REVUE DE STATISTIQUE APPLIQUÉE

F. BORGARD
D. GUÉGAN

# Études de séries chronologiques linéaires à temps discret. Comparaison de logiciels

*Revue de statistique appliquée*, tome 44, nº 4 (1996), p. 59-80
<http://www.numdam.org/item?id=RSA_1996__44_4_59_0>

*Rev. Statistique Appliquée*, 1996, **XLIV** (4), 59-80

# ETUDES DE SÉRIES CHRONOLOGIQUES LINÉAIRES À TEMPS DISCRET. COMPARAISON DE LOGICIELS

F. Borgard*, D. Guégan**

* *F. Borgard URA 1321 – ISUP, Tour 45/55 E3, 4 Place Jussieu, 75252 Paris cedex 05*

** *D. Guégan : URA 742 – ENSAE, Timbre J120, 3 av. Pierre Larousse, 92245 Malakof cedex – e-mail : guegan : ensae.fr*

## RÉSUMÉ

Si la modélisation ARIMA figure dans la plupart des logiciels pour l'étude des séries chronologiques, les outils qui servent à l'estimation d'un modèle sont différents d'un logiciel à l'autre. Il en est de même des tests proposés pour juger de la pertinence d'un modèle et de la linéarité de la série. Nous nous proposons ici d'examiner les possibilités offertes par les logiciels que nous avons pu pratiquer.

***Mots-clés :*** *ARCH, ARIMA, Tests unitaires, Linéarité.*

## ABSTRACT

In many software products, ARIMA models are available to analyse Time Series. But the tools for estimate a model are not the same in all the softwares. Tests about model's adequacy and about serie's linearity are also different. Our purpose is to investigate facilities given by the softwares known by us.

***Keywords :*** *ARCH, ARIMA, Tests for Units Roots, Linearity.*

Pour étudier des séries chronologiques à temps discret les logiciels mettent très souvent à la disposition de l'utilisateur la modélisation ARIMA. Nous nous proposons ici d'examiner comment cette modélisation est traitée dans un certain nombre de logiciels, quels sont les éléments statistiques mis à la disposition de l'utilisateur pour déterminer si cette modélisation est suffisante ou s'il faut lui substituer une modélisation de type ARCH ou autre.

Nous retiendrons, pour la clarté de la présentation, le plan suivant :

### *1^e^ partie*

Outils opérationnels dans une modélisation ARIMA pour :

– une bonne identification des modèles

– une bonne estimation des paramètres

– une bonne pertinence des prévisions.

*2^e^ partie*

Traitements relatifs à l'identification, l'estimation, la prévision des modèles ARIMA et tests mis à disposition de l'utilisateur pour juger de la pertinence du modèle retenu, pour un certain nombre de logiciels qu'il nous a été possible d'étudier. Ces logiciels sont : ANAR, Mandrake, Rats, TSP International, Shazam, Statgraphics, SAS/ETS, Eviews.

## 1. Modélisation ARIMA de séries temporelles linéaires

Supposons observée la série $(Y_t, t \in \mathbb{Z})$ que nous souhaitons modéliser à partir d'une représentation ARIMA.

### A. Stationnarisation

Nous commençons par étudier la stationnarité de la série. Plusieurs étapes sont éventuellement nécessaires.

1) Si la série a tendance à exploser ou à subir une déflation, on peut envisager dans le cas où ses valeurs sont positives de lui faire subir la transformation de Box-Cox pour la stabiliser. Soit $f$ cette transformation, elle s'écrit :

$$\text{ou} \quad \begin{array}{ll} Z_t = f(Y_t) = \dfrac{Y_t^\lambda - 1}{\lambda} & \text{si} \quad \lambda \neq 0 \\ Z_t = f(Y_t) = \text{Log} Y_t & \text{si} \quad \lambda = 0 \end{array} \quad Y_t > 0$$

Notons que si l'on procède à une telle transformation, la série étudiée ne sera pas la série $Y_t$ mais la série $Z_t = f(Y_t)$. Les prévisions seront faites sur $(Z_t)$ et il ne faudra pas oublier de revenir à $(Y_t)$.

Pour le choix de $\lambda$ on a l'habitude de décomposer la série en sous-séries pour lesquelles on calcule la valeur moyenne et l'étendue.

Le diagramme représentant l'étendue en fonction de la valeur moyenne permet en effet de choisir une valeur de $\lambda$.

Si ce diagramme présente :

– des points alignés suivant une droite horizontale, on choisit $\lambda = 1$ (pas de transformation)

– des points alignés suivant la première bissectrice, on choisit $\lambda = 0$ (transformation logarithmique)

– des points alignés situés entre la droite horizontale et la première bissectrice, on choisit une valeur de $\lambda$ positive inférieure à 1

– des points alignés situés au-dessus de la première bissectrice une valeur de $\lambda$ négative de valeur absolue inférieure à 1.

**Exemple** : Consommation de charbon : données trimestrielles du département de l'énergie de l'Allemagne de l'Ouest 1[er] trimestre 1960-4[e] trimestre 1986; source : Harvey 1990. La figure 1 donne les graphes de la série initiale et de la série transformée par $\lambda = 0, \lambda = 0,5, \lambda = 0,5$, tandis que la figure 2 représente le diagramme moyenne-étendue pour la série initiale.

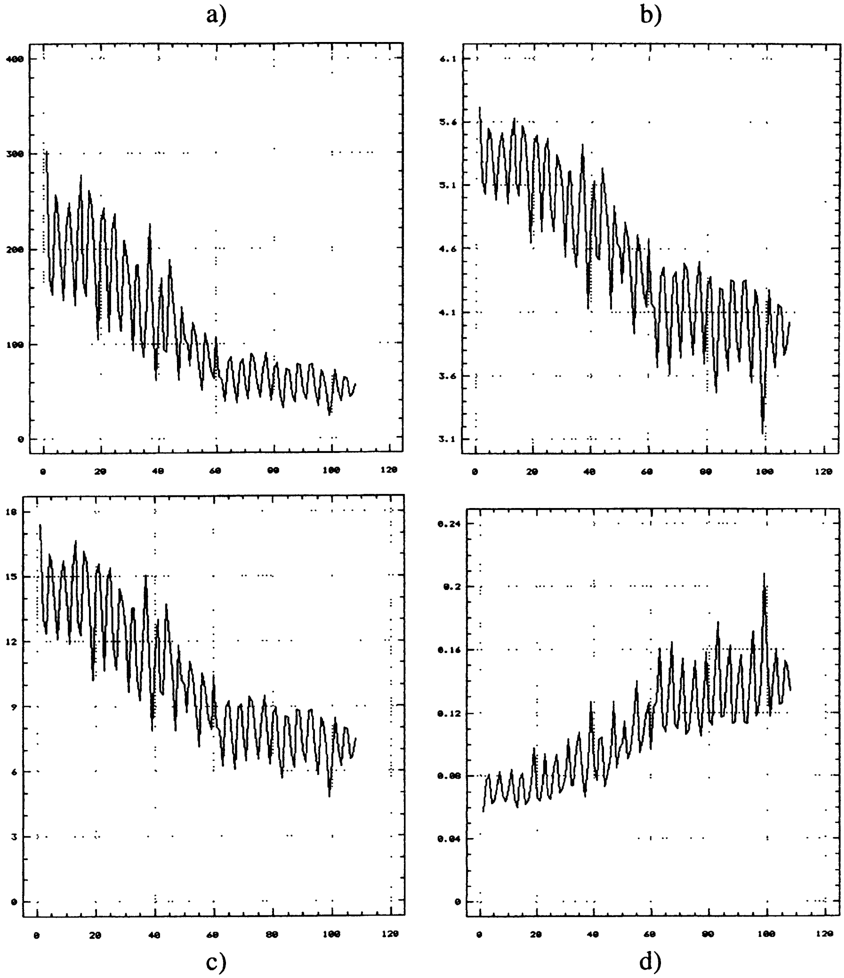



FIGURE 1

*Consommation de charbon : données trimestrielles du département de l'énergie de l'Allemagne de l'Ouest*

