110)</td>
</tr>
<tr>
<td>(\hat{\theta})</td>
<td>-0.605 (0.123)</td>
<td>-0.621 (0.113)</td>
<td>-0.447 (0.121)</td>
</tr>
<tr>
<td>(\hat{c})</td>
<td>0.003 (0.002)</td>
<td>0.003 (0.003)</td>
<td>0.04 (0.002)</td>
</tr>
<tr>
<td>(\hat{\delta})</td>
<td>-0.011 (0.002)</td>
<td>-0.011 (0.002)</td>
<td>-0.011 (0.002)</td>
</tr>
<tr>
<td>(\hat{\sigma}^2)</td>
<td>3.626</td>
<td>3.625</td>
<td>3.548</td>
</tr>
</tbody>
</table>

On peut noter que pour chacune des méthodes \(d < 0.5\)

### 3.5 Conseils bibliographiques


4 Compléments : exercices

4.1 Exercices avec correction

**Exercice 1** Soit $(\varepsilon_t)$ 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) $X_t = \varepsilon_t \cos(\varepsilon t) + \varepsilon_{t-1} \sin(\varepsilon t)$ 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 = $ constante. Sa variance est donnée par $\mathbb{V}(X_t) = [b^2 + c^2]\varepsilon^2$ où $\varepsilon^2$ est la variance du bruit, puisque le bruit est non-autocorrélé : $\mathbb{C}_{\varepsilon}(\varepsilon_t, \varepsilon_{t-1}) = 0$.

De plus, $\mathbb{C}(X_t, X_{t-1}) = b c \varepsilon^2$ qui ne dépend pas de $t$ et $\mathbb{C}(X_t, X_{t-h}) = 0$ 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 $\mathbb{V}(X_t) = [\cos^2(\varepsilon t) + \sin^2(\varepsilon t)]\varepsilon^2$ où $\varepsilon^2$ est la variance du bruit, puisque le bruit est non-autocorrélé : $\mathbb{C}_{\varepsilon}(\varepsilon_t, \varepsilon_{t-1}) = 0$.

Enfin, $\mathbb{C}(X_t, X_{t-1}) = \varepsilon^2\sin \varepsilon (t - 1)$ 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 explode en $\sqrt{t}$.

**Exercice 2** Soit $(\varepsilon_t)$ un bruit blanc et $(X_t)$ un processus vérifiant

$$X_t - \frac{7}{2} X_{t-1} + \frac{3}{2} X_{t-2} = \varepsilon_t$$

Pour tout $t \in \mathbb{Z}$

(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 $\varepsilon_t$ 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 $\Phi$. Soit $\tilde{X}_t$ est défini par

$$\tilde{X}_t = \Phi^{-1}(L) \varepsilon_t = \frac{1}{(1 - 3L)(1 - L/2)} \varepsilon_t = \left[\frac{6/5}{1 - 3L} - \frac{1/5}{1 - L/2}\right] \varepsilon_t$$

$$\tilde{X}_t = -\frac{2}{5} \sum_{k=1}^{\infty} \left(\frac{1}{3}\right)^{k-1} \varepsilon_{t+k} - \frac{1}{5} \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k \varepsilon_{t-k}$$

*Le processus $\varepsilon_t$ est alors corrélé avec $\tilde{X}_{t-1}, \tilde{X}_{t-2}, ...$ (puisque $\tilde{X}_t$ s’exprime en fonction du passé et du futur de $\varepsilon_t$) : le processus $\varepsilon_t$ 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,

$$\left(1 - \frac{1}{2}L\right) \left(1 - \frac{1}{3}L\right) \tilde{X}_t = \eta_t$$

soit $\tilde{X}_t = \frac{5}{6} \tilde{X}_{t-1} + \frac{1}{6} \tilde{X}_{t-2} = \eta_t$ où $\eta_t \sim BB (1, \sigma^2/9)$

Cette écriture correspondra à l’écriture canonique du processus $\tilde{X}_t$.

86
Exercise 3 On considère le processus aléatoire AR (2) suivant

\[ X_t = 40 + 0.4X_{t-1} - 0.2X_{t-2} + \varepsilon_t \] où \( \varepsilon_t \sim BB\left(0, \sigma^2 = 12.8\right) \)

(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 :

\[
1 - 0.4Z + 0.2Z^2 = 0 \text{ ssi } Z = \frac{0.4 \pm i\sqrt{0.64}}{0.4} \text{ c'est à dire } Z = \left\{ \begin{array}{l}
Z = 1 + 2i \\
Z = 1 - 2i
\end{array} \right.
\]

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.

\( \mathbb{E}(X_t) = 0.4 \mathbb{E}(X_{t-1}) - 0.2 \mathbb{E}(X_{t-2}) + 40 + 0 \) 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

\( \mathbb{E}(Y_tY_{t-h}) = 0.4\mathbb{E}(Y_{t-1}Y_{t-h}) - 0.2\mathbb{E}(Y_{t-2}Y_{t-h}) + \mathbb{E}(\varepsilon_tY_{t-h}) \)

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

\[
\left\{ \begin{array}{l}
\gamma_1 = 0.4\gamma_0 - 0.2\gamma_1 \\
\gamma_2 = 0.4\gamma_1 - 0.2\gamma_0
\end{array} \right.
\]

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 \( \gamma_0 \) vérifie

\[
\gamma_0 = \mathbb{E}(Y_1Y_1) = 0.4\mathbb{E}(Y_{t-1}Y_t) - 0.2\mathbb{E}(Y_{t-2}Y_t) + \mathbb{E}(\varepsilon_tY_t) = 0.4\gamma_1 - 0.2\gamma_2 + \mathbb{E}(\varepsilon_tY_t)
\]

Et comme \( \varepsilon_tY_t = 0.4\varepsilon_tY_{t-1} - 0.2\varepsilon_tY_{t-2} + \varepsilon_t^2 \), on en déduit que \( \mathbb{E}(\varepsilon_tY_t) = \sigma^2 \). Et donc \( \gamma_0 = 0.4\gamma_1 - 0.2\gamma_2 + \sigma^2 \). D'où finalement, 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

\[
\left\{ \begin{array}{l}
\gamma_1 = \gamma_0/3 = 5 \\
\gamma_2 = -0.2\gamma_0/3 = -1 \\
\gamma_3 = 0.4 \times -1 - 0.2 \times 5 = -1.4 \\
\gamma_4 = 0.4 \times -1.4 - 0.2 \times -1 = 0.36 \\
\gamma_5 = 0.4 \times -0.36 - 0.2 \times -1.4 = 0.136
\end{array} \right. \quad \text{et donc} \quad \left\{ \begin{array}{l}
\rho_1 = 0.333 \\
\rho_2 = 0.068 \\
\rho_3 = 0.093 \\
\rho_4 = 0.024 \\
\rho_5 = 0.009
\end{array} \right.
\]

(iv) Les autocorrelations partielles se calculent à l'aide des déterminants des matrices d’autocorrelations :

\[
\alpha(1) = \begin{vmatrix} \rho_1 \\ \rho_2 \end{vmatrix} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = -0.199
\]

\[
\alpha(1) = \begin{vmatrix} \rho_1 & \rho_2 \\ \rho_1 & \rho_1 \end{vmatrix} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \frac{1}{1 - \rho_1^2}
\]

\[\begin{cases} \phi_1 + \phi_2 < 1 \\ -\phi_1 + \phi_2 < 1 \\ \phi_2 > -1 \end{cases}\]

\(^{16}\)Ceci pouvait également se noter directement en notant que les conditions de stationnarité dans le cas d’un AR (2) s’écrivent...
Ce comportement de l’autocorrélogramme est caractéristique des processus mobile (à préciser). Donner la représentation canonique de
inférieur à
¯
partielles d’un processus
On peut noter que a (3) est très faible, ce qui confirme la nature AR(2) du modèle (en théorie, les autocorrélations
prenant au-delà du rang p).

**Exercise 4** Considérons deux processus \((X_t)\) et \((Y_t)\) liés par les relations suivantes
\[
\begin{align*}
Y_t &= \phi Y_{t-1} + \alpha X_t + \varepsilon_t \\
X_t &= \theta X_{t-1} + \eta_t
\end{align*}
\]

où \(\varepsilon_t\) et \(\eta_t\) sont des bruits blancs non-corrélés, de variance respective \(\sigma^2\) et \(\sigma^2\). Pour les applications pratiques, on prendra
\[
\alpha = 1.5, \phi = 0.4, \theta = 0.6, \sigma^2 = 0.016 et \sigma^2 = 0.036
\]

(i) on pose \(\omega_t = (1 - \phi L)(1 - \theta L)Y_t\). Calculer les moments de \(\omega_t\) et en déduire qu’il s’agit d’un processus moyen mobile (à préciser). Donner la représentation canonique de \(\omega_t\)

(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 \(\omega_t\) 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 \(\mathbb{E}(\omega_t) = 0\).

\[
(1 - \theta L)Y_t = Y_t - \theta Y_{t-2} = \phi Y_{t-1} + \alpha X_t + \varepsilon_t - \theta [\phi Y_{t-2} + \alpha X_{t-1} + \varepsilon_{t-1}]
\]

\[
= \phi [Y_{t-1} - \theta Y_{t-2}] + \alpha [X_t - \theta X_{t-1}] + \varepsilon_t - \theta \varepsilon_{t-1}
\]

\[
= \phi [Y_{t-1} - \theta Y_{t-2}] + \alpha [X_t - \theta X_{t-1}] + \alpha \eta_t + \varepsilon_t - \theta \varepsilon_{t-1} = \phi [1 - \theta L]Y_{t-1} + \alpha \eta_t + \varepsilon_t - \theta \varepsilon_{t-1}
\]

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 \(\omega_t\) vérifie alors,
\[
\begin{align*}
\gamma (0) &= \mathbb{E}(\omega_t^2) = \alpha^2 \sigma^2_\eta + (1 + \theta^2) \sigma^2 = 0.10276 \\
\gamma (1) &= \mathbb{E}(\omega_t \omega_{t-1}) = \theta \sigma^2 = -0.00966 \\
\gamma (h) &= 0 \text{ pour } h \geq 2
\end{align*}
\]

Ce comportement de l’autocorrélogramme est caractéristique des processus \(MA(1)\) (nullité des autocorrélations au delà du rang 1) : \(\omega_t\) est un processus moyen mobile.

Cherchons à écrire \(\omega_t\) sous forme \(MA(1)\) : \(\omega_t = u_t - \beta u_{t-1}\) où \(u_t\) est un bruit blanc de variance \(\sigma^2\). Les paramètres \(\beta\) et \(\sigma^2\) doivent alors vérifier
\[
\begin{align*}
\gamma (0) &= (1 + \beta^2) \sigma^2 \\
\gamma (1) &= -\beta \sigma^2 \\
\gamma (h) &= 0 \text{ pour } h \geq 2
\end{align*}
\]

c’est à dire que \(\beta\) doit vérifier
\[
-\frac{\beta}{1 + \beta^2} = \frac{\gamma (1)}{\gamma (0)} = \frac{-\beta \sigma^2}{\alpha^2 \sigma^2 + (1 + \theta^2) \sigma^2} = 0.0934
\]

\(\beta\) 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 (\(\beta\) étant l’inverse de la racine du polynôme caractéristique). \(\omega_t\) 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 \(\omega_t = (1 - \theta L)(1 - \phi L)Y_t = (1 - \theta L)u_t\) où \(u_t\) est un bruit blanc, c’est à dire que \(Y_t\) suit un processus ARMA(2,1).

La forme développée de la dynamique de \(Y_t\) est
\[
Y_t = (\phi + \theta) Y_{t-1} - \phi \theta Y_{t-2} + u_t - \beta u_{t-1}
\]

Il est également possible de calculer les autocorrélations \(\gamma (h)\) du processus : pour cela, on vamultiplier cette expression par \(X_{t-h}\), puis prendre l’espérance (le processus \(Y_t\) étant centré):

\[
Y_t^2 = (\phi + \theta) Y_{t-1} - \phi \theta Y_{t-2} + u_t - \beta u_{t-1} Y_t
\]

\[
= (\phi + \theta) Y_{t-1} Y_t - \phi \theta Y_{t-2} Y_{t-1} + \alpha \eta_t + \varepsilon_t - \theta \varepsilon_{t-1} + \alpha \eta_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}
\]

\[
= \phi [\phi + \theta] Y_{t-1} Y_t - \phi \theta Y_{t-2} Y_{t-1} + \alpha \eta_t + \varepsilon_t - \theta \varepsilon_{t-1} + \alpha \eta_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}
\]

\[\text{Ce qui implique que } \text{cov}(X_t, X_{t-h}) = \mathbb{E}(X_t X_{t-h}) - \mathbb{E}(X_t) \mathbb{E}(X_{t-h}) = \mathbb{E}(X_t X_{t-h}) \text{ pour tout } h. \text{ En particulier } V(X_t) = \mathbb{E}(X_t^2). \]

88
(on remplace $Y_t$ par (27) dans $u_t Y_t$ et $u_{t-1} Y_t$). En prenant l’espérance, on obtient

$$\gamma (0) = (i + \theta) \gamma (1) - \phi \theta \gamma (2) + \sigma^2 \left[ 1 - \beta (i + \theta) + \beta^2 \right]$$

De plus, en multipliant (27) par $Y_{t-1}$ on peut écrire

$$Y_{t} Y_{t-1} = (i + \theta) Y_{t-1} (Y_{t-1} - \beta u_{t-2} Y_{t-1} + u_{t-1} - \beta u_{t-1} - \beta u_{t-2})$$

$$(u_t est indépendant du passé de Y_t, en particulier indépendant de Y_{t-1}) d'où, en prenant l’espérance

$$\gamma (1) = (i + \theta) \gamma (0) - \phi \theta \gamma (1) - \beta \sigma^2$$

En multipliant (27) par $Y_{t-2}$ on a

$$\gamma (2) = (i + \theta) \gamma (1) - \phi \theta \gamma (0)$$

de façon plus générale

$$\gamma (h) = (i + \theta) \gamma (0) \gamma (h - 1) - \phi \theta \gamma (0)$$

**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

$$\Theta (L) = \sum_{k=0}^{+\infty} \theta_k L^k$$

et $\varepsilon_t$ bruit blanc de variance $\sigma^2$. Posons alors $\theta = \Theta (1)$.

2. Soit $\Theta^*$ tel que $\Theta (L) = \theta + (1 - L) \Theta^*(L)$. Donner l’expression des coefficients $\theta^*_k$ du polynôme $\Theta^*$.

3. Montrer que si $\sum k |\theta_k| < +\infty$ alors $\sum |\theta^*_k| < +\infty$.

4. 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.

5. Quelle est la limite de $V(Y_t)/t$ quand $t \to \infty$ ?

6. Considérons une décomposition quelconque du processus $Y_t$, $Y_t = \overline{T}_t + \overline{C}_t$ en une somme d’une marche aléatoire $\overline{T}_t$ et d’un processus stationnaire $\overline{C}_t$. Montrer qu’alors, on a nécessairement

$$V \left( \frac{(1 - L) \overline{T}_t}{L} \right) = \theta^2 \sigma^2$$

7. On supposera désormais que le processus $Y_t$ vérifie une équation de la forme

$$(1 - L) (1 - i \varepsilon_t) Y_t = \varepsilon_t \quad \text{où} \ 0 < \phi < 1 \quad \ (28)$$

8. Ecrire la relation précédente sous la forme $(1 - L) Y_t = \Theta (L) \varepsilon_t$ (Forme de Wold)

9. 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 ?

10. On peut 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.

11. En utilisant (28), calculer la densité spectrale de $X_t = (1 - L) Y_t$

12. 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 $X_t = (1 - L) Y_t$

13. Étudier les comportements pour $\omega = \theta$ des deux densités spectrales trouvées en (3.1) et en (3.2).

14. Conclure.

1. Par hypothèse, $(1 - L) Y_t = \Theta (L) \varepsilon_t$. Soit alors $X_t = (1 - L) Y_t$. Le polynôme $\Theta$ vérifie

$$\Theta (L) = \sum_{k=0}^{+\infty} \theta_k (L - 1) + \sum_{k=0}^{+\infty} \theta_k (L - 1) + \Theta (1)$$

Or l’aide $L^k - 1$ peut se réécrire

$$L^k - 1 = (L - 1) [1 + L + \ldots + L^{k-1}]$$
et donc
\[
\Theta (L) = \Theta (1) + (1 - L) \sum_{k=0}^{+\infty} \theta_k \left[ - (1 + L + \ldots + L^{k-1}) \right]
\]
\[
= \Theta (1) - (1 - L) \sum_{k=0}^{+\infty} \left( \sum_{j=k+1}^{+\infty} \theta_j \right) L^k
\]
D'où finalement l'écriture \( \Theta (L) = \theta + (1 - L) \Theta^* (L) \) où
\[
\Theta^* (L) = \sum_{k=0}^{+\infty} \theta_k^* L^k \text{ avec } \theta_k^* = - \sum_{j=k+1}^{+\infty} \theta_j
\]

(1 - ii) En utilisant cette expression,
\[
\sum_{k=0}^{+\infty} \theta_k^* = \sum_{k=0}^{+\infty} \sum_{j=k+1}^{+\infty} \theta_j = [\theta_1 + \theta_2 + \theta_3 + \ldots] + [\theta_2 + \theta_3 + \ldots] + [\theta_3 + \ldots] + \ldots = \sum_{k=0}^{+\infty} j \theta_j
\]
et
\[
\sum_{k=0}^{+\infty} |\theta_k^*| \leq \sum_{k=0}^{+\infty} k |\theta_k|
\]
Donc finalement 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 \( T_t \) défini par \( (1 - L) T_t = \theta \varepsilon_t \) est une marche stationnaire :
\[
T_t = T_{t-1} + \theta \varepsilon_t = \theta \sum_{k=0}^{t} \varepsilon_{t-k} + T_{t-1}
\]

(1 - iv) La variance de \( Y_t \) peut se décomposer en
\[
V (Y_t) = V (C_t) + 2cov (C_t, T_t) + V (T_t)
\]
avec \( V (C_t) \) = constante et \( V (T_t) \) est équivalent à \( \theta^2 \sigma_{\varepsilon}^2 t \), car \( T_t = T_0 + \theta [\varepsilon_1 + \ldots + \varepsilon_t] \). De plus,
\[
\text{cov} (T_t, C_t) \leq \sqrt{V (T_t)} \sqrt{V (C_t)} = O \left( \sqrt{t} \right)
\]
et finalement, on a
\[
\frac{V (Y_t)}{t} \to \theta^2 \sigma_{\varepsilon}^2 \text{ quand } t \to \infty
\]
En effectuant exactement le même genre de calcul sur \( Y_t = \overline{T_t} + \overline{C_t} \), p. a,
\[
V (Y_t) = V \left( \overline{C_t} \right) + 2cov \left( \overline{C_t}, \overline{T_t} \right) + V \left( \overline{T_t} \right)
\]
et on obtient alors
\[
\lim_{t \to \infty} \frac{V (Y_t)}{t} = \lim_{t \to \infty} \frac{V (\overline{T_t})}{t} = \theta^2 \sigma_{\varepsilon}^2
\]
D'où finalement de résultat souhaité, \( V(\Delta T_t) = \theta^2 \sigma^2. \)

**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

\[
f(\omega) = \frac{1}{2\pi} \left| \Theta(e^{i\omega}) \right|^2 \sigma^2 = \frac{\sigma^2}{2\pi} + \frac{1}{2\pi} \left| (1 - e^{i\omega}) B(e^{i\omega}) \right|^2 \sigma^2
\]

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

\[
\frac{\sigma^2}{2\pi} = \frac{\sigma^2}{2\pi} |\Theta(1)|^2 \quad \text{soit} \quad \frac{\sigma^2}{\sigma^2} = |\Theta(1)|^2
\]

Et comme \( \left| (1 - e^{i\omega}) B(e^{i\omega}) \right|^2 \sigma^2 \geq 0 \), alors \( f(\omega) \geq f(0) = \sigma^2/2\pi : f \) 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

\[
f(\omega) = \frac{\sigma^2}{2\pi} \left| 1 - \phi e^{i\omega} \right|^2 \quad \text{et} \quad f(0) = \frac{\sigma^2}{2\pi} \left( 1 - \phi^2 \right)
\]

Montrons qu'alors \( 0 \) n'est pas un minimum global : si \( \omega = \pi/2 \) alors \( \left| 1 - \phi e^{i\omega} \right|^2 = 1 + \phi^2 > (1 - \phi)^2 \) puisque \( 0 < \phi < 1 \), et donc

\[
\frac{1}{(1 - \phi)^2} > \frac{1}{1 + \phi^2} \quad \text{et donc} \quad f(0) > f \left( \frac{\pi}{2} \right)
\]

\((Y_t)\) 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**

**Exercise 6 (i) Soit \((\varepsilon_t)\) un bruit blanc et \( X_t \) défini par**

\[
X_t = \sum_{k=0}^{t} \lambda^k (\varepsilon_{t-k} - \varepsilon_{t-k-1}) \quad \text{où} \quad \lambda \in \mathbb{R}
\]

Le processus \( X_t \) est-il stationnaire ? (éventuellement avec certaines restrictions sur les paramètres)

(ii) Soit \((\varepsilon_t)\) 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

\[
X_t = \lambda^0 [\varepsilon_t - \varepsilon_{t-1}] + \lambda^1 [\varepsilon_{t-1} - \varepsilon_{t-2}] + \lambda^2 [\varepsilon_{t-2} - \varepsilon_{t-3}] + \ldots + \lambda^{t-1} [\varepsilon_{1} - \varepsilon_0] + \lambda^{t} [\varepsilon_0 - 0]
\]

avec la convention \( \varepsilon_{-1} = 0 \). Cette relation peut se réécrire, en mettant en facteur les \( \varepsilon_{t-i} \),

\[
X_t = \lambda^0 \varepsilon_t + [\lambda^1 - \lambda^0] \varepsilon_{t-1} + [\lambda^2 - \lambda^1] \varepsilon_{t-2} + [\lambda^3 - \lambda^2] \varepsilon_{t-3} + \ldots + [\lambda^{t-1} - \lambda^{t-2}] \varepsilon_1 + [\lambda^t - \lambda^{t-1}] \varepsilon_0
\]

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,

\[
V(X_t) = \sigma^2 + (\lambda - 1)^2 \sum_{k=0}^{t-1} \lambda^{2k} \sigma^2 = \sigma^2 + (\lambda - 1)^2 \frac{1 - \lambda^{2t}}{1 - \lambda^2} \sigma^2
\]

91
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 $E(Y_t) = 0$ (=constante), et $\gamma(0) = E(Y_t^2) = E(\varepsilon_t^2) = \sigma^2$ (=constante). Enfin, pour tout $h \neq 0$, $\gamma(h)$ vérifie

$$
\gamma(h) = E(Y_{t+h}Y_t) = E\left((-1)^{2t-h}\varepsilon_t\varepsilon_{t-h}\right) = \pm E(\varepsilon_t\varepsilon_{t-h}) = 0
$$

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 $E(Z_t) = 0$ (=constante), on peut noter que la variance, elle, ne sera pas constante : $E(Z_t^2) = 4\sigma^2$ pour $t$ pair, et $E(Z_t^2) = 0$ pour $t$ impair (car $Z_t$ 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 stationnaires, 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 $\varepsilon_t$ un bruit blanc, centré de variance $5/18$, et considérons le processus $Y_t$ défini par

$$
Y_t = 2\eta_{t-1} + \varepsilon_t
$$

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 $\eta_t$ est un bruit blanc non corrélaté avec $\varepsilon_t$, de variance $1/6$ (avec de plus $\text{cov}(\varepsilon_t, \eta_{t-h}) = 0$ pour tout $h \in Z$)

(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 préciserà,

(ii) donner la représentation canonique de $X_t$,

(iii) en déduire une représentation de $X_t$ du type

$$
X_t = -\sum_{i=1}^{\infty} \alpha_i X_{t-i} + u_t \quad \text{avec} \quad \sum_{i=1}^{\infty} |\alpha_i| < \infty
$$

et où $u_t$ est un bruit blanc que l’on explicitera (donner sa volatilité), non corrélaté 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 : $E(\omega_t) = 0 + 0 = 0 = 0$. Sa variance est donnée par

$$
\gamma(0) = E(\omega_t^2) = E\left((\varepsilon_t + \eta_t - 2\eta_{t-1})^2\right) = E(\varepsilon_t^2 + \eta_t^2 + 4\eta_{t-1}^2 + 2\varepsilon_t\eta_t - 4\eta_t\eta_{t-1} - 4\varepsilon_t\eta_{t-1}) = \sigma^2 + 5\sigma^2 + 0 + 0 = \sigma^2 + 5\sigma^2
$$

car $E(\varepsilon_t\eta_t) = E(\varepsilon_t\eta_{t-1}) = 0$ (les processus sont indépendants) et $E(\eta_t\eta_{t-1}) = 0$ (car $\eta_t$ est un bruit blanc). Donc $E(\omega_t^2) = \sigma^2 + 5\sigma^2$ qui est une constante : $\gamma(0) = 5/18 + 5/6 = 20/18 = 10/9$. De plus

$$
\gamma(1) = E(\omega_t\omega_{t-1}) = E\left((\varepsilon_t + \eta_t - 2\eta_{t-1})(\varepsilon_{t-1} + \eta_{t-1} - 2\eta_{t-2})\right) = E(\varepsilon_t\varepsilon_{t-1} + \varepsilon_t\eta_{t-1} - 2\varepsilon_t\eta_{t-2} + \eta_t\varepsilon_{t-1} + \eta_t\eta_{t-1} - 2\eta_t\eta_{t-2} - 2\eta_{t-1}\varepsilon_{t-2} - 4\eta_t\eta_{t-1} - 4\varepsilon_t\eta_{t-1} - 4\eta_t\eta_{t-2}) = 0 + 0 + 0 + 0 - 2\sigma^2 + 0 = -2\sigma^2
$$

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 $\omega_t$ vérifie

$$
\rho(1) = \frac{\gamma(1)}{\gamma(0)} = -\frac{9}{310} = -\frac{3}{10} \quad \text{et} \quad \rho(h) = 0 \text{ pour } h \geq 2
$$

Le processus $\omega_t$ est un processus $MA(1)$. Aussi, il existe un bruit blanc $\nu_t$ et un paramètre $\theta$ tel que

$$
\omega_t = \nu_t - \theta\nu_{t-1}
$$
Le processus \( X_t \) vérifie
\[
X_t = Y_t + \eta_t = [2Y_{t-1} + \varepsilon_t] + \eta_t = 2[X_{t-1} - \eta_{t-1}] + \varepsilon_t + \eta_t
\]
\[
= 2X_{t-1} + [\varepsilon_t + \eta_t - 2\eta_{t-1}] = 2X_{t-1} + \omega_t
\]
donc, d’après la question précédente, il existe bruit blanc \( \nu_t \) et un paramètre \( \theta \) 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 : \( \theta \) (composante moyenne-mobile) et \( \sigma^2_\nu \), 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
\[
\left\{ \begin{array}{l}
\gamma(0) = \mathbb{E}(\omega^2_t) = [1 + \theta^2]\sigma^2_\nu = 10/9 \\
\gamma(1) = \mathbb{E} (\omega_t\omega_{t-1}) = -\theta\sigma^2_\nu = -1/3
\end{array} \right.
\]
En divisant la première équation par la seconde, on obtient que
\[
\frac{1 + \theta^2}{\theta} = \frac{10/3}{1/3} = \frac{10}{3}
\]
ce qui donne l’équation de degré 2 suivante, \( 3\theta^2 - 10\theta + 3 = 0 \), dont les racines sont
\[
\theta = \frac{10 \pm \sqrt{10^2 - 4 \times 3 \times 3}}{2 \times 3} = \frac{10 \pm \sqrt{64}}{6} = \frac{10 \pm 8}{6} = \left\{ \begin{array}{l}
3/3 = 1 \\
3/3 = 1/3
\end{array} \right.
\]
Pour mettre le processus sous forme canonique on choisit \( \theta \) 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 : \( -\theta\sigma^2_\nu = -1/3 \) donc \( \sigma^2_\nu = 1 \).
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 avions choisi la racine de \( \Theta \) à l’extérieur du disque unité, le polynôme \( \Theta(L) \) est inverible en fonction des opérateurs passés, et
\[
\Theta^{-1}(L) \Phi(L) = \frac{1}{1 - \theta L} (1 + \phi L) = (1 + \phi L) \sum_{i=0}^{\infty} \theta^i L^i = 1 + \sum_{i=1}^{\infty} \left[ \frac{\theta^{i-1} \phi + \theta^i}{\sigma_\nu} \right] L^i
\]
Avec cette écriture, on obtient finalement
\[
X_t = -\sum_{i=1}^{\infty} \alpha_i X_{t-i} + \nu_t \text{ avec } \left\{ \begin{array}{l}
\alpha_i = \theta^{i-1} \phi + \theta^i = 2/3^{i-1} + 1/3^i \\
\nu_t \sim BB(0, 1)
\end{array} \right.
\]
Cette écriture est utile pour faire de la prévision : la meilleure prévision possible, faite à la date \( T \) pour un horizon \( h \) est
\[
\hat{X}_{T+h} = -\sum_{k=1}^{h-1} \alpha_k X_{T+h-k} - \sum_{k=h}^{T+h} \alpha_k X_{T+h-k} = T_{\hat{X}}_{T+h} = \sum_{k=1}^{h-1} \alpha_k T_{X_{T+h-k}}
\]

Remarque 22 L’exercice suivant utilisait les sorties informatiques disponibles sur http://www.crest.fr/pageperso/lfa/cha

Exercise 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 \( Z_t = \Delta^2 X_t = \Delta Y_t = X_t - 2X_{t-1} + X_{t-2} \)
- \( Y_{2t} \) la série \( Y_{2t} = Y_t - Y_{t-2} \) et \( Z_{2t} \) la série \( Z_{2t} = Z_t - Z_{t-2} \)
- \( Y_{4t} \) la série \( Y_{4t} = Y_t - Y_t \) et \( Z_{4t} \) la série \( Z_{4t} = Z_t - Z_{t-4} \)
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 (\(Y_t\)) 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. \(Y_2\) n'est pas stationnaire. En revanche la série \(Y^{(4)}\) 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 \(Z^{(4)}\), que la seule série stationnaire. Or \(Y^{(4)} = (1 - L) (1 - L^4) X_t\) et \(Z^{(4)} = (1 - L)^2 (1 - L^3) X_t = (1 - L) Y^{(4)}\) : puisque \(Y^{(4)}\) est déjà stationnaire, il est inutile de différencier davantage. Nous allons modéliser \(Y^{(4)} = (1 - L) (1 - L^4) X_t\).

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 significatif, 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 non-stationary') : 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 \(\alpha\) 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 pas être retenu.

L'estimation du modèle (32) - modèle \(AR(2)\) pour \(Y_2\) - 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 \(\alpha\) 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 \(Y^{(4)}\) - 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 \(\theta\) 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 $Y_t$ (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 $Y_t$ correspond au modèle (33) : ce modèle peut être retenu.

Le modèle $ARMA (4, 0)$ appliqué à $Y_t$ 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 $Y_t$ 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 $Y_t$ : tous les coefficients sont là aussi significatifs. Et si l’on considère l’autocorrélogramme des résidus obtenus après régression (slide (111)), on peut noter qu’aucune autocorrélation 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 $Y_t$ 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 $AR (4)$ appliqué à $Y_t$ (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 $(1 - L)(1 - L^4)(1 - 0.8L^2)X_t = (1 - 0.6L^2)\varepsilon_t$ est à dire $Y_t$, modélisé 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

\[\begin{array}{llllll}
\text{[1]} & \text{slide (96)} & (1 - L)(1 - L^4)X_t = (1 - \beta L)\eta_t \\
\text{[2]} & \text{slide (111)} & (1 - L)(1 - L^4)(1 - \alpha L^2)X_t = (1 - \beta_1 L - \beta_2 L^2 - \beta_3 L^3)\varepsilon_t \\
\text{[3]} & \text{slide (106)} & (1 - L)(1 - L^4)(1 - \alpha_1 L - \alpha_2 L^2 - \alpha_3 L^3 - \alpha_4 L^4)X_t = u_t \\
\text{[4]} & \text{slide (88)} & (1 - L)(1 - L^4)(1 - \alpha L)X_t = (1 - \beta L^2)v_t \\
\text{[5]} & \text{slide (131)} & (1 - L)(1 - L^4)(1 - \alpha_2 L^2 - \alpha_3 L^4)X_t = (1 - \beta_1 L - \beta_2 L^2 - \beta_3 L^3)\varepsilon_t \\
\text{[6]} & \text{slide (123)} & (1 - L)(1 - L^4)(1 - \alpha L^2)X_t = (1 - \beta_1 L - \beta_2 L^2 - \beta_3 L^3 - \beta_4 L^4)\varepsilon_t
\end{array}\]

où $\eta_t, \varepsilon_t, u_t$ et $v_t$ sont des bruit blancs.