*a) série initiale; b)* $\lambda = 0$

*c)* $\lambda = 0,5$*; d)* $\lambda = 0,5$

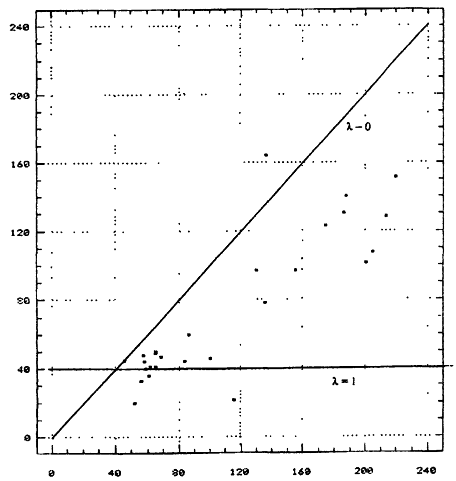



FIGURE 2
Consommation de charbon : données trimestrielles du département
de l'énergie de l'Allemagne de l'Ouest.
Diagramme Moyenne-étendue

L'utilisation du diagramme moyenne-étendue sur la série consommation de charbon indique, ici, de faire la transformation de Box et Cox avec une valeur de $\lambda$ comprise entre 0 et 1.

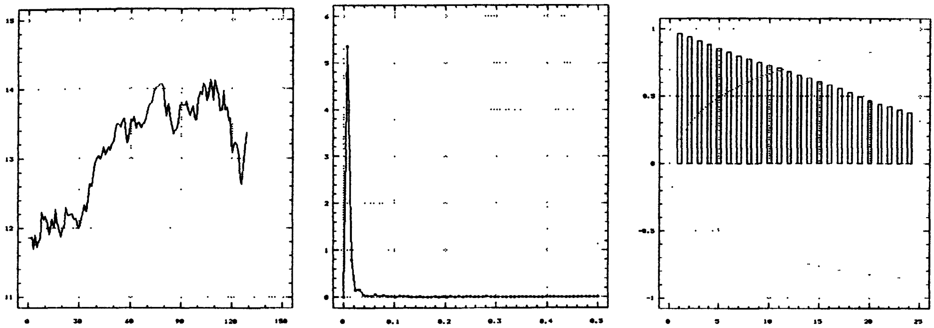
2) Si la série comporte une tendance polynomiale avec éventuellement une saisonnalité de période $S$ connue, on peut stationnariser la série par différenciation (on suppose alors que la variance est stable). Soit $B$ l'opérateur retard, $BY_t = Y_{t-1}$. On va appliquer successivement à la série $Y_t$ les opérateurs $(1 - B)^d$ et $(1 - B^S)^D$ qui auront pour but de supprimer tendance et saisonnalité. Précisons ce qui se passe pour $d = 1$ et $D = 1$.

a) On applique l'opérateur $(1 - B)$ à la série $(Y_t)$ lorsque le périodogramme estimateur de la pseudo-densité spectrale présente un pic important à l'origine et des valeurs «écrasées» ailleurs, et lorsque le corrélogramme estimateur de la pseudo-fonction d'autocorrélation présente des valeurs toutes positives qui décroissent lentement. On supprime ainsi une tendance.

**Exemple** : le dernier cours de cotation en Bourse de IBM à Paris du 31 avril 1976 à octobre 1976; source : Makridakis et Wheelwright (1978) (*cf.* figure 3).

b) On applique l'opérateur $(1 - B)^S = (1 - B)(1 + B + \ldots + B^{s-1})$ à la série $Y_t$ lorsque le périodogramme présente un pic important à l'origine et des pics en $1/S, 2/S, \ldots, (S-1)/S$, et lorsque le corrélogramme présente des valeurs positives en décroissance faible en $S, 2S, 3S$.... On peut noter éventuellement la présence de sous-harmoniques.

On supprime ainsi une tendance d'ordre 1 et une saisonnalité d'ordre $S$.



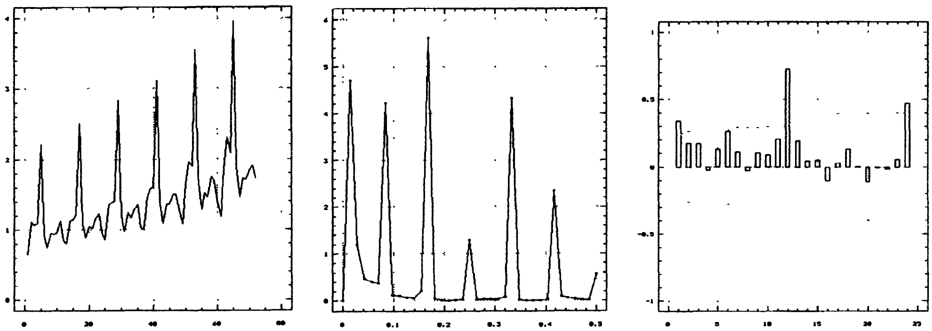
a) b) c)

FIGURE 3

*Dernier cours de cotation en Bourse de IBM à Paris*
*a) série initiale; b) périodogramme; c) corrélogramme*

**Exemples**

$\alpha$) Chiffre d'affaires d'un grand magasin (série mag). : 144 observations mensuelles. Source ASU 1984 (*cf.* figure 4).



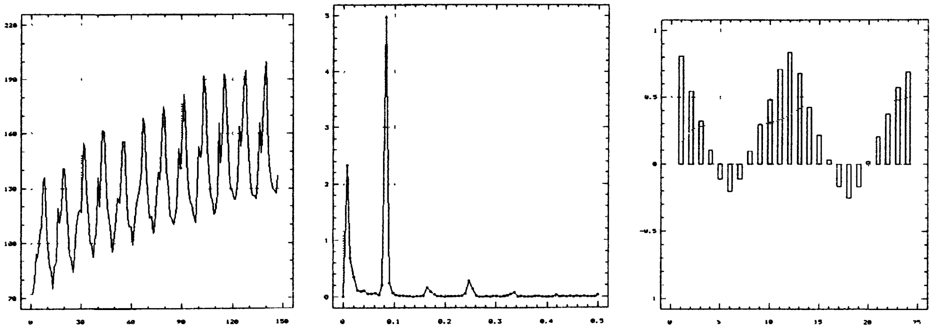
a) b) c)

FIGURE 4

*Chiffre d'affaires d'un grand magasin : données mensuelles*
*a) série initiale; b) périodogramme; c) corrélogramme*

$\beta$) Indice de circulation routière : 147 observations mensuelles. Source : Ministère des Transports (*cf.* figure 5).

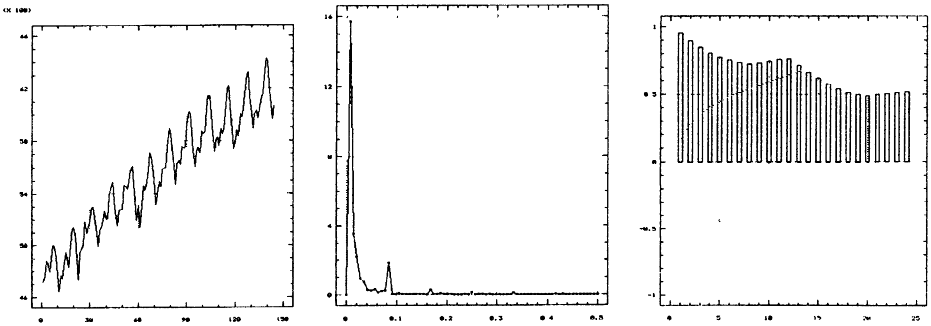
c) On applique l'opérateur $(1-B)(1-B^S) = (1-B)^2(1+B+\ldots+B^{S-1})$ lorsque le périodogramme présente des pics à l'origine et des pics en $1/S, 2/S, \ldots$ Mais le pic situé à l'origine est beaucoup plus important que ceux en $1/S, 2/S, \ldots$, et lorsque le corrélogramme a une allure très caractéristique avec des valeurs positives présentant une forme géométrique d'amplitude décroissante. On supprime ainsi une tendance d'ordre 1 ou 2 et une saisonnalité d'ordre $S$.



a) b) c)

FIGURE 5
*a) série initiale; b) périodogramme; c) corrélogramme*

**Exemple** : série Air-line de Box et Jenkins : Logarithme du nombre mensuel de passagers en trafic aérien international de janvier 1949 à décembre 1960. Source : Box et Jenkins 1976 (*cf.* figure 6).



a) b) c)

FIGURE 6
*Série Air-line : logarithme du nombre mensuel de passagers en trafic aérien international*
*a) série initiale; b) périodogramme; c) corrélogramme*

## Remarques

1) La comparaison des dispersions de la série initiale et de la série différenciée permet d'apprécier l'effet de stationnarisation. Certains logiciels, à cette étape, proposent une régression de la série différenciée sur son passé et testent à l'aide du test de Portmanteau (4) ou de Ljung-Box (20) si les résidus de cette régression peuvent être considérés comme un bruit blanc (c'est en particulier le cas du prologiciel Mandrake).

2) L'étude d'une surdifférenciation peut se faire avec le corrélogramme de la fonction d'autocorrélation inverse fournie par certains logiciels. (ex. : Anar, Mandrake, SAS, Shazam...) Supposons la série $(Y_t)$ modélisée suivant le modèle

ARMA suivant :

$$\Phi(B)Y_t = \Theta(B)\varepsilon_t$$

où $\Phi(B)$ et $\Theta(B)$ sont les polynômes autorégressif et moyenne mobile et où $\varepsilon_t$ est un bruit blanc.

La fonction d'autocorrélation inverse de $Y_t$ est la fonction d'autocorrélation du processus $\tilde{Y}_t$ modélisé par

$$\Theta(B)\tilde{Y}_t = \Phi(B)\eta_t$$

où $\eta_t$ est un bruit blanc de même variance que $\varepsilon_t$.

Si l'on surdifférencie $Y_t$ et que l'on considère les modélisations suivantes :

1) $Z_t = (1 - B)Y_t$

$Z_t$ aura la modélisation $\Phi(B)Z_t = (1 - B)\Theta(B)\varepsilon_t$

$\tilde{Z}_t$ la modélisation $\Theta(B)(1 - B)\tilde{Z}_t = \Phi(B)\eta_t$ la fonction d'autocorrélation inverse de $Z_t$ aura donc les caractéristiques de la fonction d'autocorrélation d'une série qui devient stationnaire par différenciation $1 - B$

2) $Z_t = (1 - B^S)Y_t$

$Z_t$ aura la modélisation $\Phi(B)Z_t = (1 - B^S)\Theta(B)\varepsilon_t$

$\tilde{Z}_t$ la modélisation $(1 - B^S)\Theta(B)\tilde{Z}_t = \Phi(B)\eta_t$ la fonction d'autocorrélation inverse de $Z_t$ aura donc les caractéristiques de la fonction d'autocorrélation d'une série qui devient stationnaire par différentiation $1 - B^S$.

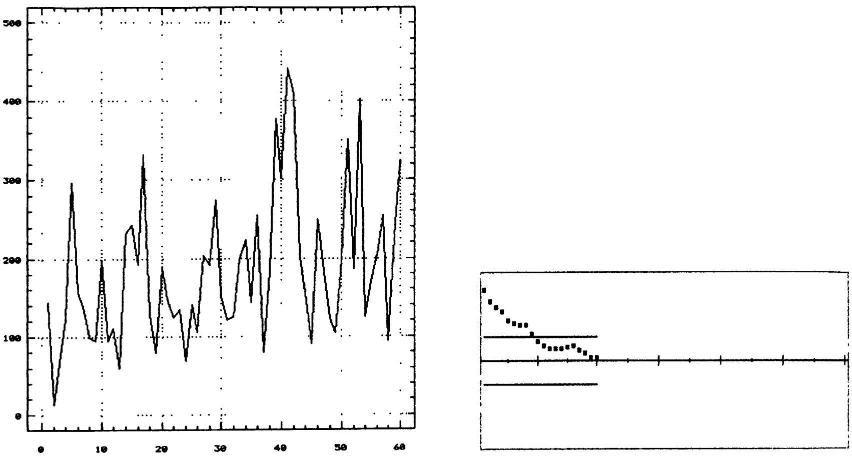
**Exemples**

a) Surdifférenciation, due à l'opérateur $1 - B$, de la série chiffre d'affaires d'un grand magasin.

Le graphe de la série différenciée en $1 - B^{12}$, comme le montre la figure 7 en a), est le graphe d'une série stationnaire. Cette différenciation suffit donc à stationnariser la série. Si on différencie à nouveau en utilisant la différenciation $1 - B$, la fonction d'autocorrélation inverse de cette série deux fois différenciée présente, comme le montre la figure 7 en b), des valeurs toutes positives qui décroissent lentement avec le temps. On a donc les caractéristiques d'une série surdifférenciée par l'opérateur $1 - B$.

b) Surdifférenciation, due à l'opérateur $1 - B^S$ (avec $S = 4$), de la série du dernier cours de cotation en Bourse de IBM.

Le graphe de la série différenciée en $1 - B$, comme le montre la figure 8 en a) est le graphe d'une série stationnaire. Cette différenciation suffit donc à stationnariser la série. Si on différencie à nouveau en utilisant la différenciation $1 - B^4$, la fonction d'autocorrélation inverse, comme le montre la figure 8 en b), a des valeurs positives décroissant lentement pour les décalages temporels multiples de 4. La série de cotation en Bourse de IBM différenciée par $(1 - B)(1 - B^4)$ est donc surdifférenciée. La surdifférenciation étant due à l'opérateur $1 - B^4$.