Les indicateurs de choix de modèles donnent les résultats suivants :

\[
\begin{array}{lllllll}
R^2 & \overline{R^2} & \log L & \text{Akaike} & \text{Schwarz} & F\text{-stat} \\
\text{[1]} & 0.2893 & 0.2893 & -101.6273 & -0.10115 & -0.07025 \\
\text{[2]} & 0.3322 & 0.3032 & -94.7231 & -0.13313 & -0.00763 & 11.44084 \\
\text{[3]} & 0.3487 & 0.3196 & -91.8485 & -0.13792 & -0.01044 & 11.95692 \\
\text{[4]} & 0.3273 & 0.3179 & -97.4337 & -0.15038 & -0.08811 & 35.02662 \\
\text{[5]} & 0.9920 & 0.9916 & -92.8359 & -0.15744 & -0.00056 & 2117.092 \\
\text{[6]} & 0.9913 & 0.9910 & -99.1440 & -0.08737 & 0.036229 & 2709.168
\end{array}
\]


### 4.3 Examen de 2002/2003

**Exercice 9** Le processus $MA (2)$ suivant est-il stationnaire (au second ordre) ?

\[Y_t = [1 - 2.4L + 0.8L^2]\varepsilon_t\text{ où }\mathbb{E}(\varepsilon_t) = 0\]

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

\[
\gamma (h) = \mathbb{E}(X_t X_{t-h})
\]

\[
= \mathbb{E}(|\varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}| \varepsilon_{t-h} + \theta_1 \varepsilon_{t-h-1} + \ldots + \theta_q \varepsilon_{t-h-q})
\]

\[
= \begin{cases} 
\theta_h + \theta_{h+1} \varepsilon_t + \ldots + \theta_q \varepsilon_{t-h} | \sigma^2 & \text{si } 1 \leq h \leq q \\
0 & \text{si } h > q
\end{cases}
\]

95
avec, pour $h = 0$, la relation

$$
\gamma (0) = [1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_h^2] \sigma^2
$$

Dans le cas d’un processus $MA(2)$,

$$
\gamma (h) = \begin{cases}
[1 + \theta_1^2 + \theta_2^2] \sigma^2 & \text{pour } h = 0 \\
\theta_1 [1 + \theta_2^2] \sigma^2 & \text{pour } h = 1 \\
\theta_2 \sigma^2 & \text{pour } h = 2 \\
0 & \text{pour } h \geq 3
\end{cases}
$$

D’où la fonction d’autocovariance dans le cas particulier $Y_t = [1 - 2.4L + 0.8L^2] \varepsilon_t$ où $V (\varepsilon_t) = 1$,

$$
\gamma (0) = 1 + 2.4^2 + 0.8^2 = 7.4 \\
\gamma (1) = 2.4 - 8 \times 2.4 = 0.48 \\
\gamma (2) = 0.8 \\
\gamma (h) = 0 & \text{pour } h \geq 3
$$

**Exercise 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 $(\varepsilon_t)$ 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 correspond au cas où l’autocorrélation est négative : si $X_{t-1}$ est positif, $X_t$ semble négatif (en moyenne), et réciproquement, ce qui correspond à un cas où $corr (Z_{t-1}, Z_t) \leq 0$. Le comportement semble le contraire dans le cas de gauche : $corr (Z_{t-1}, Z_t) \geq 0$. La série de gauche correspond au cas $\rho = 0.85$ et le cas de droite $\rho = -0.7$.

**Exercise 11** (i) Dans le cas d’un processus stationnaire $AR(2)$, $X_t = \alpha X_{t-1} + \beta X_{t-2} + \varepsilon_t$ où $(\varepsilon_t)$ est un bruit blanc, rappeler les valeurs des $\rho (1)$ et $\rho (2)$ en fonction de $\alpha$ et $\beta$.

(ii) Lors de la modélisation d’une série d’observations $X_1, \ldots, X_T$, l’autocorrélogramme suivant a été obtenu,

<table>
<thead>
<tr>
<th>Autocorr.</th>
<th>Partial Corr.</th>
<th>AC</th>
<th>PAC</th>
<th>Q-Stat</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.60555</td>
<td>-0.60555</td>
<td>1</td>
<td>1834.56</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.68580</td>
<td>0.50387</td>
<td>2</td>
<td>4188.07</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>-0.51135</td>
<td>-0.00083</td>
<td>3</td>
<td>5497.80</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.498081</td>
<td>0.00121</td>
<td>4</td>
<td>6739.72</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>-0.40768</td>
<td>-0.00091</td>
<td>5</td>
<td>7571.90</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

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 \( (\varepsilon_t) \) est un bruit blanc, les deux premières équations de Yule-Walker sont

\[
\begin{bmatrix}
\rho(1) \\
\rho(2)
\end{bmatrix} =
\begin{bmatrix}
1 & \rho(1) \\
\rho(2) & 1
\end{bmatrix}
\begin{bmatrix}
\alpha \\
\beta
\end{bmatrix},
\]
c'est à dire

\[
\begin{align*}
\rho(1) &= \alpha + \beta \rho(1) \\
\rho(2) &= \alpha \rho(1) + \beta
\end{align*}
\]
soit

\[
\begin{align*}
\rho(1) &= \frac{\alpha}{1 - \beta} \\
\rho(2) &= \frac{\alpha^2}{1 - \beta}
\end{align*}
\]

(ii) L'équation de Yule-Walker peut s'inverser, de la façon suivante

\[
\begin{bmatrix}
\alpha \\
\beta
\end{bmatrix} =
\begin{bmatrix}
1 & \rho(1) \\
\rho(2) & 1
\end{bmatrix}^{-1}
\begin{bmatrix}
\rho(1) \\
\rho(2)
\end{bmatrix}
= \frac{1}{1 - \rho(1)^2}
\begin{bmatrix}
1 & -\rho(1) \\
-\rho(1) & 1
\end{bmatrix}
\begin{bmatrix}
\rho(1) \\
\rho(2)
\end{bmatrix},
\]

ce qui donne, respectivement

\[
\alpha = \frac{\rho(1)[1 - \rho(2)]}{1 - \rho(1)^2} \quad \text{et} \quad \beta = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2},
\]
dont des estimateurs naturels sont

\[
\hat{\alpha} = \frac{\hat{\rho}(1)[1 - \hat{\rho}(2)]}{1 - \hat{\rho}(1)^2} \quad \text{et} \quad \hat{\beta} = \frac{\hat{\rho}(2) - \hat{\rho}(1)^2}{1 - \hat{\rho}(1)^2},
\]

AN : dans l'exemple considéré, \( \hat{\rho}(1) \approx -0.60555 \) et \( \hat{\rho}(1) \approx 0.68580 \), et donc

\[
\hat{\alpha} \approx \frac{-0.60555[1 - 0.68580]}{1 - 0.60555^2} \approx -0.3 \quad \text{et} \quad \hat{\beta} \approx \frac{0.68580 - 0.60555^2}{1 - 0.60555^2} \approx 0.5.
\]

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): 45000
Included observations: 4997 after adjusting endpoints
Convergence achieved after 7 iterations

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>-0.148003</td>
<td>1.648750</td>
<td>-0.09009</td>
<td>0.9283</td>
</tr>
<tr>
<td>AR(2)</td>
<td>0.549109</td>
<td>0.494613</td>
<td>1.110179</td>
<td>0.2670</td>
</tr>
<tr>
<td>AR(3)</td>
<td>-0.077412</td>
<td>0.821452</td>
<td>-0.093105</td>
<td>0.9258</td>
</tr>
<tr>
<td>MA(1)</td>
<td>-0.150609</td>
<td>1.649037</td>
<td>-0.091332</td>
<td>0.9272</td>
</tr>
</tbody>
</table>

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 .152 .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 ci-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.

En particulier, il pourra être intéressant de commenter l’hypothèse de stationnarité de la série à l’aide de plusieurs arguments.

Les autocorrélations 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), \((1 - \phi L)(1 - \theta L) X_t = (1 - \alpha L)(1 - \beta L) \epsilon_t\), on peut noter que \(\hat{\alpha} = 0.99\) : le polynôme 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,
Une des conditions importantes lors d’une modélisation ARMA est que les résidus \( \varepsilon_t \) 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 \( \rho \neq 1 \).

**Exercise 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],

\[
\begin{align*}
\text{[1]} \quad X_t &= \varepsilon_t - \alpha_1 \varepsilon_{t-1} - \alpha_2 \varepsilon_{t-2} - \alpha_3 \varepsilon_{t-3} \\
\text{[2]} \quad X_t &= \eta_t + \beta_1 \eta_{t-1} + \beta_2 \eta_{t-2} + \beta_3 \eta_{t-3}
\end{align*}
\]

où \( \varepsilon_t \) et \( \eta_t \) sont des bruits blancs.

La sortie obtenue lors de l’estimation des paramètres est la suivante, ainsi que l’autocorrélogramme de \( X_t \).

---

**Conditional Least Squares Estimation**

| Parameter | Std Estim. | Std Error | t Value | Pr > |t| |
|-----------|------------|-----------|---------|-------|
| MA(1)     | -0.642247  | 0.01133   | -56.64595 | 0.0000 |
| MA(2)     | 0.487406   | 0.013856  | 35.17769 | 0.0000 |
| MA(3)     | -0.794482  | 0.010962  | -72.47537 | 0.0000 |

---

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

\[
\begin{align*}
\text{[1]} \quad X_t &= \varepsilon_t + 0.6422 \varepsilon_{t-1} - 0.4872 \varepsilon_{t-2} + 0.7944 \varepsilon_{t-3} \quad \text{où } (\varepsilon_t) \text{ est un bruit blanc} \\
\text{[2]} \quad X_t &= \varepsilon_t - 0.6422 \varepsilon_{t-1} + 0.4872 \varepsilon_{t-2} - 0.7944 \varepsilon_{t-3} \quad \text{où } (\varepsilon_t) \text{ est un bruit blanc}
\end{align*}
\]

Dans le cas d’un processus \( MA(q) \), écrit sous la forme [1], l’autocovariance à l’ordre \( q \) est donnée par

\[
\gamma(q) = \text{cov}(X_t, X_{t-q}) = \mathbb{E}(X_tX_{t-q}) = \mathbb{E}(\varepsilon_t \varepsilon_{t-q}) = \text{E}(\theta_q \varepsilon_{t-q} \varepsilon_{t-q} - \theta_q \varepsilon_{t-q} - \theta_q \varepsilon_{t-q} - \theta_q \varepsilon_{t-q} - \theta_q \varepsilon_{t-q}) = -\theta_q^2 \sigma^2 \text{ car } \mathbb{E}(\varepsilon_t \varepsilon_{t-j}) = \sigma^2 \delta(i = j)
\]

L’autocorrélation étant du même signe que l’autocovariance, alors \( \rho(q) \) est toujours du signe opposé à \( \theta_q \) 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

\[
\begin{align*}
\text{[2]} \quad X_t &= \varepsilon_t - 0.6422 \varepsilon_{t-1} + 0.4872 \varepsilon_{t-2} - 0.7944 \varepsilon_{t-3} \quad \text{où } (\varepsilon_t) \text{ est un bruit blanc}
\end{align*}
\]

**Exercise 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
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 test des modèles de la forme suivante, [1] \( X_t = aX_{t-1} + \epsilon_t \) (sans constante, testé en haut) ou [2] \( X_t = \mu + aX_{t-1} + \epsilon_t \) (avec constante, en bas), par exemple.

Le processus \( \epsilon_t \) 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] \( \mathbb{E}(X_t) = a\mathbb{E}(X_0) = 0 \), donc si \( a \) est différent de 1, \( \mathbb{E}(X_t) \neq 0 \),
- [2] \( \mathbb{E}(X_t) = \mu + a\mathbb{E}(X_0) + 0 \), donc si \( a \) est différent de 1, \( \mathbb{E}(X_t) = \mu/(1-a) \).

Le graphique ci-dessous 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 base, 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

**Exercise 17** On considère \( (Y_t) \) un processus vérifiant

\[
Y_t = X_t + \epsilon_t, \tag{34}
\]

où \( \epsilon_t \) est un bruit blanc, et où \( (X_t) \) vérifie

\[
X_t = \phi X_{t-1} + \eta_t + \theta \eta_{t-1}, \tag{35}
\]

où \( \eta_t \) est un bruit blanc tel que \( \mathbb{E}(\epsilon_t \eta_s) = 0 \) 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’écrit
\[
Y_t = X_t + \varepsilon_t = [\phi X_{t-1} + \eta_t + \theta \eta_{t-1}] + \varepsilon_t
\]
\[
= \phi [Y_{t-1} - \varepsilon_{t-1}] + \eta_t + \theta \eta_{t-1} \text{ car } Y_{t-1} = X_{t-1} + \varepsilon_{t-1},
\]
\[
= \phi Y_{t-1} + \varepsilon_t - \phi \varepsilon_{t-1} + \eta_t + \theta \eta_{t-1} = \phi Y_{t-1} + u_t.
\]

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 \(E(u_t) = 0\) (car \(E(\varepsilon_t) = E(\eta_t) = 0)\),

\[
\mathbb{E}\left(u_t^2\right) = \mathbb{E}\left(\left(\varepsilon_t - \phi \varepsilon_{t-1} + \eta_t + \theta \eta_{t-1}\right)^2\right)
\]
\[
= \mathbb{E}\left(\left[\varepsilon_t - \phi \varepsilon_{t-1} + \eta_t + \theta \eta_{t-1}\right] \left[\varepsilon_t - \phi \varepsilon_{t-1} + \eta_t + \theta \eta_{t-1}\right]\right)
\]
\[
= \mathbb{E}\left(\varepsilon_t^2 + \phi^2 \varepsilon_{t-1}^2 + \eta_t^2 + \theta^2 \eta_{t-1}^2 - 2\phi \varepsilon_t \varepsilon_{t-1} + 2\varepsilon_t \eta_t + 2\theta \varepsilon_t \eta_{t-1} - 2\phi \varepsilon_{t-1} \eta_t - 2\theta \varepsilon_{t-1} \eta_{t-1} + 2\theta \eta_t \eta_{t-1}\right).
\]

En utilisant la linéarité de l’espérance, et en utilisant le fait que \(E(\varepsilon_{t-1}) = 0\) ((\(\varepsilon_t\) est un bruit blanc) \(E(\eta_{t-1}) = 0\), et
\[
E(\varepsilon_t \eta_t) = E(\eta_{t-1} \varepsilon_{t-1}) = E(\eta_t \varepsilon_{t-1}) = E(\eta_{t-1} \varepsilon_{t-1}) = 0,
\]
(par hypothèse \(E(\varepsilon_t \eta_t) = 0\) pour tout \(s, t\), cette expression se simplifie en
\[
E(u_t^2) = E(\varepsilon_t^2) + \phi^2 E(\varepsilon_{t-1}^2) + E(\eta_t^2) + \theta^2 E(\eta_{t-1}^2) = (1 + \phi^2) \sigma_\varepsilon^2 + (1 + \theta^2) \sigma_\eta^2,
\]
en notant \(\sigma_\varepsilon^2\) et \(\sigma_\eta^2\), respectivement, les variances des deux bruits blancs (et on détaillant outargisement les calculs...). De même
\[
E(u_t u_{t-1}) = E(\phi \varepsilon_t \varepsilon_{t-1} + \eta_t \eta_{t-1} - \theta \phi \varepsilon_t \eta_{t-1} - \theta \varepsilon_t \eta_{t-1} + \theta \varepsilon_{t-1} \eta_t - \theta \phi \varepsilon_{t-1} \eta_t - \theta \eta_{t-1} \eta_t + \theta \eta_{t-1} \eta_t)
\]
\[
= E(\phi \varepsilon_t \varepsilon_{t-1} + \varepsilon_t \varepsilon_{t-1} + \varepsilon_t \eta_{t-1} + \varepsilon_t \eta_{t-1} - \phi \varepsilon_t \varepsilon_{t-1} - \phi \varepsilon_{t-1} \varepsilon_t - \phi \eta_t \eta_{t-1} + \phi \eta_{t-1} \eta_t)
\]
\[
= E(\phi^{2} \varepsilon_t \varepsilon_{t-1} + \varepsilon_t \varepsilon_{t-1} + \varepsilon_t \eta_{t-1} + \theta \varepsilon_t \varepsilon_{t-1} - \phi \varepsilon_t \varepsilon_{t-1} - \phi \varepsilon_{t-1} \varepsilon_t - \phi \eta_t \eta_{t-1} + \theta \varepsilon_t \varepsilon_{t-1} - \phi \varepsilon_t \varepsilon_{t-1} - \phi \eta_{t-1} \eta_t).
\]

En utilisant encore une fois la linéarité de l’espérance, et le fait que \(E(\varepsilon_t \varepsilon_t) = E(\eta_t \eta_t) = 0\) pour \(s \neq t\) (hypothèse de bruit blanc), et que, par hypothèse, \(E(\varepsilon_t \eta_t) = 0\) pour tout \(s, t\) (mêmes égales), on note qu’il ne reste plus que deux termes non nuls,
\[
E(u_t u_{t-1}) = -\phi E(\varepsilon_t \eta_{t-1}) + \theta E(\eta_t \varepsilon_{t-1}) = \theta \sigma_\eta^2 - \phi \sigma_\varepsilon^2.
\]
Et de façon plus général, un calcul identique à celui détaillé ci-dessus donne
\[
E(u_t u_{t-h}) = 0 \text{ pour } h \geq 2.
\]

On en déduit, tout d’abord que, pour tout \(h\), et pour tout \(t\), \(E(u_t u_{t-h})\) 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 \(\rho_h(h) = E(u_t u_{t-h})/E(u_t^2) = 0\) 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 “\(\theta = -\phi\)”, 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’écritrait \(\theta = \phi\) 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
\[
u_t = v_t + \alpha v_{t-1} \text{ où } \alpha \text{ est solution de } \frac{\alpha}{1 + \alpha^2} = \frac{\theta \sigma_\eta^2 - \phi \sigma_\varepsilon^2}{(1 + \phi^2) \sigma_\varepsilon^2 + (1 + \theta^2) \sigma_\eta^2}
\]
(cce 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
\[Y_t = \phi Y_{t-1} + \varepsilon_t + \alpha \varepsilon_{t-1},\]
ou \((\varepsilon_t)\) 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
\[
\rho (h) = \phi^h \rho (1) \text{ pour } h \geq 2.
\]

**Exercise 18** *Montrer que les deux processus MA \((1)\) suivant ont la même fonction d’autocorrélations,
\[X_t = \varepsilon_t + \theta \varepsilon_{t-1} \text{ et } Y_t = \eta_t + \frac{1}{\theta} \eta_{t-1},\]  
\text{ où } (\varepsilon_t) \text{ et } (\eta_t) \text{ sont des bruits blancs au sens } L^2, \text{ 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 \((X_t)\) sont données par \((X_t \text{ étant centré, i.e. } \E X_t = 0)\)
\[
\rho_X (1) = \frac{\E (X_t X_{t-1})}{\E (X_t^2)} = \frac{\theta \E (\varepsilon_t^2)}{(1 + \theta^2) \E (\varepsilon_t^2)} = \frac{\theta}{1 + \theta^2},
\]
et
\[
\rho_X (h) = \frac{\E (X_t X_{t-h})}{\E (X_t^2)} = \frac{0}{(1 + \theta^2) \E (\varepsilon_t^2)} = 0, \text{ pour } h \geq 2.
\]

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 \(\kappa = 1/\theta\)
\[
\rho_Y (1) = \frac{\E (Y_t Y_{t-1})}{\E (Y_t^2)} = \frac{\kappa \E (\varepsilon_t^2)}{(1 + \kappa^2) \E (\varepsilon_t^2)} = \frac{\kappa}{1 + \kappa^2} = \frac{1/\theta}{1 + 1/\theta^2} = \frac{\theta}{\theta^2 + 1},
\]
(en multipliant en haut et en bas par \(\theta^2\)) et
\[
\rho_Y (h) = \frac{\E (Y_t Y_{t-h})}{\E (Y_t^2)} = \frac{0}{(1 + \kappa^2) \E (\eta_t^2)} = 0, \text{ pour } h \geq 2.
\]
C’est à dire que pour tout \(h, \rho_Y (h) = \rho_X (h) : \text{les deux processus ont la même fonction d’autocorrélation.}

**Exercise 19** 1) *Supposons que \((Y_t)\) soit un processus AR \((1)\) stationnaire, et que \((X_t)\) soit un processus MA \((1)\), soit
\[Y_t = \phi Y_{t-1} + \varepsilon_t \text{ et } X_t = \eta_t - \theta \eta_{t-1},\]
où \((\varepsilon_t)\) et \((\eta_t)\) 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’écrit
\[(1 - L)^d \Phi (L) X_t = \Theta (L) \varepsilon_t, \text{ où } (\varepsilon_t) \text{ est un bruit blanc. Malheureusement, ce processus n’est pas observable directement, et l’on ne peut observer que } X^*_t = X_t + Y_t \text{ où } (Y_t) \text{ est une erreur d’observation (erreur de mesure ou bruit exogène) supposé suivre un processus ARMA } (\alpha, \beta), \text{ dont le bruit associé } (\eta_t) \text{ est supposé indépendant de } (\varepsilon_t). \]

*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 \(\phi X_{t-1}\)),
\[
Z_t = \eta_t - \theta \eta_{t-1} + \phi Y_{t-1} + \varepsilon_t + \phi X_{t-1} - \phi X_{t-1} = \phi Z_{t-1} + [\theta \eta_{t-1} + \varepsilon_t - \phi X_{t-1}].
\]
or $X_{t-1} = \eta_{t-1} - \theta \eta_{t-2}$, donc, par substitution,

$$Z_t = \phi Z_{t-1} + \left[ -\theta \eta_{t-1} + \epsilon_t - \phi \left( \eta_{t-1} - \theta \eta_{t-2} \right) \right] = \phi Z_{t-1} + u_t,$$

où $u_t = \epsilon_t - (\theta + \phi) \eta_{t-1} + \phi \theta \eta_{t-2}$. 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 $ARIMA(1,1)$, montrons que $(u_t)$ est un processus $MA(1)$. En notant que $(u_t)$ est centré, on en déduit sa fonction d'autocovariance,

$$\mathbb{E} (u_t^2) = \mathbb{E} \left( \left( \epsilon_t - \phi \epsilon_{t-1} + \eta_t + \theta \eta_{t-1} \right)^2 \right) = \mathbb{E} \left( \left( \epsilon_t - \phi \epsilon_{t-1} + \eta_t + \theta \eta_{t-1} \right) \left( \epsilon_t - \phi \epsilon_{t-1} + \eta_t + \theta \eta_{t-1} \right) \right) = \mathbb{E} (\epsilon_t^2) + \phi^2 \mathbb{E} (\epsilon_{t-1}^2) + 2 \epsilon_t \eta_t + 2 \epsilon_t \eta_{t-1} - 2 \phi \epsilon_{t-1} \eta_t - 2 \phi \epsilon_{t-1} \eta_{t-1} + 2 \theta \epsilon_t \eta_t - 2 \phi \epsilon_{t-1} \eta_{t-1} + 2 \theta \eta_t \eta_{t-1} - 2 \phi \epsilon_{t-1} \eta_{t-1}. \quad \text{En utilisant la linéarité de l'espérance, et en utilisant le fait que} \quad \mathbb{E} (\epsilon_t \eta_t) = \mathbb{E} (\epsilon_t \eta_{t-1}) = \mathbb{E} (\epsilon_t \eta_t) = 0, \quad \text{et} \quad \mathbb{E} (\epsilon_t \eta_{t-1}) = \mathbb{E} (\epsilon_t \eta_{t-1}) = \mathbb{E} (\epsilon_t \eta_{t-1}) = 0, \quad \text{(par hypothèse } \mathbb{E} (\epsilon_t \eta_t) = 0 \text{ pour tout } s, t, \text{ cette expression se simplifie en} \quad \mathbb{E} (u_t^2) = \mathbb{E} (\epsilon_t^2) + \phi^2 \mathbb{E} (\epsilon_{t-1}^2) + \theta^2 \mathbb{E} (\eta_t^2) + (1 + \phi^2) \sigma_\epsilon^2 + (1 + \theta^2) \sigma_\eta^2, \quad \text{en notant } \sigma_\epsilon^2 \text{ et } \sigma_\eta^2, \text{ respectivement, les variances des deux bruits blancs (et en détaillant outragement les calculs...). \quad \text{De même} \quad \mathbb{E} (u_t u_{t-1}) = \mathbb{E} \left( \left( \epsilon_t - \phi \epsilon_{t-1} + \eta_t + \theta \eta_{t-1} \right) \left( \epsilon_{t-1} - \phi \epsilon_{t-2} + \eta_{t-1} + \theta \eta_{t-2} \right) \right) = \mathbb{E} (\epsilon_t \epsilon_{t-1} - \phi \epsilon_t \epsilon_{t-2} + \epsilon_t \eta_{t-1} - \phi \epsilon_t \eta_{t-2} + \epsilon_{t-1} \eta_t - \phi \epsilon_{t-1} \eta_{t-2} + \eta_t \eta_{t-1} - \phi \eta_t \eta_{t-2} + \theta \epsilon_{t-1} \eta_{t-1} - \phi \eta_{t-1} \eta_{t-2} + \theta \eta_{t-1} \eta_{t-2}). \quad \text{En utilisant encore une fois la linéarité de l'espérance, et le fait que} \quad \mathbb{E} (\epsilon_t \epsilon_{t-1}) = \mathbb{E} (\eta_t \eta_{t-1}) = 0 \text{ pour } s \neq t \text{ (hypothèse de bruit blanc), et que, par hypothèse, } \mathbb{E} (\epsilon_t \eta_t) = 0 \text{ pour tout } s, t \text{ (même égaux), on note qu'il ne reste plus que deux termes non nuls,} \quad \mathbb{E} (u_t u_{t-1}) = -\phi \mathbb{E} (\epsilon_t^2) + \theta \mathbb{E} (\eta_t^2) = \theta \sigma_\eta^2 - \phi \sigma_\epsilon^2. \quad \text{Et de façon plus général, un calcul identique à celui détaillé ci-dessus donne} \quad \mathbb{E} (u_t u_{t-1}) = 0 \text{ pour } h \geq 2. \quad \text{On en déduit, tout d'abord que, pour tout } h, \text{ et pour tout } t, \mathbb{E} (u_t u_{t-h}) \text{ ne dépend pas de } t : \text{ le processus } (u_t) \text{ est effectivement un processus stationnaire. De plus, compte tenu de la forme particulière des autocorrélations, on peut noter que } p_a(h) = \mathbb{E} (u_t u_{t-h}) / \mathbb{E} (u_t^2) = 0 \text{ pour } h \geq 2 : \text{ le processus } (u_t) \text{ est un processus } MA(1). \quad \text{Encore une fois, le fait que } q = 1 \text{ (et que } u_t \text{ n’est alors pas un bruit blanc) est assuré en rajoutant l’hypothèse } \theta \sigma_\eta^2 \neq \phi \sigma_\epsilon^2. \text{ On peut alors conclure que } (Y_t) \text{ est un processus } ARIMA(1,1). \quad 

2. \text{De façon analogue, supposons que } (X_t) \text{ suit un processus } ARIMA(p, d, q) \text{ tel que } (1 - L)^d \Phi(L) X_t = \Theta(L) \epsilon_t \text{ et que le bruit d’observation } Y_t \text{ suit un processus } ARIMA(\alpha, \beta), \text{ tel que } A(L) Y_t = B(L) \eta_t. \text{ On s’intéresse alors à la somme de ces deux processus, } X^*_t = X_t + Y_t. \quad \text{Soit } \Phi^*(L) = \Phi(L) A(L) \text{ de degrés respectif } p^* = p + \alpha, \text{ et } q^* = q + \beta. \text{ Montrons qu’alors } (1 - L)^d \Phi^*(L) X^*_t \text{ suit un processus } MA(q^*), \text{ i.e. } (X^*_t) \text{ est un processus } ARIMA(p^*, d, q^*), \text{ ou } ARIMA(p + \alpha, d, q^*). \quad \text{Soit } \Theta^*(L) = A(L) \Theta(L) \text{ et } \Theta^{**}(L) = (1 - L)^d \Phi(L) B(L), \text{ de degrés respectifs } q^{**} = a + q \text{ et } q^{**} = d + p + \beta, \text{ et} \quad \Theta^{**}(L) = \sum_{j=0}^{q^{**}} \theta_j^{**} 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.

\[
\mathbb{E}(u_t^2) = \mathbb{E}\left[ (A(L) \Theta(L) \varepsilon_t + (1 - L)^d \Phi(L) B(L) \eta_t) (A(L) \Theta(L) \varepsilon_t + (1 - L)^d \Phi(L) B(L) \eta_t) \right] \\
= \sum_{j=0}^{q^*} \left( \theta_j^* \right)^2 \mathbb{E}(\varepsilon_{t-j}^2) + \sum_{j=0}^{q^*} \left( \theta_j^* \right)^2 \mathbb{E}(\eta_{t-j}^2)
\]

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 \( (\varepsilon_t) \) et \( (\eta_t) \) 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 précédante, ainsi que dans l'exercice 1).

\[
\mathbb{E}(u_t u_{t-1}) = \mathbb{E}\left[ (A(L) \Theta(L) \varepsilon_t + (1 - L)^d \Phi(L) B(L) \eta_t) (A(L) \Theta(L) \varepsilon_{t-1} + (1 - L)^d \Phi(L) B(L) \eta_{t-1}) \right] \\
= \sum_{j=1}^{q^*} \theta_j^* \theta_{j-1}^* \mathbb{E}(\varepsilon_{t-j}^2) + \sum_{j=1}^{q^*} \theta_j^* \theta_{j-1}^* \mathbb{E}(\eta_{t-j}^2)
\]
en utilisant encore et toujours les mêmes arguments : bruits blancs indépendants entre eux. De façon plus générale, on en déduit

\[
\mathbb{E}(u_t u_{t-k}) = \mathbb{E}\left[ (A(L) \Theta(L) \varepsilon_t + (1 - L)^d \Phi(L) B(L) \eta_t) (A(L) \Theta(L) \varepsilon_{t-k} + (1 - L)^d \Phi(L) B(L) \eta_{t-k}) \right] \\
= \sum_{j=k}^{q^*} \theta_j^* \theta_{j-k}^* \mathbb{E}(\varepsilon_{t-j}^2) + \sum_{j=k}^{q^*} \theta_j^* \theta_{j-k}^* \mathbb{E}(\eta_{t-j}^2)
\]
at condition que ces sommes soient bien définies, c'est à dire \( q^* \geq k \) et \( q^* \geq j \), sinon, ces sommes valent 0. En particulier,

\[
\mathbb{E}(u_t u_{t-q^{*+1}-1}) = \mathbb{E}\left[ (A(L) \Theta(L) \varepsilon_t + (1 - L)^d \Phi(L) B(L) \eta_t) (A(L) \Theta(L) \varepsilon_{t-q^{*+1}-1} + (1 - L)^d \Phi(L) B(L) \eta_{t-q^{*+1}-1}) \right] \\
= 0 + \sum_{j=k}^{q^*} \theta_j^* \theta_{j-k}^* \mathbb{E}(\eta_{t-j}^2)
\]
et de même pour \( \gamma (q^{*+2} + 1) \) : aussi, on en déduit facilement que \( \gamma (\max \{q^*, q^{*+2}\} + 1) = 0 \), et de façon plus générale \( \gamma (\max \{q^*, q^{*+2}\} + h) = 0 \) 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

\[
\begin{align*}
X_t &= \alpha X_{t-1} + \beta Y_{t-1} + \varepsilon_t \\
Y_t &= \delta Y_{t-1} + \gamma X_{t-1} + \eta_t
\end{align*}
\]

où \( (\varepsilon_t) \) et \( (\eta_t) \) sont des bruits blancs mutuellement indépendants (i.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

\[
Z_t = K Z_{t-1} + u_t \quad \text{où} \quad K = \begin{pmatrix} \alpha & \beta \\ \delta & \gamma \end{pmatrix} \quad \text{et} \quad u_t = (\varepsilon_t, \eta_t)'
\]

soit

\[
M(L) Z_t = u_t \quad \text{où} \quad M(L) = \begin{pmatrix} 1 - \alpha L & -\beta L \\ -\gamma L & 1 - \delta L \end{pmatrix}
\]
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
\[
\begin{pmatrix}
1 - \alpha L - \delta \gamma L \\
-\gamma L - \delta L
\end{pmatrix} Z_t = u_t,
\]
et donc
\[
\begin{pmatrix}
1 - \delta L & \beta L \\
\gamma L & 1 - \alpha L
\end{pmatrix}
\begin{pmatrix}
1 - \delta L & \beta L \\
\gamma L & 1 - \alpha L
\end{pmatrix} Z_t =
\begin{pmatrix}
1 - \delta L & \beta L \\
\gamma L & 1 - \alpha L
\end{pmatrix} u_t,
\]
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'ont 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 présentent respectivement l'estimation des coefficients, l'autocorrélogramme des résidus, puis une représentation graphique suivi de l'histogramme...
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)

dont on observe l’autocorrélogramme suivant (ainsi que le test ADF avec constante à droite)

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

\[
\begin{align*}
\text{Dependent Variable: } & Y \\
\text{Method: } & \text{Least Squares} \\
\text{Date: } & 12/03/03 \quad \text{Time: } 12:46 \\
\text{Sample(adjusted): } & 3180 \\
\text{Included observations: } & 178 \text{ after adjusting endpoints} \\
\text{Convergence achieved after } & 2 \text{ iterations} \\
\text{Variable} & \text{Coefficient} & \text{Std. Error} & \text{t-Statistic} & \text{Prob.} \\
\text{AR(1)} & 0.581741 & 0.061570 & 9.448527 & 0.0000 \\
\text{R-squared} & 0.335086 & \\
\text{Mean dependent var} & 0.021036 & \\
\text{Adjusted R-squared} & 0.335086 & \\
\text{S.D. dependent var} & 1.257401 & \\
\text{S.E. of regression} & 1.025313 & \\
\text{Akaike info criterion} & 2.893476 & \\
\text{Sum squared resid} & 186.0743 & \\
\text{Schwarz criterion} & 2.911351 & \\
\text{Log likelihood} & -256.5193 & \\
\text{Durbin-Watson stat} & 1.928592 & \\
\text{Jarque-Bera} & 6.108137 & 0.047167 & \\
\text{Series: Residuals} & \\
\text{Sample } & 3180 \\
\text{Observations } & 178 \\
\end{align*}
\]
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

\[
\begin{align*}
(1) & \quad (1 - \alpha L)(1 - \beta L)X_t = \varepsilon_t \\
(2) & \quad (1 - L)(1 - \gamma L)X_t = \eta_t
\end{align*}
\]

(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 \(\alpha, \beta, \gamma\), ainsi que la variance des bruits.