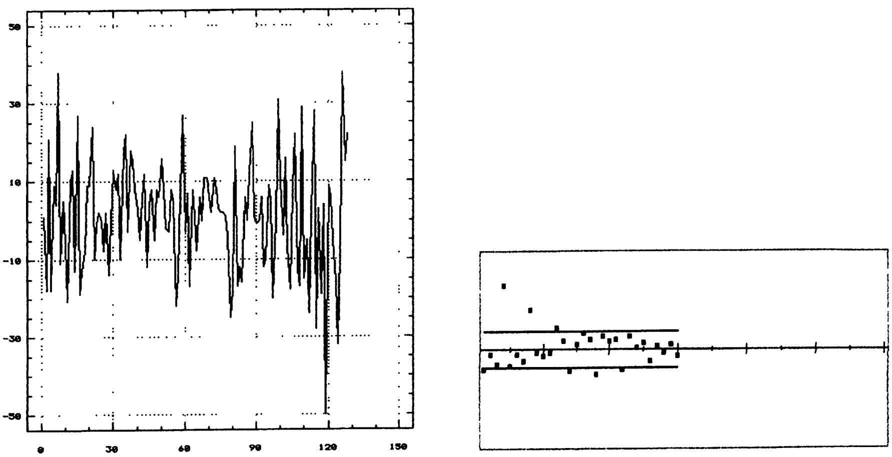

a) b)

FIGURE 7

*Chiffre d'affaires d'un grand magasin : données mensuelles*

*a) série différenciée en* $1 - B^{12}$

*b) fonction d'autocorrélation inverse de la série différenciée en* $(1 - B)(1 - B^{12})$




a) b)

FIGURE 8

*Dernier cours de cotation en Bourse de IBM à Paris*

*a) série différenciée en* $1 - B$

*b) fonction d'autocorrélation inverse de la série*

*différenciée en* $(1 - B)(1 - B^4)$

### B. Modélisation ARMA

Supposons que la série considérée soit stationnaire et de longueur $n$, alors l'étude conjointe de la fonction d'autocorrélation et de la fonction d'autocorrélation partielle permet de proposer une modélisation ARMA.

#### *1. Modélisation moyenne mobile*

L'ordre de la série moyenne mobile sera déterminé à partir de la fonction d'autocorrélation $\rho(h)$. En effet si $Y_t \sim MA(q)$ alors $\rho(h) = 0$ pour $h \geqslant q+1$. Dans ce cas là, l'estimateur empirique $\widehat{\rho}(h)$ de $\rho(h)$, suit une loi asymptotiquement normale d'espérance nulle et de variance

$$\frac{1}{n}\left(1 + 2\sum_{j=1}^{q} \rho^2(j)\right)^{1}.$$

On obtient ainsi les intervalles de confiance dits de Bartlett [5-b] pour cette fonction d'autocorrélation empirique (logiciels : Statgraphics, ANAR, Mandrake, SAS, Shazam, TSP). Ces intervalles de confiance ne sont pas donnés dans ITSM, RATS, Eviews.

#### *2. Modélisation autorégressive*

L'ordre de la série autorégressive est déterminé à partir dela fonction d'autocorrélation partielle $r(h)$.

En effet si $Y_t \sim AR(p)$ alors $r(h) = 0$ pour $h \geqslant p+1$. Dans ce cas là, l'estimateur empirique $\widehat{r}(h)$ de $r(h)$ suit asymptotiquement une loi normale d'espérance nulle et de variance $\frac{1}{n}$. On obtient alors facilement les intervalles de confiance dit de Quenouille [5-e] pour cette fonction d'autocorrélation empirique.

#### *3. Modélisation ARMA*

Une série $Y_t$ de modèle ARMA $(p, q)$ ayant une modélisation AR $(\infty)$ et MA $(\infty)$, on est amené à utiliser la modélisation ARMA dès que l'écriture AR ou MA comporte un nombre important de termes significatifs. Pour des modèles ARMA $(p, q)$ avec $p$ et $q$ inférieurs à deux, on connait la configuration de la fonction d'autocorrélation et de la fonction d'autocorrélation partielle théoriques. On examine donc si les valeurs significatives du corrélogramme et du corrélogramme partiel correspondent à cette configuration.

Autrement, le plus souvent, on modélise en deux étapes :

*Etape 1* – On estime un autorégressif d'ordre $p$ sur la série. On analyse les résidus de cette modélisation qui présentent alors non une structure de bruit blanc

---

1 Dans le cas d'un bruit blanc fort.

mais de moyenne mobile. On détermine l'ordre de cette moyenne mobile. Soit $q$ cet ordre.

*Etape 2* – On estime un modèle ARMA $(p, q)$ sur la série.

Il existe aussi la méthode dite «Méthode du Coin» basée sur l'ordre minimum $p$ des équations de récurrence satisfaites par les coefficients du développement MA $(\infty)$ de la série à partir d'un certain indice et de l'ordre minimum $q$ des équations de récurrence satisfaites par les coefficients du développement AR $(\infty)$ de la série à partir d'un certain indice (17).

**Remarque**

Lors de la modélisation on voit se différencier les logiciels qui admettent des modèles «percés» et ceux qui ne les admettent pas[2] – (un modèle percé est un modèle pour lequel on peut imposer aux paramètres $\theta_i$ de l'écriture moyenne mobile et aux paramètres $\phi_j$ de l'écriture autorégressive d'être contraints à zéro) – ces derniers, dans le cas de séries saisonnières, ne permettent que l'estimation de modèles multiplicatifs[3]

## *C. Estimation des paramètres*

Pour estimer les paramètres, la plupart des logiciels utilisent une approximation du maximum de vraisemblance.

On distingue plusieurs types de procédure :

– Estimation en deux étapes : estimation préliminaire, puis estimation fine : ANAR, Mandrake, ITSM.

– Estimation en une étape : Eviews, Rats, SAS, Shazam, Statgraphics, TSP.

Les valeurs initiales des paramètres sont soit des valeurs petites arbitraires (ITSM, Shazam) soit des valeurs calculées par le logiciel en fonction de la fonction d'autocorrélation (SAS, Statgraphics, TSP, Eviews).

Certains logiciels permettent l'introduction de valeurs initiales arbitraires par l'utilisateur (Rats, SAS, TSP) ou des valeurs déterminées à partir d'estimations précédentes (Eviews).

Les logiciels proposent en général plusieurs méthodes d'optimisation, le choix du critère de convergence et le nombre maximum d'itérations.

En ce qui concerne la présence d'une constante dans la modélisation :

– soit la constante est toujours présente : Anar, Mandrake.

– soit, il faut demander sa présence : ITSM, Rats, SAS, Shazam, Statgraphics, TSP.

---

[2] On suppose la série stationnaire $Y_t$ écrite : $Y_t - \varphi_1 Y_{t-1} \dots \varphi_p Y_{t-p} = \theta_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} \dots \theta_q \varepsilon_{t-q}$ avec $\varepsilon_t$ bruit blanc.

[3] Modèle multiplicatif général pour une série $Y_t$ de saisonnalité de période $S$ :

$$(1 - \varphi_1 B \dots - \varphi_p B^p)(1 - \varphi_{1s} B^S \dots - \varphi_{PS} B^{SP}) Y_t$$
$$= \theta_0 + (1 - \theta_1 B \dots - \varphi_q B^q)(1 - \theta_{1S} B^S \dots - \theta_{QS} B^{SQ}) \varepsilon_t$$

## *D. Tests de modèles*

Pour l'étude de la validité et de la pertinence des modèles, les logiciels proposent toujours des tests de signification des paramètres et une étude des résidus plus ou moins complète. Certains, de plus, proposent une comparaison entre prévisions internes et valeurs observées.