4) Donnez, dans les deux cas, une méthode pour obtenir une prédiction, 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 prédiction, 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 autocorrelations 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’autocorrélogramme), même s’il suffisait de noter que les premières autocorrélations sont largement significatives \((\hat{\rho}(1) \approx 0,576)\). Un modèle \(AR(1)\) ne peut pas être retenu.


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, un modèle \(AR(1)\) 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,

\[
\begin{align*}
(1) & \quad (1 - 0,99L)(1 - 0,59L)X_t = \varepsilon_t \\
(2) & \quad (1 - L)(1 - 0,58L)X_t = \eta_t
\end{align*}
\]

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 \(T + 1, T + 2, T + 3...\) sont

\[T X_{T+1} = 1,58X_T - 0,58X_{T-1}\]

\[T X_{T+2} = 1,58T X_{T+1} - 0,58X_T = 1,58 \times [1,58X_T - 0,58X_{T-1}] - 0,58X_T\]
\[= [1,58^2 - 0,58] X_T - 1,58 \times 0,58X_{T-1} = 1,92X_T - 0,92X_{T-1}\]

108
\[ T_{X_{T+3}} = 1,58T_{X_{T+2}} - 0,58T_{X_{T+1}} = 1,58 \times \left[ [1,58^2 - 0,58] \times T_T - 1,58 \times 0,58 \times T_{T-1} \right] - 0,58 [1,58 \times T_T - 0,58 \times T_{T-1}] \]
\[ = [1,58 \times [1,58^2 - 0,58] \times T_T + [0,58^2 - 1,58^2 \times 0,58] \times T_{T-1}] \]
\[ = 2,11T_T - 1,11T_{T-1}, \]

pouri

\[ T_{X_{T+4}} = 1,89X_T - 1,22X_{T-1} \]
\[ T_{X_{T+5}} = 0,70X_T - 1,09X_{T-1} \]

... etc

\[ T_{X_{T+9}} = -52,69X_T + 14,11X_{T-1}. \]

On retrouve ici le résultat général d’un modèle AR (Z), \( X_t = \alpha X_{t-1} + \beta X_{t-2} + \varepsilon_t, \)

\[ T_{X_{T+1}} = \alpha X_T + \beta X_{T-1} \]

\[ T_{X_{T+2}} = \alpha T_{X_T} + \beta X_T = \alpha [\alpha X_T + \beta X_{T-1}] + \beta X_T = [\alpha^2 + \beta] X_T + \alpha \beta X_{T-1} \]

\[ T_{X_{T+3}} = \alpha T_{X_{T+2}} + \beta X_{T+1} = \alpha [\alpha^2 + \beta] X_T + \alpha \beta X_{T-1} + [\alpha^2 + 2\beta] X_T + \beta [\alpha^2 + \beta] X_{T-1} \]

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 (\( \infty \))",

\[ X_t = (1 - \alpha L)^{-1}(1 - \beta L)^{-1} \varepsilon_t = \left( \sum_{i=0}^{\infty} \alpha^i L^i \right) \left( \sum_{j=0}^{\infty} \beta^j L^j \right) \varepsilon_t \]

\[ = \sum_{i=0}^{\infty} \omega_i \varepsilon_{t-i} \text{ où } \omega_i = \sum_{k=0}^{i} \alpha^k \beta^{i-k} \]

(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, \( T_{\Delta T+1} \) est \( \varepsilon_{T+1} \), dont la variance est \( \sigma^2 = V(\varepsilon_{T+1}) \approx 1 \). A horizon supérieur, les erreurs de prévision sont respectivement

\[ T_{\Delta T+2} = X_{T+2} - T_{X_{T+2}} = \varepsilon_{T+2} + [\alpha + \beta] \varepsilon_{T+1} \]

\[ T_{\Delta T+3} = X_{T+3} - T_{X_{T+3}} = \varepsilon_{T+3} + [\alpha + \beta] \varepsilon_{T+2} + [\alpha^2 + \alpha \beta + \beta^2] \varepsilon_{T+1} \]

\[ T_{\Delta T+4} = X_{T+4} - T_{X_{T+4}} = \varepsilon_{T+4} + [\alpha + \beta] \varepsilon_{T+3} + [\alpha^2 + \alpha \beta + \beta^2] \varepsilon_{T+2} + [\alpha^3 + \alpha \beta + \beta^2] \varepsilon_{T+1} \]

dont les variances respectives sont

\[ V_1 = \sigma^2, \ V_2 = \left( 1 + [\alpha + \beta]^2 \right) \sigma^2, \]

\[ V_3 = \left( 1 + [\alpha + \beta]^2 + [\alpha^2 + \alpha \beta + \beta^2]^2 \right) \sigma^2 \]

\[ V_4 = \left( 1 + [\alpha + \beta]^2 + [\alpha^2 + \alpha \beta + \beta^2]^2 + [\alpha^3 + \alpha \beta + \beta^2]^2 \right) \sigma^2. \]
Dans le cas du modèle AR(1) intégré, il n’est pas possible d’utiliser la forme MA(∞), 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
\[
(1 - L) X_t = (1 - \phi L)^{-1} \eta_t = \eta_t + \phi \eta_{t-1} + \phi^2 \eta_{t-2} + \phi^3 \eta_{t-3} + \ldots = X_t - X_{t-1}.
\]
On a alors
\[
T X_{T+1} = (1 + \phi) X_T - \phi X_{T-1} = X_T + \phi (X_T - X_{T-1}) = X_T + \phi [\eta_T + \phi \eta_{T-1} + \phi^2 \eta_{T-2} + \phi^3 \eta_{T-3} + \ldots]
\]
alors que
\[
X_{T+1} = X_T + \eta_{T+1} + \phi \eta_T + \phi^2 \eta_{T-1} + \phi^3 \eta_{T-2} + \ldots
\]
Aussi,
\[
T \Delta_{T+1} = X_{T+1} - T X_{T+1} = \eta_{T+1}.
\]
De façon analogue,
\[
T X_{T+2} = (1 + \phi) T X_{T+1} - \phi X_T = (1 + \phi) \left[ X_T + \phi [\eta_T + \phi \eta_{T-1} + \phi^2 \eta_{T-2} + \phi^3 \eta_{T-3} + \ldots] \right] - \phi X_T
\]
alors que
\[
X_{T+2} = X_{T+1} + \eta_{T+2} + \phi \eta_{T+1} + \phi^2 \eta_T + \phi^3 \eta_{T-1} + \ldots
\]
\[
= [X_T + \eta_{T+1} + \phi \eta_T + \phi^2 \eta_{T-1} + \phi^3 \eta_{T-2} + \ldots] + \eta_{T+2} + \phi \eta_{T+1} + \phi^2 \eta_T + \phi^3 \eta_{T-1} + \ldots
\]
\[
= X_T + \eta_{T+2} + (1 + \phi) \eta_{T+1} + (1 + \phi) \eta_T + \phi^2 (1 + \phi) \eta_{T-1} + \phi^3 (1 + \phi) \eta_{T-2} + \ldots
\]
Aussi,
\[
T \Delta_{T+2} = \eta_{T+2} + (1 + \phi) \eta_{T+1}.
\]
En itérant une nouvelle fois afin d’expliciter clairement le résultat,
\[
T X_{T+3} = (1 + \phi) T X_{T+2} - \phi T X_{T+1} = (1 + \phi) \left[ X_T + \phi (1 + \phi) \left[ \eta_T + \phi \eta_{T-1} + \phi^2 \eta_{T-2} + \phi^3 \eta_{T-3} + \ldots \right] \right] - \phi \left[ X_T + \phi \left[ \eta_T + \phi \eta_{T-1} + \phi^2 \eta_{T-2} + \phi^3 \eta_{T-3} + \ldots \right] \right]
\]
\[
= X_T + (1 + \phi + \phi^2) \left[ \phi \eta_T + \phi^2 \eta_{T-1} + \phi^3 \eta_{T-2} + \phi^4 \eta_{T-3} + \ldots \right]
\]
alors que
\[
X_{T+3} = X_{T+2} + \eta_{T+3} + \phi \eta_{T+2} + \phi^2 \eta_{T+1} + \phi^3 \eta_T + \ldots
\]
\[
= [X_T + \eta_{T+2} + (1 + \phi) \eta_{T+1} + (1 + \phi) \eta_T + \phi^2 (1 + \phi) \eta_{T-1} + \phi^3 (1 + \phi) \eta_{T-2} + \ldots] + \eta_{T+3} + \phi \eta_{T+2} + \phi^2 \eta_{T+1} + \phi^3 \eta_T + \ldots
\]
\[
= X_T + \eta_{T+3} + (1 + \phi) \eta_{T+2} + (1 + \phi + \phi^2) \eta_{T+1} + (1 + \phi + \phi^2) \eta_T + \phi^2 (1 + \phi + \phi^2) \eta_{T-1} + \phi^3 (1 + \phi + \phi^2) \eta_{T-2} + \ldots
\]
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
\[
T \Delta_{T+h} = \eta_{T+h} + (1 + \phi) \eta_{T+h-1} + (1 + \phi + \phi^2) \eta_{T+h-2} + \ldots + (1 + \phi + \ldots + \phi^{h-1}) \eta_{T+1}
\]
On en déduit alors les variances respectives sont
\[
V_1 = \sigma^2, \quad V_2 = \left( 1 + [1 + \phi]^2 \right) \sigma^2,
\]
\[
V_3 = \left( 1 + [1 + \phi]^2 + [1 + \phi + \phi^2]^2 \right) \sigma^2
\]
\[ V_4 = \left( 1 + [1 + \phi]^2 + [1 + \phi + \phi^2]^2 + [1 + \phi + \phi^2 + \phi^3]^2 \right) \sigma^2. \]

(iii) si l’on effectue les calculs, on notera que les chiffres sont ici grosso modo 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
\[ \alpha + \beta \approx 1 + \phi \] 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),
\[ V_2^{(i)} = \left( 1 + [\alpha + \beta]^2 \right) \sigma^2 \approx \left( 1 + [1 + \phi]^2 \right) \sigma^2 = V_2^{(ii)} \]
\[ V_3^{(i)} - V_2^{(i)} = \left[ \alpha^2 + \alpha \beta + \beta^2 \right] \sigma^2 \approx \left( [\alpha + \beta]^2 - \alpha \beta \right) \sigma^2 \]
\[ 
\approx \left( [1 + \phi]^2 - \phi \right) \sigma^2 = \left( 1 + \phi + \phi^2 \right) \sigma^2 = V_3^{(ii)} - V_2^{(ii)} \]

Du point de vue de la prévision, il est donc équivalent de considérer un modèle AR(1) : les prévision et les intervalles de confiances associés sont quasiment identiques.

**Exercice 22** On dispose de \(T\) observations d’une série temporelle, \(X_1, \ldots, X_T\). On suppose que l’on se trouve à la date \(T\) et que l’on souhaite prévoir la valeur \(X_{T+h}\). On notera \(\hat{T}X_{T+h}\) 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 \(\beta\) une constante, appelée constante de lissage, appartenant à \([0, 1]\), et considérons
\[ T X_{T+1} = (1 - \beta) \sum_{j=0}^{T-1} \beta^j X_{T-j}. \]  

1) Donner une interprétation du coefficient \(\beta\), en particulier que signifie \(\beta\) proche de 0 et \(\beta\) proche de 1 ?
2) On supposera alors que \(T X_{T+h} = T X_{T+1}\). Montrer que la prévision \(T X_{T+h}\) peut s’écrire comme une moyenne pondérée entre la prévision faite à la date précédente (en \(T-1\), i.e. \(T-1 X_{T+h}\)) et la dernière observation \(X_T\). Réécrire cette expression comme la dernière prévision corrigée d’un terme proportionnel à la dernière erreur de prévision,
\[ T X_{T+h} = T X_{T+1} + \alpha X_T - \alpha X_{T+1}. \]  

**Donner l’expression du coefficient \(\alpha\).**
3) On considérera \(\mu^*\) défini par
\[ \mu^* = \arg \min_{\mu} \sum_{j=0}^{T-1} \beta^j (X_{T-j} - \mu)^2. \]  

**Donner l’expression de \(\mu^*\).** Montrer en particulier que
\[ T X_{T+h} = \frac{1}{1 - \beta^*} \mu^*. \]  

De cette dernière relation, on notera que pour \(T\) suffisamment grand (c’est-à-dire si l’on a beaucoup d’observations), \(T X_{T+h} \approx \mu^*\), c’est-à-dire que \(T X_{T+h}\) 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 \(\beta\). On dira que la prévision est "optimale" si \(\varepsilon_T = X_T - T X_{T-1}\) 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. \(\beta\) est ici un coefficient permettant de pondérer plus ou moins les observations plus ou moins lointaines. Aussi, si \(\beta \to 0\), on notera que toutes les observations, hormis la première \((j = 0)\) seront très peu pondérées : dans cas, on préviendra les toutes dernières observations, au détriment des plus anciennes, dont le poids sera très très faible. De façon analogue, si \(\beta \to 1\), même les anciennes observations auront un poids non négligeables.

2. A la date \(T - 1\), une prévision à horizon \(h\) a été effectuée, en posant

\[
T_{-1}X_{T+h} = (1 - \beta) \sum_{j=0}^{T-2} \beta^j X_{T-1-j} = (1 - \beta) \left[ X_{T-1} + \beta X_{T-2} + \beta^2 X_{T-3} + \ldots \right].
\]

De façon analogue, la prévision faite à la date \(T\), en considérant

\[
T X_{T+h} = (1 - \beta) \sum_{j=0}^{T-1} \beta^j X_{T-j} = (1 - \beta) \left[ X_{T} + \beta X_{T-1} + \beta^2 X_{T-2} + \beta^3 X_{T-3} + \ldots \right] = (1 - \beta) X_T + \beta (1 - \beta) \left[ X_{T-1} + \beta X_{T-2} + \beta^2 X_{T-3} + \ldots \right]
\]

aussi, \(T X_{T+h}\) est obtenu comme moyenne pondérée, avec respectivement des poids \((1 - \beta)\) et \(\beta\), de la dernière observation \(X_T\), et de l’ancienne prévision \(T_{-1}X_{T+h}\). On retrouve d’ailleurs ici que si \(\beta \to 0\), on ne prend en compte que la dernière observation et pas la dernière prévision (on n’intègre plus le passé "lointain"), alors que si \(\beta \to 1\), on est assez peu sensible à la toute dernière observation, compte tenu du poids important du passé "lointain".