### *1. Etude des résidus*

On a en général : le graphe des résidus, la fonction d'autocorrélation, la fonction d'autocorrélation partielle, le test de Portmanteau (4) ou de Ljung-Box (20) (ou des statistiques équivalentes pour Anar et Mandrake). Souvent un test de normalité (ajustement graphique du «Normality Plot» (Statgraphics) ou le test de Jarque-Bera (15) utilisant le coefficient d'asymétrie (skewness) et le coefficient d'aplatissement (kurtosis) (Eviews, Shazam)).

Des logiciels proposent une étude plus poussée de la structure résiduelle :

– Périodogramme cumulé (Statgraphics, ITSM, Rats, TSP) [5-g].

– Test de Fisher-Kappa (ITSM, Rats, TSP) [5-g].

– Test de corrélation LM de Breusch-Godfrey (Eviews, Rats, Shazam, TSP) (13), (14).

– Test d'hétéroscédasticité conditionnelle Arch LM test (Eviews Rats, Shazam) (12).

### *2. Stabilité du modèle*

Certains logiciels intègrent les tests de Chow (Breakpoint et Forecast) pour juger de la stabilité du modèle (Eviews, Shazam...) (6) et (7).

### *3. Etude de prévisions*

Plusieurs possibilités : certains logiciels permettent la comparaison entre les valeurs observées de la série et des prévisions internes données par le modèle estimé (Anar, Mandrake, Rats, SAS, Shazam, Eviews).

D'autres ne donnent une estimation des prévisions qu'à partir de la dernière valeur observée de la série (ITSM, Statgraphics).

## *E. Points spécifiques*

Pour la détection des valeurs aberrantes qui peuvent influencer la modélisation et les prévisions finales certains logiciels proposent des procédures (ex. : Mandrake).

Lorsque certaines séries peuvent être considérées comme la superposition d'un processus faiblement stationnaire et d'une tendance qui ne peut être retirée par une simple différenciation, certains logiciels donnent la possibilité de retirer la tendance, de modéliser la partie ARIMA de la série, puis de réinsérer la tendance et de faire des prévisions (ex. : Mandrake, SAS).

Certains logiciels offrent la possibilité d'introduire dans la modélisation ARIMA des fonctions d'intervention, des fonctions de transfert et d'écrire des équations de cointégration (ITSM, Mandrake, Rats, SAS, Shazam); d'autres permettent d'introduire une modélisation de type ARCH ou GARCH (Rats, SAS, Shazam, TSP) pour les résidus. Enfin certains logiciels proposent de plus une modélisation ARIMA multivariée (ITSM, Rats, SAS, Shazam, TSP).

## 2. Présentation de logiciels[4]

Pour les différents logiciels, nous ne présentons que les procédures traitant de la modélisation ARIMA.

Les logiciels auxquels nous avons eu l'occasion de nous intéresser ont pour la plupart une version fonctionnant sous Dos (seuls Eviews et Shazam n'existent que sous Windows). Mais ils ont éventuellement aussi une version fonctionnant sous Windows (Statgraphics, ITSM).

Pour Rats, SAS, Shazam, TSP, il faut écrire, sous éditeur, un programme d'enchaînement des procédures; programme qui peut être complété par des programmes personnels.

Pour Statgraphics, Anar, Mandrake, ITSM, on sélectionne, par la touche «enter» ou par cliquage, des procédures. On a, de plus, à sa disposition des feuilles de calcul. Pour Eviews, le logiciel fonctionne avec une fenêtre de commandes et des boîtes de dialogues. Ce logiciel offre quelques possibilités de programmer.

### *A. MANDRAKE*

Ce logiciel concerne uniquement l'analyse et la prévision des séries temporelles univariées basées sur la méthode de Box et Jenkins avec ou sans variables exogènes.

Il a trois groupes de traitement :

– Modélisation automatique.

– Utilitaires prétraitements et post-traitements.

– Modélisation pas à pas.

#### *1. Prétraitements*

– Regroupement des données et échantillonnage.

– Génération de séries par expression arithmétique.

– Remplacement des valeurs manquantes et/ou aberrantes.

– Détection de rupture de tendance.

– Correction de tendance.

[4] Le logiciel Anar, à quelques variantes près, correspond à la modélisation «pas à pas» de Mandrake et ne sera donc pas détaillé ici.

*2. Post-traitements*

– Modélisation ARMAX[5]

– Ajustement saisonnier : décomposition de la série en tendance, saisonnalité et composante irrégulière.

*3. Modélisation automatique*

Elle permet d'obtenir plusieurs modèles classés suivant le critère BIC (1). On obtient pour chaque modèle ses caractéristiques, les étapes successives de la modélisation et des prévisions avec encadrement. Un module de relance permet à l'utilisateur de proposer d'autres modèles, d'autres différenciations, et de les comparer avec les modèles proposés.

Les utilitaires permettent de stocker des résultats pour reprendre ou achever une étude antérieure.

*4.* La modélisation pas à pas comporte deux étapes d'estimation : l'estimation préliminaire et l'estimation fine.

L'estimation préliminaire permet de déterminer la longueur de la série autorégressive ou (et) moyenne mobile et les coefficients $\theta_i$ ou $\varphi_j$ significatifs. L'estimation des coefficients est effectuée par la résolution d'un système linéaire qui étend aux modèles ARMA la méthode de Yule-Walker [5-e]. Il y a toujours vérification de l'inversibilité du modèle obtenu. Si le modèle n'est pas inversible il est remplacé par un modèle proche inversible. Pour tout coefficient inférieur en module à son écart-type, le logiciel propose de ne pas le garder dans la modélisation.

L'estimation fine : on garde si on le désire le modèle retenu précédemment et les valeurs des paramètres de l'estimation préliminaire comme valeurs initiales ou bien on peut supposer un autre modèle (éventuellement un modèle multiplicatif) avec des valeurs initiales pour les paramètres. On peut aussi imposer des modèles «percés».

L'estimation est basée sur la méthode du maximum de vraisemblance où le maximum de la vraisemblance est approché par la somme des carrés des résidus. Le calcul d'optimisation utilise l'algorithme de Gauss-Markov. Le logiciel donne

– les valeurs estimées des coefficients avec leurs écarts-types,

– la part de la variance de la série différenciée expliquée par le modèle,

– l'écart-type résiduel,

– les autocorrélations des résidus jugées significatives,

– un test de Fisher de stationnarité de la variance des résidus,

– un test de «Randomness» pour les résidus (test basé sur les séquences successives).

On peut obtenir : le graphe, les fonctions d'autocorrélation, d'autocorrélation partielle, d'autocorrélation inverse des résidus.

---

[5] Modèle de la forme : $\phi(B)Y_t = \sum_i \Psi_i(B)X_{it} + \theta(B)\varepsilon_t$ c'est-à-dire une écriture autorégressive de $Y_t$ en fonction de variables exogènes $X_{it}$ avec une équation d'erreur moyenne mobile.

Les prévisions sont données avec encadrement; il y a possibilité de dates de départs successives avec comparaison des valeurs prévisionnelles et des valeurs observées.

## *B. RATS*

L'analyse et la prévision des séries temporelles univariées par la méthode de Box et Jenkins n'est qu'une petite partie de ce logiciel qui est avant tout un logiciel de régression (régression linéaire des moindres carrés ordinaires ou pondérés, variables instrumentales, doubles moindres carrés, «stepwise regression», «ridge régression», régression linéaire avec correction pour des erreurs autocorrélées d'ordre 1, modèles Logit et Probit, régression non linéaire par la méthode des moindres carrés, estimation de systèmes d'équations utilisant la méthode des moindres carrés généralisée (Generalized Least Squares) ou les triples moindres carrés...).

### *Modélisation ARIMA*

La procédure «Boxjenk», procédure d'estimation et de modélisation ARIMA des séries univariées, permet d'intégrer dans la modélisation une fonction de transfert ou une fonction d'intervention. (Après avoir éventuellement testé la causalité).

Il existe, de plus, une procédure de modélisation multivariée VAR ainsi qu'une procédure d'estimation de Filtre de Kalman. L'utilisation des instructions de tests («Hypothesis testing») ou de méthode de maximum de vraisemblance («Maximize») permet de tester l'homoscédasticité des résidus et d'estimer des modèles ARCH.

### *Procédure BOXJENK*

Pour cette procédure, il y a une seule étape d'estimation. Les valeurs initiales des paramètres sont calculées soit par le logiciel, soit introduites par l'utilisateur.

Les valeurs initiales proposées sont données : pour un modèle autorégressif, par la résolution des équations de Yule-Walker; pour un modèle moyenne mobile par l'algorithme de Gauss-Seidel.

Le logiciel pour le calcul utilise l'algorithme de Gauss-Markov et, si l'option «Startup» n'est pas précisée, ne fait pas de calculs de «Backforecasting»[6] et considère les premières valeurs résiduelles comme nulles. Dans le cas d'un modèle à moyenne mobile, le logiciel vérifie que la partie moyenne mobile est inversible. Si ce n'est pas le cas, il ne donne pas d'estimation de modèle.

Le logiciel donne les valeurs estimées des coefficients avec leurs écarts-types et leurs statistiques de Student. Il n'est pas possible d'imposer des modèles «percés» au départ mais il est possible d'effectuer des tests d'hypothèses sur les coefficients en utilisant les instructions «Test» et «Restrict».

---

[6] Le «backforecasting» consiste à prolonger la série dans le passé suivant le modèle envisagé et d'obtenir ainsi des valeurs résiduelles dès la première valeur observée de la série.

Pour juger de l'adéquation d'un modèle, le logiciel donne le coefficient DW de Durbin-Watson (10), (11)[7] , la statistique $Q$ de Ljung-Box. Pour tester la corrélation d'ordre $n$ sur les résidus, il faut écrire le test LM de Godfrey et Breusch (13) et (14).

Pour l'étude des résidus, on dispose du graphe de la fonction d'autocorrélation, de la fonction d'autocorrélation partielle.

On peut obtenir le périodogramme cumulé et des tests de normalité en écrivant les instructions correspondantes.

Les prévisions se font à partir de n'importe quelle date de départ et pour n'importe quel horizon; mais elles sont données sans encadrement et elles ne sont pas comparées aux valeurs observées.

### *C. TSP International*

Ce logiciel est très développé en régression linéaire et non linéaire (régression linéaire des moindres carrés ordinaires ou pondérés, variables instrumentales, doubles moindres carrés, régression linéaire avec correction pour des erreurs autocorrélées d'ordre 1, estimation d'un système d'équations non linéaires par la méthode des simples, doubles ou triples moindres carrés, méthode généralisée des moments, modèles Tobit, Probit, Logit...). Pour les séries temporelles, le logiciel T.S.P. permet d'étudier outre les modèles ARIMA univariés, les filtres de Kalman, les modèles VAR avec les tests de causalité, les modèles ARCH et GARCH. Il propose de plus les tests de racines unitaires de Dickey Fuller (8), (9) et de co-intégration de Johansen (16).

#### *Modélisation ARIMA*

##### *1. Procédure «Bjident» pour l'identificaton du modèle*

Cette procédure donne la fonction d'autocorrélation, la fonction d'autocorrélation partielle de la série différenciée et de la série originelle mais ne donne pas le périodogramme. Pour chaque valeur de la fonction d'autocorrélation, on a l'écart-type et la $Q$-valeur (rapport de la valeur estimée et de son écart-type). Le graphe est donné avec les intervalles de Bartlett [5-B].

##### *2. Procédure «Bjest» pour l'estimation du modèle*

Le logiciel propose l'utilisation du «Backforecasting». L'utilisateur peut soit laisser le logiciel déterminer les valeurs initiales des paramètres soit proposer des valeurs initiales pour certains paramètres et des valeurs fixes pour les autres (par exemple zéro) ce qui permet d'utiliser des modèles «percés». La procédure d'optimisation utilise le processus iteratif de Gauss-Marquard. Les coefficients estimés sont donnés avec leurs écarts-types.

La fonction d'autocorrélation des résidus est donnée, avec pour chaque valeur, la $Q$-statistique et la $P$-valeur. On dispose aussi du périodogramme cumulé des résidus.

[7] Le DW de Durbin-Watson n'est pas très adapté car il correspond à un test d'autocorrélation d'ordre 1 sur des résidus gaussiens estimés à partir de la méthode des moindres carrés ordinaires.

Pour juger de l'adéquation du modèle, on dispose du coefficient de détermination $R^2$ et de la statistique de Durbin-Watson.

*Procédure «Bjfrcst» pour les prévisions*

Les prévisions peuvent être obtenues à partir de plusieures dates de départ. Elles sont données avec leurs intervalles de confiance et sont comparées s'il y a lieu avec les valeurs observées.

On peut enchaîner les deux procédures «Bjest» et «Bjrcst».

### *D. ITSM*

Ce logiciel est un logiciel d'analyse et de prévisions de séries temporelles d'après la méthode de Box et Jenkins. Il comporte huit programmes (d'inégale longueur); les trois plus importants sont «Pest», «Trans» et «Arar».

«Pest» permet l'étude, la modélisation, la prévision des séries univariées.

«Trans» permet l'introduction et l'estimation de fonctions de transfert et «Arar» l'étude et la prévision de séries multivariées écrites sous la forme autorégressive.

Parmi les autres programmes, il y a en particulier «Longmem» qui permet d'étudier des modèles à mémoire longue.

**– Programme Pest :**

Deux étapes d'estimation : l'estimation préliminaire et l'estimation fine.

L'estimation préliminaire permet :

– de déterminer la longueur de la série autorégressive et/ou moyenne mobile,

– de déterminer les paramètres significatifs.

MA $(q)$ : les estimateurs initiaux sont donnés par la méthode d'innovation de Brockwell et Davis [5-a].

AR $(p)$ : les estimateurs initiaux sont donnés par les équations de Yule-Walker.

ARMA $(p, q)$ le modèle est d'abord écrit sous forme moyenne mobile longue, les paramètres étant alors estimés par la méthode d'innovation, puis les estimateurs du modèle ARMA $(p, q)$ sont donnés par identification.

L'estimation fine permet d'introduire éventuellement une structure multiplicative en écrivant des contraintes sur les coefficients. On peut, à cette étape, imposer à certains paramètres de ne pas figurer et à d'autres paramètres de prendre des valeurs fixées. On a le choix entre deux méthodes d'estimation (méthode du Maximum de vraisemblance, méthode des moindres carrés).

Le logiciel donne les valeurs estimées avec leurs écarts-types. On dispose aussi du test de Portmanteau, de la statistique de BIC, des tests de «randomness» sur les résidus, des tests de Kappa-Fisher et Bartlett-Kolmogorov-Smirnov [5-h], de l'histogramme des résidus,de la fonction d'autocorrélation et leur fonction d'autocorrélation partielle des résidus.