On peut d’ailleurs réécrire

\[
T X_{T+h} = (1 - \beta) X_T + \beta_{T-1} X_{T+h} = (1 - \beta) X_T + [1 - (1 - \beta)]_{T-1} X_{T+h}
\]

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

\[
\sum_{j=0}^{T-1} \beta^j (X_{T-j} - \mu) = 0
\]

soit \(\mu \left( \sum_{j=0}^{T-1} \beta^j \right) = \sum_{j=0}^{T-1} \beta^j X_{T-j}\)

d’où finalement

\[
\mu^* = \frac{1 - \beta^T}{1 - \beta} \sum_{j=0}^{T-1} \beta^j X_{T-j}.
\]

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 \((T \text{ grand})\), on note que \(T X_{T+h} \approx \mu^*\) (d’autant plus que \(\beta\) 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 \( \varepsilon_T = X_T - T_{-1} X_T \) est l’innovation du processus \( (X_t) \). Et donc, en utilisant (8),

\[
T X_{T+h} = T_{-1} X_{T+h} + (1 - \beta) [X_T - T_{-1} X_{T+h}].
\]

En remplaçant \( T_{-1} X_T \) par \( X_T - \varepsilon_T \), ou, de façon plus générale, \( T X_{T+h} = T X_{T+1} \) par \( X_{T+1} - \varepsilon_{T+1} \), on peut réécrire cette expression sous la forme

\[
T X_{T+1} = T_{-1} X_T + (1 - \beta) [X_T - T_{-1} X_T] \text{ ou } X_{T+1} - \varepsilon_{T+1} = X_T - \varepsilon_T + (1 - \beta \varepsilon_T)
\]

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 \(N(0,1)\) indépendantes.

Ceci permet de simuler aisément un vecteur \( \varepsilon \sim N(0, I) \in \mathbb{R}^d \) en simulant \( d \) variables uniformes indépendantes. Pour information, si l’on cherche à simuler un vecteur \( X \sim N(m, \Sigma) \) où la matrice \( \Sigma \) 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 \( \Sigma \), i.e. \( T \) telle que \( TT' = \Sigma \) (ex: décomposition de Cholesky) : comme \( \Sigma \) est inversible, \( T \) aussi,

On peut noter que \( Z = T^{-1} (X - m) \sim N(0, I) \), 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 \( \text{ARMA} \)

**Simulation d’un processus \( \text{MA}(q) \)** Nous avons vu dans la partie précédente comment simuler une échantillon i.i.d. gaussien, \( \varepsilon_1, \ldots, \varepsilon_n \) correspondant à notre bruit blanc. De façon générale, pour simuler un processus \( \text{MA}(q) \) l’algorithme est le suivant :

(i) simuler \( q + n \) valeurs initiales d’un bruit blanc \( \varepsilon_{-q+1}, \ldots, \varepsilon_0, \varepsilon_1, \ldots, \varepsilon_n \)

(ii) à l’étape \( i + 1 \), on pose \( x_{i+1} = \varepsilon_{i+1} + \theta_1 \varepsilon_i + \theta_2 \varepsilon_{i-1} + \ldots + \theta_q \varepsilon_{i-q+1} \).

**Exemple 30** Les graphiques suivant correspondent à la simulation d’un processus \( \text{MA}(1) \) de coefficient \( \theta_1 = 0.8 \), avec \( n = 20 \) en haut et \( n = 250 \) en bas, et avec \( x_t \) à gauche, et \( \varepsilon_t \) à droite.

---

**Simulation d’un processus \( \text{AR}(p) \)** Un processus \( \text{AR}(p) \) étant un processus Markovien, simuler un processus \( \text{AR}(p) \) 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 \( (\varepsilon_t) \) de variables indépendantes et de même loi à valeurs

(ii) une application mesurable \( \Phi \) telle que pour tout \( t \in \mathbb{N} \), \( X_{t+1} = \Phi(t, X_t, \varepsilon_t) \)

Pour simuler un processus \( \text{AR}(p) \) l’algorithme est le suivant :

(i) choisir ou simuler une valeur initiale \( x_0 \) et simuler un bruit blanc \( \varepsilon_1, \ldots, \varepsilon_n \)

(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 \( \text{AR}(1) \) de coefficient \( \phi_1 = -0.8 \),
avec n = 20 en haut et n = 250 en bas, et avec xₜ à gauche, et εₜ à droite.

De façon générale, pour simuler un processus $AR(p)$ l’algorithme est le suivant :

(i) choisir ou simuler $p$ valeurs initiales $x₀, x₋₁, ..., x₋ₚ₊₁$ et simuler un bruit blanc $ε₁, ..., εₙ$

(ii) à l'étape $i+1$, on pose $x_{i+1} = φ₁x₁ + φ₂x₋₁ + ... + φₚx₋ₚ₊₁ + ε_{i+1}$.

Simulation d'un $ARMA(p, q)$ Un processus $ARMA(p, q)$, de la forme $Φ(L) Xₜ = Θ(L) εₜ$ 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 $ARCH$

[À 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

$$dXₜ = b(Xₜ) dt + a(Xₜ) dWₜ, \quad X₀ = x.$$  \hspace{1cm} (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

$$dXₜ = rXₜdt + XₜσdWₜ$$

dont la solution est

$$Xₜ = X₀ \exp \left( \left[ r - \frac{1}{2} σ² \right] t + σWₜ \right)$$

Pour simuler $Xₜ$, il suffit alors de simuler la valeur de $(Wₜ)$ à la date $t$, i.e. une loi normale $N(0, t)$.

Pour simuler $(Xₜ)$, on a recours à une approximation de $(Xₜ)$, correspondant à une discrétisation de l’équation (48). Pour cela, considérons une partition de $[0, T]$ en $n$ subdivision, de taille $T/n$, et notons $tᵢ = iT/n$. Soit $δ_n = T/n$ le pas de temps, et on noter $ΔWₙ^ᵢ = Wᵢₙ - Wᵢ₋₁$. La discrétisation d’Euler consiste à écrire

$$Xₙᵢ = Xₙᵢ₋₁ + ∫_{tᵢ₋₁}^{tᵢ} b(Xₛ) ds + ∫_{tᵢ₋₁}^{tᵢ} a(Xₛ) dWₛ ≈ Xₙᵢ₋₁ + b(Xₙᵢ₋₁) δ_n + a(Xₙᵢ₋₁) ΔWₙᵢ.$$
ce qui conduit à considérer le processus suivant
\[
\begin{align*}
\tilde{X}_0 &= X_0 \\
\tilde{X}_{t_i} &= \tilde{X}_{t_{i-1}} + b \left( \tilde{X}_{t_{i-1}} \right) \delta_n + a \left( \tilde{X}_{t_{i-1}} \right) \Delta W_i
\end{align*}
\]  
(49)

et simuler \(\tilde{X}_t\) revient à simuler les accroissements \(\Delta W_i \sim \mathcal{N}(0, \delta_n)\), i.e ce qui revient à simuler les lois gaussiennes indépendantes, centrées, de variance \(\delta_n\).

Sous l’hypothèse où \(a\) et \(b\) sont lipschitziennes, il est alors possible de montrer (application du lemme de Gronwall) que, pour \(p \geq 1\),
\[
\mathbb{E} \left( \sup_{t \in [0,T]} \left\| X_t - \tilde{X}_t \right\|_p \right)^{1/p} \leq \frac{C}{\sqrt{n}}
\]
c’est à dire que l’on a une convergence uniforme, dans \(L^p\), du processus discretisé \(\tilde{X}_t\) vers le processus \(X_t\) en \(1/\sqrt{n}\). Ou de façon plus générale, en rajoutant des conditions de régularité sur \(a\) et \(b\), on a que
\[
\mathbb{E} \left\| g(X_T) - g(\tilde{X}_T) \right\| \leq \frac{C}{n}
\]
(en utilisant le théorème d’Itô).

Remarque 23 L’approximation utilisée ici, correspondant au schéma d’Euler, repose sur l’approximation
\[
\int_{t_{i-1}}^{t_i} a(X_s) \, dW_s \approx a(X_{t_{i-1}}) \Delta W_i
\]
mais une approximation à un ordre supérieur est également possible. En utilisant le fait que
\[
\int_{t_{i-1}}^{t_i} \Delta W_i \, dW_s = \frac{1}{2} \left[ (\Delta W_i) - \delta_n \right],
\]
peut approximer
\[
\int_{t_{i-1}}^{t_i} a(X_s) \, dW_s \approx a(X_{t_{i-1}}) \Delta W_i + \frac{1}{2} a'(X_{t_{i-1}}) a(X_{t_{i-1}}) \left[ (\Delta W_i)^2 - \delta_n \right]
\]
Aussi, le schéma d’approximation (49) peut se réécrire
\[
\begin{align*}
\tilde{X}_0 &= X_0 \\
\tilde{X}_{t_i} &= \tilde{X}_{t_{i-1}} + b \left( \tilde{X}_{t_{i-1}} \right) \delta_n + a \left( \tilde{X}_{t_{i-1}} \right) \Delta W_i + \frac{1}{2} a'(X_{t_{i-1}}) a(X_{t_{i-1}}) \left[ (\Delta W_i)^2 - \delta_n \right]
\end{align*}
\]
(au 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 dérivables, et pour \(p \geq 1\)
\[
\max_{i=1, \ldots, n} \mathbb{E} \left( \left\| X_{t_i} - \tilde{X}_{t_i} \right\|_p \right)^{1/p} \leq \frac{C}{n}
\]

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://ysdx.myetang.com/soft.htm. La page étant en chinois, je déclare toute responsabilité sur l’usage de cette version n’est pas autorisée par les éditeurs.
6.1.1 Les données sous EViews


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 ‘Serie’.

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 \(\Phi\) 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.
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 intéressant 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 retrouver sur le document Word et d’y coller l’image.

6.2 La régression linéaire sous Eviews

Considérons la sortie suivante, obtenue sous EViews,

\[
\begin{array}{l}
\text{LS // Dependent Variable is } Y \\
\text{Sample: 1 14} \\
\text{Included observations: 14} \\
\hline \\
\text{Variable} & \text{Coefficient} & \text{Std. Error} & \text{T-Statistic} & \text{Prob.} \\
\text{(1)} & \text{(2)} & \text{(3)} & \text{(4)} \\
\hline \\
C & 32.89132 & 11.66331 & 2.820068 & 0.0182 \\
X1 & 0.801901 & 0.298436 & 2.687012 & 0.0228 \\
X2 & -0.381362 & 0.156581 & -2.435564 & 0.0351 \\
X3 & -0.037132 & 0.052023 & -0.713768 & 0.4917 \\
\hline \\
\text{R-squared} & 0.702687 & \text{Mean dependent var} & 17.71429 & \text{(11)} \\
\text{Adjusted R-squared} & 0.613493 & \text{S.D. dependent var} & 4.177385 & \text{(12)} \\
\text{S.E. of regression} & 2.597069 & \text{Akaike info criterion} & 2.143723 & \text{(13)} \\
\text{Sum squared resid} & 67.44767 & \text{Schwartz criterion} & 2.326311 & \text{(14)} \\
\text{Log likelihood} & -30.87120 & \text{F-statistic} & 7.878181 & \text{(15)} \\
\text{Durbin-Watson stat} & 3.186886 & \text{Prob(F-statistic)} & 0.005452 & \text{(16)} \\
\end{array}
\]

Ce type de sortie correspond à des sorties obtenues en économétrie, lors de l’estimation des paramètres du modèle linéaire

\[ Y_i = \beta_1 X_{i1} + \beta_2 X_{i2}^2 + \beta_3 X_{i3}^3 + \beta_4 X_{i4}^4 + \varepsilon_i \]

qui se réécrit de façon matricielle \( Y_i = X_i \beta + \varepsilon_i \). Supposons que nous avons un échantillon \( Y_1, \ldots, Y_n \) et \( X_{i1}, \ldots, X_{in} \) de \( n \) observations. On supposera l’hypothèse \((H)\) sur les résidus \( \varepsilon_i \), de stationnarité au second ordre :

- les résidus sont centrés \( \mathbb{E}(\varepsilon_i) = 0 \)
- la variance des résidus est constante : \( \text{Var}(\varepsilon_i) = \sigma^2 \) (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 \( \hat{Y}_i \) ajustée sont représentés ci-dessous à gauche. La sortie de droite donne différentes statistiques sur \( \hat{\varepsilon}_i = Y_i - \hat{Y}_i \),

Pour information, les observations étaient les suivantes,

\[
\begin{array}{cccccccccccccccc}
& 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 \\
X_1 & 2 & 1 & 3 & 6 & 7 & 8 & 8 & 5 & 5 & 8 & 4 & 9 & 12 & 7 \\
X_2 & 45 & 43 & 43 & 47 & 42 & 41 & 32 & 33 & 41 & 38 & 32 & 31 & 35 & 29 \\
\hline \\
\hat{\varepsilon} & -0.84 & 1.61 & -3.18 & 1.61 & -3.70 & 1.12 & -1.20 & 0.14 & 4.49 & -2.76 & 1.08 & -0.90 & 2.29 & 0.24 \\
\end{array}
\]

(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 \( \beta \)

**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 \( \mathbb{E}(\varepsilon_1) = 0 \) 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 \( \beta \), dont la variance est minimale

Cet estimateur sera qualifié de BLUÉ - Best Linear Unbiased Estimator (meilleur estimateur linéaire sans biais). Sous l’hypothèse où \( (\varepsilon_1) \) est gaussien, soit \( \varepsilon_1 \sim N(0, \sigma^2) \), on a le résultat suivant

**Propriété 12** \( \hat{\beta} \) est un estimateur gaussien, centré en \( \beta \) de variance \( (X'X)^{-1}\sigma^2 \)

La variance \( \sigma^2 \) des résidus peut être estimée par
\[ s^2 = \frac{1}{n - p - 1} \| Y - X\hat{\beta} \|^2 = \frac{1}{n - p - 1} \| \varepsilon \|^2 \]
ou \( p \) 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 \( \hat{Y}_i = X_i \hat{\beta} \).

**Exemple 32** Dans l’exemple présenté dans l’introduction, on a
\[
X = \begin{pmatrix}
1 & 2 & 45 & 121 \\
1 & 1 & 43 & 132 \\
\vdots & \vdots & \vdots & \vdots \\
1 & 7 & 29 & 180
\end{pmatrix}, \quad X'X = \begin{pmatrix}
14 & 85 & 532 & 2094 \\
85 & 631 & 3126 & 13132 \\
532 & 3126 & 2066 & 78683 \\
2094 & 13132 & 78683 & 317950
\end{pmatrix}
\]
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
\[ (X'X)^{-1} X'Y = \begin{pmatrix}
248 \\
1622 \\
9202 \\
37592
\end{pmatrix} = \begin{pmatrix}
32.891 \\
0.802 \\
-0.381 \\
-0.037
\end{pmatrix} \]

**Exemple 33** Toujours dans cet exemple, la variance des résidus est estimée par
\[ s^2 = \frac{1}{14 - 3 - 1} \sum_{i=1}^{14} \varepsilon_i^2 = \frac{67.45}{10} = 6.745 \]

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

$$\frac{1}{n} \sum (Y_i - \bar{Y})^2 = \frac{1}{n} \sum (Y_i - \hat{Y}_i)^2 + \frac{1}{n} \sum (\hat{Y}_i - \bar{Y})^2$$

qui peut aussi s’écrire

$$||\hat{\varepsilon}||^2 = ||Y - X\hat{\beta}||^2 + ||X\hat{\beta} - X\bar{\beta}||^2$$

On peut alors montrer (théorème de Cochran) que

$$\frac{||Y - X\bar{\beta}||^2}{\sigma}$$

suit un $\chi^2 (n-p)$

Donc, en remplaçant $\sigma$ par son estimateur, on peut montrer, toujours sous l’hypothèse de normalité de $\varepsilon_t$ que

$$t_{n-p-1} = \frac{(\hat{\beta}_i - \bar{\beta}_i)}{\sqrt{\sum (Y_i - \hat{Y}_i)^2 [(X'X)^{-1}]_{ii}}} \sqrt{n-p-1}$$

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

$$\sigma \sqrt{[(X'X)^{-1}]_{ii}}$$

La statistique $t_{n-p-1}$ doit alors être comparée à la valeur donnée dans la table de Student.