Les prévisions se font en fin d'observation de la série. On dispose tout d'abord des valeurs prévisionnelles du processus ARMA centré, puis du processus non centré (éventuellement après inversion de la transformation de Box et Cox). Le logiciel fournit les éléments de calcul pour construire les intervalles de prédiction.

De plus pour juger de l'adéquation du modèle, le logiciel donne la fonction d'autocorrélation, la fonction d'autocorrélation partielle, le périodogramme ainsi que l'écritre Ar $(\infty)$ et MA $(\infty)$ du modèle.

Il est à noter que le logiciel ne fournit pas pour la fonction d'autocorrélation les intervalles de Bartlett.

### *E. Shazam*

Shazam est un logiciel d'économétrie. La modélisation ARIMA n'est qu'une très petite partie de ce logiciel (régression des moindres carrés ordinaires ou pondérés, «stepwise regression», modèles à erreurs corrélées d'ordre 1 ou 2, à erreurs ARMA – modèle de Box-Cox autorégressif, modèle de cointégration et tests de racine unitaire – modèles hétéroscédastiques – modèles Logit, Probit, Tobit – Systèmes d'équations avec variables instrumentales... Il propose les tests unitaires de Dickey-Fuller, le test ARCH, les tests de co-intégration de Johansen.

*Modélisation ARIMA*

Pour la procédure Arima trois sous procédures :

*Identification*

Cette procédure donne la fonction d'autocorrélation avec les intervalles de Bartlett, la fonction d'autocorrélation inverse (le maximum de décalage temporel autorisé est 20), la fonction d'autocorrélation partielle et teste si la série est un bruit blanc (test de ljung-Box-Pierce) (20). L'option All permet de faire ces opérations pour toutes les différenciations envisagées pour la série. Cette étude peut être complétée éventuellement, en utilisant l'instruction «Coint» qui donne les tests unitaires.

*Estimation*

Lors de cette procédure, l'utilisateur peut proposer des valeurs initiales pour les paramètres et imposer à certains d'être nuls ce qui permet d'utiliser des modèles «percés». Lorsqu'il n'y a pas de valeurs initiales proposées, le logiciel démarre son processus itératif de calcul avec des valeurs voisines de zéro. Lors du processus itératif les calculs se font avec l'algorithme de Marquardt et une méthode de «backforecasting». Les coefficients estimés sont donnés avec leurs écarts-types. La fonction d'autocorrélation des résidus est donnée, avec, pour chaque valeur, la $Q$-statistique et la $P$-valeur ainsi que la corrélation croisée des résidus avec la série différenciée. Pour juger de l'adéquation du modèle, on dispose de plus du coefficient de détermination $R^2$, des critères AIC et BIC (1). De plus avec l'instruction

«Chowtest», on peut étudier la stabilité du modèle et la stabilité de la variance résiduelle et avec l'instruction «Het», étudier l'hétéroscédasticité.

*Forecasting*

Pour le modèle spécifié, cette procédure permet de faire des prévisions à partir de n'importe quelle date de départ et pour n'importe quel horizon. Elles sont données avec des intervalles de prédiction et sont comparées aux valeurs observées s'il y a lieu.

### F. Statgraphics

Statgraphics est un logiciel de statistique non spécialisé qui comporte plusieurs parties

– statistique descriptive, estimation et tests,

– Anova et régression,

– séries temporelles et contrôle de qualité,

– variables catégorielles (mesures d'association et modèle Loglinéaire),

– méthodes non paramétriques et analyse des données.

En séries temporelles, il propose les méthodes de lissage et de décomposition en tendance, saisonnalité et composante irrégulière ainsi que la modélisation ARIMA.

*Modélisation ARIMA*

Cette modélisation est dans la rubrique «Time series analysis».

On peut indépendamment pour la série elle-même ou la série différenciée demander le graphe, la fonction d'autocorrélation avec les intervalles de Bartlett, la fonction d'autocorrélation partielle, la transformation de Box et Cox, le périodogramme, le périodogramme cumulé.

Pour la modélisation ARIMA proprement dite, Statgraphics n'admet pas de modèles percés. Par contre, il propose à l'utilisateur d'utiliser ou non la méthode du «backforecasting». Les paramètres estimés sont donnés avec leurs écarts-types, le $T$ de Student, le «probability level». Il donne, pour le modèle, la variance des résidus et la valeur du test de Portmanteau.

*Etude des résidus*

Pour les résidus, il donne le graphe, la fonction d'autocorrélation, la fonction d'autocorrélation partielle, le périodogramme, le périodogramme cumulé.

*Prévisions*

Elles ne sont données qu'en fin d'observation de la série – mais avec leurs intervalles de prédiction.

Il n'est pas prévu de test de Chow de stabilité de modèle. Celui-ci peut cependant être mis en oeuvre par l'utilisateur en utilisant la feuille de calcul et les procédures adéquates; même remarque pour le test Arch.

### *G. Time Series Modelling and Forecasting de SAS/ETS*

Cette partie du module de SAS/ETS concerne exclusivement l'analyse des séries temporelles. En plus des modèles ARIMA qui peuvent être utilisés avec des fonctions d'intervention et des fonctions de transfert, on trouve les modèles de régression avec erreurs corrélées, la méthode des moyennes mobiles du Census X11, la méthode Espace-Etat des filtres de Kalman et les lissages exponentiels.

*Modélisation ARIMA*

Procédure *Identify* pour l'identification du modèle :

Cette procédure donne la fonction d'autocorrélation avec les intervalles de Bartlett, la fonction d'autocorrélation inverse, la fonction d'autocorrélation partielle et teste si la série est un bruit blanc (test de Ljung-Box).

*Pour l'estimation*, deux possibilités :

Si le modèle envisagé est autorégressif : procédure «Autoreg» qui propose quatre méthodes de calcul pour l'estimation des paramètres (Yule-Walker, Yule-Walker itéré [5-d], Moindres carrés non conditionnels, Maximum de vraisemblance) [5-f]. Avec l'option backstep, on peut retirer les paramètres non significatifs.

Autrement procédure Estimate de la procédure ARIMA. Cette méthode propose trois méthodes d'estimation (Moindres carrés, Moindres carrés non conditionnels et Maximum de Vraisemblance), et permet d'imposer des modèles «percés». Les coefficients estimés sont donnés avec leur matrice de variance-covariance et avec leurs $t$-ratio. La fonction d'autocorrélation des résidus est donnée avec des $Q$-statistique et des $P$-valeurs. Pour juger de l'adéquation du modèle on dispose du critère d'information d'Akaike et du critère de Schwartz [17-b], mais on a aussi la possibilité d'étudier le périodogramme des résidus avec la procédure «Spectral Analysis» et de tester si le résidu est un bruit blanc gaussien avec les tests de Fisher-Kappa et Bartlett-Kolmogorov-Smirnov.

*Prévision* :

Elle peut se faire avec la procédure ARIMA, avec la procédure «Autoreg» ou avec la procédure «Forecast». Ces trois procédures donnent les prévisions avec leurs intervalles de prédiction.

### *H. Eviews*

Eviews est un logiciel d'économétrie fonctionnant sous Windows et faisant suite au logiciel Micro TSP créé en 1981. Il fonctionne par l'intermédiaire d'une fenêtre de commandes et de boîtes de dialogues. Il propose la régression simple, double, pondérée, avec variables à retards échelonnés; les modèles à corrections d'erreurs ARMA; les modèles Logit, Probit; les systèmes d'équations; les modèles VAR. Il propose de nombreux tests : tests unitaires de Dickey-Fuller, test LM de Breusch-Godfrey, tests de Chow, test ARCH, test Reset de Ramsey (25), tests de causalité de Granger (21), test de cointégration de Johansen.