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

\[
t^{(1)} = \frac{32.89132}{11.66331} = 2.820068 \text{ et } t^{(2)} = \frac{0.801901}{0.298436} = 2.687012
\]

\[
t^{(3)} = \frac{-0.383162}{0.156581} = 2.435564 \text{ et } t^{(4)} = \frac{-0.037132}{0.052023} = 0.713768
\]

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, $t_{0.05} = 2.228$ : 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 $\alpha$
- accepter à tort $H_0$ : cas (ii), erreur de seconde espèce - probabilité $\beta$

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 $\alpha$ et $\beta$ varie en sens inverse, aussi, plus $1 - \beta$ sera petit, plus $\alpha$ sera petit : minimiser 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 :

<table>
<thead>
<tr>
<th>$t$</th>
<th>0.00</th>
<th>0.05</th>
<th>0.10</th>
<th>0.20</th>
<th>0.30</th>
<th>0.50</th>
<th>0.60</th>
<th>0.70</th>
<th>0.80</th>
<th>0.90</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1.093</td>
<td>1.372</td>
<td>1.812</td>
<td>2.228</td>
<td>2.764</td>
<td>3.169</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 Boz-Pierce, $H_0$ correspond à l’hypothèse de bruit blanc : si $p$ est faible, on rejette l’hypothèse de bruit blanc des résidus

122
6.2.3 Coefficient $R^2$ (5) – (6)

**Coefficient $R^2$** Le coefficient $R$ est le coefficient de corrélation entre la série $Y_1, ..., Y_n$ et la série estimée $\hat{Y}_1, ..., \hat{Y}_n$. Le carré de ce coefficient s’interprète alors en termes de variance expliquée :

$$R^2 = \frac{\text{Variance expliquée par la régression}}{\text{Variance totale de } Y} = \frac{\sum (Y_i - \overline{Y})^2 - \sum (Y_i - \hat{Y}_i)^2}{\sum (Y - \overline{Y})^2}$$

Le coefficient $R^2$ est le cosinus de l’angle formé par $Y - \overline{Y}$ et $\hat{Y} - \overline{Y}$ dans $\mathbb{R}^n$, où $\overline{Y}$ est le vecteur $\overline{Y} \times \mathbb{I}$.

**Remarque 25** Il est toujours possible d’augmenter ’artificiellement’ un $R^2$ : considérons $p + 1$ variables $X'_i$, alors on a toujours

$$R^2 \left( X'_1, ..., X'_p, X'_{p+1} \right) \geq R^2 \left( X'_1, ..., X'_p \right)$$

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

$$R^2 = 1 - \frac{\sum_{i=1}^n \hat{e}_i^2}{\sum_{i=1}^n (Y_i - \overline{Y})^2}$$

Il faut toutefois prendre des précaution 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 $i = 1, ..., n$. Pour $n = 30$, on obtient le modèle

$$Y_i = 5.92 + 0.03 X_i \text{ avec un } R^2 = 0.94$$

Ce modèle peut paraître très significatif ($R^2 = 94\%$) mais ne pas être utilisé pour autant : Durbin Watson : $DW = 0.057$.

**Coefficient $R^2$ ajusté - $\overline{R}^2$** Le $R^2$ ajusté, noté $\overline{R}^2$, permet de prendre en compte le nombre de variables utilisées dans le modèle,

$$\overline{R}^2 = \frac{(n - 1) R^2 - p}{n - p} = 1 - \frac{1}{n - p} \frac{\sum_{i=1}^n \hat{e}_i^2}{\sum_{i=1}^n (Y_i - \overline{Y})^2}$$

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 $\sum_{i=1}^n \hat{e}_i = 67.45$ et que $\sum_{i=1}^n (Y_i - \overline{Y})^2 = 226.86$, on alors

$$R^2 = 1 - \frac{67.45}{226.86} = 0.702$$

et le $R^2$ ajusté est

$$\overline{R}^2 = 1 - \frac{14 - 1}{14 - 4} \left( 1 - R^2 \right) = 1 - \frac{14 - 1}{14 - 4} (1 - 0.702) = 0.613$$
6.2.4 Somme des carrés (7) – (8)
La variable (8) est la somme des carrés des résidus,
\[ SQR = SCR = s^2 = \sum_{i=1}^{n} \varepsilon_i^2 \]
La variable (7) est l’écart-type ‘de la régression’, c’est à dire σ, écart-type des erreurs ε. Cette valeur est estimé par
\[ SER = \sqrt{s^2} \; \text{où} \; s^2 = \frac{1}{14 - 3 - 1} \sum_{i=1}^{14} \varepsilon_i^2 = \frac{67.45}{10} = 6.745 \]

Exemple 38 En reprenant l’exemple de l’introduction, on a SCR = \sum_{i=1}^{n} \hat{\varepsilon}_i = 67.45 et donc
\[ SER = s \; \text{où} \; s^2 = \frac{1}{14 - 3 - 1} \sum_{i=1}^{14} \varepsilon_i^2 = \frac{67.45}{10} = 6.745 \]
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 sont normalement distribuées, centrées, de variance σ². L’idépendance des ε_i implique que \( ε = (ε_1, ..., ε_n)' \) suit un vecteur gaussien, \( \mathcal{N}(0, σ^2 I_{n \times n}) \). Aussi, Y suit un vecteur gaussien \( Y \sim \mathcal{N}(X\beta, σ^2 I_{n \times n}) \). Aussi, conditionnellement aux vecteurs \( X_j, j = 1, ..., p \), aux vecteurs β des coefficients de la régression, et à σ², la distribution des \( Y_i \) est normale :
\[ f(Y_i|X_{1,i}, ..., X_{p,i}, \beta, σ^2) = \frac{1}{\sqrt{2πσ^2}} \exp \left( -\frac{(Y_i - X'_iβ)^2}{2σ^2} \right) \]
en notant \( X_i = (X_{1,i}, ..., X_{p,i})' \). Les observations étant supposées indépendantes, la distribution jointe des \( Y_i \) est
\[ f(Y_1, Y_2, ..., Y_n) = f(Y_1) \times f(Y_2) \times \cdots \times f(Y_n) = \frac{1}{(2πσ^2)^{n/2}} \exp \left( -\frac{(Y - Xβ)'(Y - Xβ)}{2σ^2} \right) \]
Aussi, la vraisemblance du modèle est
\[ \ell(\beta, σ^2|Y, X) = \frac{1}{(2πσ^2)^{n/2}} \exp \left( -\frac{(Y - Xβ)'(Y - Xβ)}{2σ^2} \right) \]
et la log-vraisemblance s’écrit alors
\[ \log \ell(\beta, σ^2|Y, X) = -\frac{n}{2} \log 2π - \frac{n}{2} \log σ^2 - \frac{(Y - Xβ)'(Y - Xβ)}{2σ^2} \]
Aussi, la log-vraisemblance empirique du modèle, donnée dans (9), est
\[ L = -\frac{n}{2} \log 2π - \frac{n}{2} \log s^2 - \frac{(Y - Xβ)'(Y - Xβ)}{2σ^2} \; \text{où} \; s^2 = \frac{1}{n - p - 1} \sum_{i=1}^{n} \varepsilon_i^2 = \frac{1}{n - p - 1} (Y - X\beta)'(Y - X\beta) \]

Exemple 39 En reprenant l’exemple de l’introduction, on a \( s^2 = 6.745 \) et donc
\[ L = -\frac{14}{2} \log 2π - \frac{14}{2} \log 6.745 - \frac{14 - 3 - 1}{2} = -30.871 \]
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

\[ \varepsilon_t = \rho \varepsilon_{t-1} + \eta_t \text{ où } \eta_t \sim N \left(0, \sigma_\eta^2\right) \]

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

\[ DW = \frac{\sum_{t=2}^{n} \left( \varepsilon_t - \tilde{\varepsilon}_{t-1}\right)^2}{\sum_{t=1}^{n} \varepsilon_t^2} \]

où les \( \tilde{\varepsilon}_t \) 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 \( \rho \) 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.

<table>
<thead>
<tr>
<th>n</th>
<th>( t=1 )</th>
<th>( t=2 )</th>
<th>( t=3 )</th>
<th>( t=4 )</th>
<th>( t=5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>1,00</td>
<td>1,09</td>
<td>0,95</td>
<td>1,34</td>
<td>0,92</td>
</tr>
<tr>
<td>16</td>
<td>1,00</td>
<td>1,19</td>
<td>0,96</td>
<td>1,34</td>
<td>0,92</td>
</tr>
<tr>
<td>17</td>
<td>1,13</td>
<td>1,18</td>
<td>1,01</td>
<td>1,34</td>
<td>1,31</td>
</tr>
<tr>
<td>18</td>
<td>1,16</td>
<td>1,29</td>
<td>1,07</td>
<td>1,35</td>
<td>1,39</td>
</tr>
<tr>
<td>19</td>
<td>1,23</td>
<td>1,40</td>
<td>1,09</td>
<td>1,36</td>
<td>1,46</td>
</tr>
<tr>
<td>20</td>
<td>1,28</td>
<td>1,41</td>
<td>1,14</td>
<td>1,36</td>
<td>1,50</td>
</tr>
<tr>
<td>21</td>
<td>1,32</td>
<td>1,46</td>
<td>1,14</td>
<td>1,36</td>
<td>1,56</td>
</tr>
<tr>
<td>22</td>
<td>1,37</td>
<td>1,55</td>
<td>1,15</td>
<td>1,36</td>
<td>1,60</td>
</tr>
<tr>
<td>23</td>
<td>1,26</td>
<td>1,31</td>
<td>1,15</td>
<td>1,48</td>
<td>1,50</td>
</tr>
<tr>
<td>24</td>
<td>1,27</td>
<td>1,35</td>
<td>1,15</td>
<td>1,49</td>
<td>1,51</td>
</tr>
<tr>
<td>25</td>
<td>1,38</td>
<td>1,45</td>
<td>1,12</td>
<td>1,50</td>
<td>1,61</td>
</tr>
<tr>
<td>26</td>
<td>1,41</td>
<td>1,46</td>
<td>1,12</td>
<td>1,50</td>
<td>1,62</td>
</tr>
<tr>
<td>27</td>
<td>1,47</td>
<td>1,50</td>
<td>1,13</td>
<td>1,50</td>
<td>1,63</td>
</tr>
<tr>
<td>28</td>
<td>1,47</td>
<td>1,52</td>
<td>1,13</td>
<td>1,50</td>
<td>1,63</td>
</tr>
<tr>
<td>29</td>
<td>1,51</td>
<td>1,54</td>
<td>1,14</td>
<td>1,50</td>
<td>1,64</td>
</tr>
<tr>
<td>30</td>
<td>1,53</td>
<td>1,55</td>
<td>1,15</td>
<td>1,51</td>
<td>1,65</td>
</tr>
<tr>
<td>31</td>
<td>1,55</td>
<td>1,56</td>
<td>1,15</td>
<td>1,51</td>
<td>1,66</td>
</tr>
<tr>
<td>32</td>
<td>1,57</td>
<td>1,57</td>
<td>1,15</td>
<td>1,51</td>
<td>1,66</td>
</tr>
<tr>
<td>33</td>
<td>1,60</td>
<td>1,58</td>
<td>1,16</td>
<td>1,51</td>
<td>1,67</td>
</tr>
<tr>
<td>34</td>
<td>1,63</td>
<td>1,59</td>
<td>1,16</td>
<td>1,51</td>
<td>1,67</td>
</tr>
<tr>
<td>35</td>
<td>1,65</td>
<td>1,60</td>
<td>1,17</td>
<td>1,51</td>
<td>1,68</td>
</tr>
<tr>
<td>36</td>
<td>1,66</td>
<td>1,61</td>
<td>1,17</td>
<td>1,51</td>
<td>1,68</td>
</tr>
<tr>
<td>37</td>
<td>1,67</td>
<td>1,62</td>
<td>1,17</td>
<td>1,51</td>
<td>1,69</td>
</tr>
<tr>
<td>38</td>
<td>1,69</td>
<td>1,63</td>
<td>1,17</td>
<td>1,51</td>
<td>1,69</td>
</tr>
<tr>
<td>39</td>
<td>1,70</td>
<td>1,64</td>
<td>1,18</td>
<td>1,51</td>
<td>1,70</td>
</tr>
<tr>
<td>40</td>
<td>1,70</td>
<td>1,64</td>
<td>1,18</td>
<td>1,51</td>
<td>1,70</td>
</tr>
</tbody>
</table>

\( \bar{Y} \) est le nombre de variables expliquées (constante exclue).

\( \bar{Y} \) 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’Aikaïké - 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.

\[
AIC = \log \left( \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2 \right) + \frac{2p}{n}
\]

Le critère de Schwarz - BIC ou SC
Le critère de Schwarz est un critère légèrement différent de celui introduit par Aikaïké,

\[
SC = \log \left( \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2 \right) + \frac{p \log n}{n}
\]

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 : \( H_0 : \beta_1 = \ldots = \beta_p = 0 \) contre \( H_1 : \) il existe \( j \) 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 $\hat{Y}$ la solution des moindres carrés ordinaires, soit $\hat{Y} = X (X'X)^{-1} X'Y$, et $\hat{Y}_0$ la solution des moindres carrés contraints (contrainte $A \beta = \theta$).

On peut alors montrer que si $H_0$ est vraie,

$$
\frac{n - p}{p - 1} \left\| Y - \hat{Y}_0 \right\|^2 - \left\| Y - \hat{Y} \right\|^2 = F (p - 1, n - p)
$$

où $p$ est le nombre de variables, incluant 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 dépassees ($F = \chi^2(p)$) |
|-----|-----|-----|-----|-----|-----|
| $F_{p=0.1}$ | $F_{p=0.05}$ | $F_{p=0.01}$ | $F_{p=0.005}$ | $F_{p=0.001}$ |
| 1.00 | 1.64 | 2.71 | 3.84 | 4.60 |

Exemple 40 En reprenant l’exemple de l’introduction, la statistique de Fisher, permettant de tester la significativité globale de la régression à $3$ variables $(X_1, X_2, X_3)$, est donnée par

$$
F = \frac{n - p}{p - 1} \frac{R^2}{1 - R^2} = \frac{0.702}{1 - 0.702} = 7.878
$$

Cette valeur est à comparer avec le $F$ de Fisher à 3 et 10 degrés de liberté : $F_{3,10}^{0.05} = 3.71$. $F$ étant plus grande que $F_{3,10}^{0.05}$, nous rejetons l’hypothèse de nullité de tous les coefficients : 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 $\bar{s}^2$ n’est pas ici exactement la même que $s^2$. En effet,

$$
\bar{s}^2 = \frac{1}{n - p} \sum_{i=1}^{n} \hat{e}_i^2 = \frac{1}{14 - 3 - 1} \sum_{i=1}^{14} \hat{e}_i^2 = \frac{67.45}{10} = 6.745
$$

19References inclues toujours la constante comme variable explicative.
alors que
\[ s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \varepsilon_i^2 = \frac{1}{14-1} \sum_{i=1}^{14} \varepsilon_i^2 = \frac{67.45}{13} = 5.188 \]

**Test de Jarque-Berra** Un test possible est celui de Bera et Jarque (1984), basé sur le skewness (coefficient d'asymétrie de la distribution) et la kurtosis (aplatissement - épaisseur des queues).

En notant \( \mu_k \) le moment d'ordre \( k \) de la distribution, \( \mu_k = \mathbb{E} \left( (X - \mathbb{E}(X))^k \right) \), on appelle skewness le coefficient \( s = \mu_3/\mu_2^{3/2} \) et kutosis \( k = \mu_4/\mu_2^2 \). Sous des hypothèses de normalités, on a normalité des estimateurs du skewness et de la kurtosis,

\[ s \overset{\mathcal{L}}{\sim} \mathcal{N} \left( 0, \sqrt{6/n} \right) \) et \( k \overset{\mathcal{L}}{\sim} \mathcal{N} \left( 3, \sqrt{24/n} \right) \) quand \( n \to \infty \)

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 \( \chi^2 \) à 2 degrés de liberté. Aussi, si \( BJ \geq \chi^2_{2} - \alpha \) (2) on rejette l'hypothèse \( H_0 \) de normalité des résidus au seuil \( \alpha \).