*Modélisation ARIMA*

*Stationnarisation*

Dans la fenêtre de commandes avec les instructions «Ident» et «Uroot» il y a possibilité d'une part d'avoir la fonction d'autocorrélation avec les $Q$ statistiques (mais sans les intervalles de Bartlett), la fonction d'autocorrélation partielle; d'autre part les tests unitaires pour détecter les problèmes de tendance. Le périodogramme n'est pas prévu.

*Identification – Estimation*

Dans la boîte de dialogues «Equation», on écrit le modèle ARIMA supposé. On peut écrire un modèle «percé». Les valeurs initiales des paramètres sont normalement calculées par Eviews. Mais on peut donner des valeurs initiales qui sont un pourcentage des valeurs obtenues lors d'une estimation précédente.

On obtient les paramètres estimés avec leurs écarts-types et leurs $T$-statistiques ainsi que la matrice de variance et covariance. Le logiciel donne les racines des polynômes $\Phi(B)$ et $\Theta(B)$ ce qui permet de voir si le modèle estimé est bien inversible. Le logiciel donne le coefficient de détermination, les critères d'Akaike et de Schwartz. Il compare la série différenciée et la série ajustée sur la série différenciée.

Pour les résidus, il donne la fonction d'autocorrélation (et les $Q$ statistiques associées), la fonction d'autocorrélation partielle, l'histogramme, le test de normalité de Jarque-Bera, le test de corrélation LM de Breusch-Godfrey, le test ARCH d'hétéroscédasticité (mais ne prévoit pas de modélisation ARCH).

Pour l'étude de la stabilité du modèle : les tests de Chow (Forecast et Breakpoint) ainsi que le test des Cusum [17-c].

*Prévisions*

Les prévisions se font à partir de n'importe quelle date de départ, on a graphiquement les intervalles de prédiction. Mais, il n'est pas prévu d'étude systématique de comparaison dans le cas de prévisions internes des valeurs observées et des valeurs prédites.

## Conclusion

Les logiciels offrent des outils différents pour étudier la différenciation nécessaire (fonction d'autocorrélation inverse, tests de racines unitaires, périodogrammes). Pour l'estimation de la partie moyenne mobile du modèle ARIMA ils ont des procédures qui amènent en général à des estimateurs différents. Pour l'étude de la pertinence des modèles, ils permettent en général de comparer réalisations et prévisions internes. Mais ils ne sont que quelques-uns à donner directement des tests de stabilité de modèle. Ils proposent tous au minimum pour les résidus : le graphe, la fonction d'autocorrélation et la fonction d'autocorrélation partielle. Mais ils sont encore peu nombreux à donner directement un test d'homoscédasticité conditionnelle.

## Bibliographie


[1] Akaike H. (1969), «Fitting Autoregressive Models for Prediction», *Annals of the Institute of Statistical Mathematics*, 21, 243-247.

[2] Akaike H. (1977), «On Entropy Maximisation Principle», dans *Applications of Statistics*, Editeur Krisnaiah, 27-41, North-Holland.

[3] Box G. et Jenkins G. (1970), «Time Series Analysis : Forecasting and Control», Holden-Day.

[4] Box G. et Pierce D. (1970), «Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models», *Journal of the American Statistical Association* 65, 1509-1526.

[5] Brockwell P.J. et Davis R.A. (1991), «Time Series : Theory and Methods», Springer-Verlag a-172, b-223-224, c-239-240, d-239-245, e-241, f-256-258, g-338, h-339.

[6] Chow G. (1970), «Technological Change and the Demand for Computers", *American Economic Review*, Vol. 65.

[7] Chow G. (1984), «Random and Changing Coefficient Models» Handbook of Econometrics, Vol. 2 1213-1245. Griliches Z. et Intriligators Y. éditeurs Amsterdam North-Holland.

[8] Dickey D. et Fuller W. (1979), «Distribution of the Estimators for Autoregressive Time Series with a Unit Root», Journal of the American Statistical Association 74, 427-431.

[9] Dickey D. et Fuller W. (1981), «Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root», *Econometrica*, 49, 1057-1072.

[10] Durbin J. et Watson G. (1950) et (1951), «Testing for Serial Correlation in Least-Squares Regression», *Biometrika*, Vol. 37 et Vol. 38.

[11] Durbin J. (1970), «Testing for Serial Correlation in Least-Squares Regression when Some of the Regressors are lagged Dependent», *Biometrika*, Vol. 58.

[12] Engle R. (1982), «Autoregressive Conditional Heteroskedasticity, with Estimates of the Variance of United Kingdom Inflations», *Econometrica*, 50, 987-1008.

[13] Godfrey L. (1979), «Testing the Adequacy of a Time Series Model», *Biometrika*, 64, 67-72.

[14] Godfrey L. (1981), «On the Invariance of Lagrange Multiplier Test with Respect to Certain Changes in the Alternative Hypothesis», *Econometrica*, 49, 1443-1455.

[15] Jarque C. et Bera A. (1980), «Efficient Tests for Normality, Homoskedasticity and Serial Independence of Regression Residual» *Econometrics Letters*, 6, 255-259.

[16] Johansen S. (1991), «Estimation and Hypothesis Testing of Co-Integration Vectors in Gaussian Vector Autoregressive Models», *Econometrica* 59, 1551-1580.

[17] Gourieroux C. et Monfort A. (1990), «Séries temporelles et Modèles dynamiques», *Economica* a-202, b-239, c-714.

[18] Harvey L. (1989), «Forecasting, Structural Time Series Models and the Kalman Filter», Cambridge University Press.

[19] Hosking J.R. (1980), «Lagrange Multiplier Tests of Time Series Models», *Journal of the Royal Statistical Society*. B, 42, 170-181.

[20] Ljung G. et Box G. (1978), «On a Measure of Lack of Fit in Times Series Models», *Biometrika*, 65, 297-303.

[21] Lutkepohl H., «Introduction to Multiple Time Series Analysis», Springer-Verlag. 93-95 et 378-379.

[22] McLeod A. et Li W. (1983), «Diagnostic Checking ARMA Time Series Models using Squared Residual Autocorrelations», *Journal of Time Series Analysis*, 4, 269-273.

[23] Phillips P. (1987), «Time Series Regression with a Unit Root», *Econometrica*, 55, 277-301.

[24] Phillips P. et Perron P. (1988), «Testing for a Unit Root in Time Series Regression», *Biometrika*, 75, 335-346.

[25] Ramsey, (1969), «Tests for Specification Errors in Classical Linear Least Squares Regression Analysis», *Journal of the Royal Statistical Society*, Series B, 31, 350-371.

[26] Wei W. (1990), «Time Series Analysis Univariate and Multivariate Methods. Addison Wisley.

[27] White H. (1980), «A Heteroskedasticity-Consistent Covariance Estimator and a Direct Test for Heteroskedasticity», *Econometrica*, Vol. 48, 817-838.

[28] Whittle P. (1951), «Hypothesis Testing in Time Series», Analysis Hareferer, New-York.

[29] Said S. et Dickey D. (1984), «Testing for Unit Roots in ARMA of Unknown Order», *Biometrika* 71, 599-607.