**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 \( \chi^2 \).

### 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 (\( Z = \alpha Y + \beta X + \gamma + c \)), 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
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 EViews, 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  0.191  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 EViews, 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

```plaintext
PROC REG; MODEL Y = X1 X2 X3;
```
SAS fournit alors la sortie suivante

### Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Prob&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3</td>
<td>159.40948</td>
<td>53.13649</td>
<td>7.878</td>
<td>0.0055</td>
</tr>
<tr>
<td>Error</td>
<td>10</td>
<td>67.44767</td>
<td>6.74477</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C Total</td>
<td>13</td>
<td>226.85714</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 | Std Error | T for H0: Parameter = 0 | Prob > |T| |
|-----------|-----|--------------------|-----------|-------------------------|--------|---|
| INTERCEP  | 1   | 32.89132428        | 11.66331015| 2.82007                 | 0.0182 |
| X1        | 1   | 0.80190069         | 0.29843584 | 2.68701                 | 0.0228 |
| X2        | 1   | -0.38136236        | 0.15658069 | -2.43556                | 0.0351 |
| X3        | 1   | -0.0371332        | 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 d’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
- le coefficient de variation : C.V. = 14.66087. Ce coefficient correspond au rapport Root MSE/Dep Mean exprimé en pourcentage, CV = 100 × 2.59707/17.1429
- le $R^2$ : R-square = 0.7027
- le $R^2$ ajusté, $R^2_{adj}$ = Adj R-sq = 0.6135

La régression linéaire sous RATS

La régression linéaire se fait à l’aide de la commande suivante

```r
linreg Y residus; #constant X1 X2 X3;
```

### RATS fournit alors la sortie suivante

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coeff</th>
<th>Std Error</th>
<th>T-Stat</th>
<th>Signif</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Constant</td>
<td>32.89132428</td>
<td>11.66331015</td>
<td>2.82007</td>
<td>0.01815860</td>
</tr>
<tr>
<td>2. X1</td>
<td>0.80190069</td>
<td>0.29843584</td>
<td>2.68701</td>
<td>0.02281643</td>
</tr>
<tr>
<td>3. X2</td>
<td>-0.38136236</td>
<td>0.15658069</td>
<td>-2.43556</td>
<td>0.03511440</td>
</tr>
<tr>
<td>4. X3</td>
<td>-0.0371332</td>
<td>0.05202312</td>
<td>-0.71377</td>
<td>0.49169355</td>
</tr>
</tbody>
</table>

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, $R^2$ : 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 = 67.44766031
- 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, 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

$$
H_0 : \rho(1) = \rho(2) = \ldots = \rho(h) = 0
$$

$$
H_a : \text{il existe } i \text{ tel que } \rho(i) \neq 0
$$

Pour effectuer ce test, on utilise la statistique de Box et Pierce (1970) $Q$, donnée par

$$
Q_h = T \sum_{k=1}^{h} \rho_h^2
$$

où $h$ est le nombre de retards, $T$ est le nombre d’observations et $\rho_h$ l’autocorrélation empirique. Asymptotiquement, sous $H_0$, $Q_h$ suit un $\chi^2$ à $h$ degrés de liberté. Nous rejetons l’hypothèse de bruit blanc au seuil $h$ si $Q$ est supérieure au quantile d’ordre $(1-\alpha) = 0.05$ de la loi du $\chi^2$ à $h$ degrés de liberté.

Une statistique ayant de meilleurs propriétés asymptotiques peut être utilisée :

$$
Q'_h = T(T+2) \sum_{k=1}^{h} \frac{\rho_h^2}{T-k}
$$

qui suit asymptotiquement, sous $H_0$ une loi du $\chi^2$ 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 $\chi^2$ 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 ρ(h) = \( \text{corr}(X_t, X_{t-h}) \), estimé à partir de l’échantillon par

\[
\hat{\rho}_T(h) = \frac{\hat{\gamma}_T(h)}{\hat{\gamma}_T(0)} \quad \text{où} \quad \hat{\gamma}_T(h) = \frac{1}{T-h} \sum_{t=1}^{T-h} (X_t - \overline{X_T})(X_{t-h} - \overline{X_T})
\]

La fonction d’autocorrélation partielle (2) correspond à \( \psi_X(h) \) défini par

\[
\psi_X(h) = \text{corr}(X_{t-h} - \text{EL}(X_{t-h} | X_{t-1}, \ldots, X_{t-h+1}), X_t - \text{EL}(X_t | X_{t-1}, \ldots, X_{t-h+1}))
\]

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’autocorrélogramme), 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 texte\(^{20}\).

6.4 Estimation d’un modèle ARMA sous Eviews

Considérons la série – (représentée à gauche), et considérons la modélisation ARMA (2, 1), dont l’estimation donne les résultats ci-dessous à droite

\[\begin{array}{cccc}
\text{Variable Coefficient} & \text{Std. Error} & \text{T-Statistic} & \text{Prob.} \\
AR(1) & 0.504043 & 0.067464 & 7.473705 & 0.0000 \\
AR(2) & -0.398390 & 0.033359 & -11.94246 & 0.0000 \\
MA(1) & -0.504204 & 0.067464 & -7.473705 & 0.0000 \\
\end{array}\]

Cette sortie \((\text{obtenue sous EViews})\) 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 ± 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)

---

\(^{20}\)Toute autre suggestion de méthode sera la bienvenue.
Comme nous pouvons le noter, nous retrouvons des racines très proches de celle obtenue 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’apparait comme significatif, même si les \(R^2\) et la statistique de Fisher sont identiques pour les 2 modèle. En fait, toutes les statistiques sont identiques pour les deux modèles : nous obtenons exactement le même modèle.

6.4.2 Tests sur les erreurs \(\varepsilon_t\)

Un grand nombre de tests et d’information peuvent être obtenus sur le processus \(\varepsilon_t\).

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 moyenne mobile). Sur l’exemple ci-dessous, la \(p\)-value (\(toujours très élevée\)) permet de valider l’hypothèse de bruit-blanc de l’erreur \(\varepsilon_t\).
Distribution (marginale) des erreurs

Les statistique de la distribution des erreurs permettent de tester l’hypothèse de normalité des erreurs $\varepsilon_t$. 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 $\varepsilon_t$. 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 $\varepsilon_t$. 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 $\varepsilon_t$ suivent un modèle $MA (2)$. 

135
Modélisation ARCH (1) des erreurs Il est également possible de tester si les erreurs $\varepsilon_t$ suivent un modèle ARCH (1).

La sortie obtenue sous EVIEWS présente la statistique de Fisher et le $R^2$ du modèle $\Phi (L) X_t = \Theta (L) \varepsilon_t$ où $\varepsilon_t$ suit un processus ARCH (1). De plus, EVIEWS donne les résultats de l’estimation des paramètres du modèle ARCH (1) ajusté sur la série $\varepsilon_t$,

$$\varepsilon_t \sim \eta_t \sqrt{c + \alpha \varepsilon^2_{t-1}}$$

où $\eta_t$ est un bruit blanc.

Si la constante est significative sur l’exemple présenté, le paramètre $\alpha$ 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 de 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), ** et *** 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 autocorrélations.

#### ARIMA Procedure

Name of variable = $Z$, (t)

Mean of working series = 0.042511 (4)

Standard deviation = 1.12708 (2)

Number of observations = 75 (h)

<table>
<thead>
<tr>
<th>Lag</th>
<th>Covariance</th>
<th>Correlation</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.238320</td>
<td>1.00000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-0.443324</td>
<td>-0.35800</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.120113</td>
<td>-0.09700</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.087635</td>
<td>0.07077</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.279699</td>
<td>0.22579</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-0.133568</td>
<td>-0.10786</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.245910</td>
<td>0.19858</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.046701</td>
<td>0.03771</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.239016</td>
<td>0.152213</td>
<td></td>
</tr>
</tbody>
</table>

Lag Covariance Correlation -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1 Std

La sortie suivante donne l’autocorrélogramme inverse (10) et l’autocorrélogramme partiel (11) de la série $Z_t$.

#### ARIMA Procedure

Name of variable = $Z$, (t)

Mean of working series = 0.042511 (4)

Standard deviation = 1.12708 (2)

Number of observations = 75 (h)

<table>
<thead>
<tr>
<th>Lag</th>
<th>Correlation</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.00000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-0.15030</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.06272</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.29795</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-0.38516</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.279699</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.139227</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.11243</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.10786</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.140868</td>
<td></td>
</tr>
</tbody>
</table>

Lag Correlation -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

Reverse Autocorrelations

<table>
<thead>
<tr>
<th>Lag</th>
<th>Correlation</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.00000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-0.77158</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.61314</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.50580</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.38195</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.28373</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.24951</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.21145</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.28373</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.23720</td>
<td></td>
</tr>
</tbody>
</table>

Lag Correlation -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

Partial Autocorrelations

<table>
<thead>
<tr>
<th>Lag</th>
<th>Correlation</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.00000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-0.35800</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.25827</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.38516</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-0.29795</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.279699</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.139227</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.11243</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.10786</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.140868</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lag</th>
<th>Correlation</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.00000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-0.35800</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.25827</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.38516</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-0.29795</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.279699</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.139227</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.11243</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.10786</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.140868</td>
<td></td>
</tr>
</tbody>
</table>
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 \( \chi^2 \) 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 de degrés de liberté associé au test du \( \chi^2 \). Enfin, la colonne (15) donne la probabilité que le \( \chi^2 \) soit supérieur à \( Q_K \). Enfin, (16) rappelle les autocorrélations empiriques (soit (6)).

La colonne (17) [A FINIR]

La colonne (17) [A FINIR]
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.

```sas
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 6.5 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 4.3 5.9 6.7 5.3 5.7 5.7
  3.0 3.4 4.3 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 4.4 6.2 5.1 4.4 4.4 4.9
  2.9 2.4 5.7 5.1 4.0 5.7 7.7 6.3 5.3 5.7 5.7 5.7
  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 2.3 3.2 3.9 4.3 6.2 3.9 3.2 3.0 2.2 2.2
  2.4 1.9 2.1 4.5 3.3 3.4 4.1 5.7 4.8 5.0 3.8 3.4
  1.7 3.2 2.7 3.0 3.4 3.8 5.6 6.9 3.5 3.2 3.0 3.4
  1.4 2.3 2.5 3.1 3.5 4.5 5.7 5.0 4.4 4.8 4.1 3.9
  2.1 2.9 2.7 4.3 3.9 4.1 6.5 6.4 4.6 1.5 1.5 1.9
  1.8 1.9 3.7 4.4 3.8 5.6 5.7 5.1 5.4 4.7 2.5 1.5
  1.8 2.5 2.4 1.4 3.7 3.7 4.9 5.1 3.7 5.4 3.0 1.8
  2.1 2.4 2.3 3.2 3.5 4.5 4.9 4.2 4.7 5.7 3.2 1.8
  2.0 1.7 3.2 4.2 4.4 3.4 3.9 5.3 5.6 3.2 2.3 2.2
  1.3 2.3 2.7 3.3 3.7 7.0 3.8 4.7 4.2 2.9 1.7 1.3
  1.8 2.0 2.3 6.0 2.4 5.3 3.5 2.7 2.5 1.4 1.2
  1.5 2.0 3.1 0.3 3.5 4.6 4.6 3.1 2.1 1.4 1.3 ;
  proc arima data=air; identifier var=ozone crosscorr= x1(12) summer winter;
  estimate q=(1)(12) input=( x1 summer winter )
    noconstant method=ml itprint;
  forecast lead=12 id=date interval=month;
  quit;
```

```sas
title1 "Sales of Passenger Cars";
title2 'The OUT= Data Set';
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 print data=out;
  run;
proc print data=est;
  run;
proc gplot data=out;
  plot vehicles * date = 1 / vref=0
    haxis= '1jan80'd to '1jan92'd by year;
  where _type_ ^= 'RESIDUAL' & date >= '1jan90'd;
  format date year4. ;
  run;
```

```sas
title2 'Plot of Residuals';
symbol1 i=needle;
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;
proc print data=out;
  run;
```

```sas
title2 'Plot of Forecast from WINTERS Method';
symbol1 i=none v=star h=2 / for_type='ACTUAL'
  symbol2 i=spline v=plus h=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;
  format date year4. ;
  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 phase 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

```sas
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

```sas
identify var=N_serie;
identify var=N_serie(1);
identify var=N_serie(12);
identify var=N_serie(1,1);
identify var=N_serie(1,12);
```

La forme du modèle ARIMA à estimer peut être assez complexe. Pour cela, différentes notation sont possibles sous SAS

```sas
estimate P=1; on considère le polynôme AR \( \Phi(L) = (1 - \phi_1 L) \)
estimate P=4; on considère le polynôme AR \( \Phi(L) = (1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \phi_4 L^4) \)
estimate P=(1 4); on considère le polynôme AR \( \Phi(L) = (1 - \phi_1 L - \phi_4 L^4) \)
estimate P=(1)(4); on considère le polynôme AR \( \Phi(L) = (1 - \phi_1 L) (1 - \phi_4 L^4) \)
estimate P=(1)(1 4); 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

```sas
identify CENTER la série est centré
identify NOMISS les valeurs manquantes sont exclues
identify ALPHA= permet de spécifier le niveau de significativité des tests
identify ESACF affiche les résultats de la procédure ESACF
identify SCAN affiche les résultats de la procédure SCAN
identify P= détermine la dimension AR dans les méthodes ESACF et SCAN
identify Q= détermine la dimension MA dans les méthodes ESACF et SCAN
identify NLAG= permet de spécifier le nombre des retards dans les autocorrélogrammes
identify STATIONARITY permet d’effectuer des tests de stationnarité sur la série
```

139
6.8.2 La lecture des sorties SAS

Les sorties SAS sont les suivantes

<table>
<thead>
<tr>
<th>Iter</th>
<th>Loglike</th>
<th>MA1,1</th>
<th>MA2,1</th>
<th>NUM1</th>
<th>NUM2</th>
<th>NUM3</th>
<th>Lambda</th>
<th>R Crit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-245.88481</td>
<td>-0.26684</td>
<td>0.76665</td>
<td>-1.33062</td>
<td>-0.23936</td>
<td>-0.08021</td>
<td>0.05581</td>
<td>0.00001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Iter</th>
<th>Loglike</th>
<th>MA1,1</th>
<th>MA2,1</th>
<th>NUM1</th>
<th>NUM2</th>
<th>NUM3</th>
<th>Lambda</th>
<th>R Crit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-245.88481</td>
<td>-0.26684</td>
<td>0.76665</td>
<td>-1.33062</td>
<td>-0.23936</td>
<td>-0.08021</td>
<td>0.05581</td>
<td>0.00001</td>
</tr>
</tbody>
</table>

Variance Estimate = 0.63450562
Std Error Estimate = 0.79655861
AIC = 501.769629
SBC = 518.360229
Number of Residuals = 204

Forecasts for variable OZONE

<table>
<thead>
<tr>
<th>Obs</th>
<th>Forecast</th>
<th>Std Error</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>217</td>
<td>1.4205</td>
<td>0.7966</td>
<td>-0.1407</td>
<td>2.9817</td>
</tr>
<tr>
<td>218</td>
<td>1.8446</td>
<td>0.8244</td>
<td>0.2287</td>
<td>3.4604</td>
</tr>
<tr>
<td>219</td>
<td>2.4567</td>
<td>0.8244</td>
<td>0.8408</td>
<td>4.0725</td>
</tr>
<tr>
<td>220</td>
<td>2.8590</td>
<td>0.8244</td>
<td>1.2431</td>
<td>4.4748</td>
</tr>
<tr>
<td>221</td>
<td>3.1501</td>
<td>0.8244</td>
<td>1.5342</td>
<td>4.7659</td>
</tr>
<tr>
<td>222</td>
<td>2.7211</td>
<td>0.8244</td>
<td>1.1058</td>
<td>4.3370</td>
</tr>
<tr>
<td>223</td>
<td>3.3147</td>
<td>0.8244</td>
<td>1.6989</td>
<td>4.9306</td>
</tr>
<tr>
<td>224</td>
<td>3.4787</td>
<td>0.8244</td>
<td>1.8629</td>
<td>5.0946</td>
</tr>
<tr>
<td>225</td>
<td>2.9405</td>
<td>0.8244</td>
<td>1.3247</td>
<td>4.5564</td>
</tr>
<tr>
<td>226</td>
<td>2.3587</td>
<td>0.8244</td>
<td>0.7429</td>
<td>3.9746</td>
</tr>
<tr>
<td>227</td>
<td>1.8588</td>
<td>0.8244</td>
<td>0.2429</td>
<td>3.4746</td>
</tr>
<tr>
<td>228</td>
<td>1.2898</td>
<td>0.8244</td>
<td>-0.3260</td>
<td>2.9057</td>
</tr>
</tbody>
</table>

Tous les détails sur la procédure ARIMA sont disponibles dans la doc SAS ou sur internet.