JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

MICHEL TERRAZA

Prévision à court terme des séries temporelles économiques. L'estimation des modèles ARMA par l'algorithme de la plus grande pente

Journal de la société statistique de Paris, tome 122, nº 3 (1981), p. 143-161

<http://www.numdam.org/item?id=JSFS_1981__122_3_143_0>

© Société de statistique de Paris, 1981, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (http://publications-sfds.math.cnrs.fr/index.php/J-SFdS) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

PRÉVISION A COURT TERME DES SÉRIES TEMPORELLES ÉCONOMIQUES. L'estimation des modèles ARMA par l'algorithme de la plus grande pente

Michel TERRAZA
Assistant en sciences économiques. Laboratoire d'économétrie, Faculté de droit et des sciences économiques de Montpellier.

L'estimation des paramètres des processus ARMA s'est considérablement développée au cours de la dernière décennie. Les méthodes de calcul utilisées à cette fin sont généralement complexes. L'algorithme de MARQUARDT employé par BOX et JENKINS en est un exemple. Nous proposons dans cet article le recours à la technique de la plus grande pente pour réaliser cette estimation. L'algorithme de calcul est simple et malgré quelques inconvénients présente des avantages certains : adaptation des coefficients du modèle ARMA aux modifications lentes de la structure de la chronique, prévision de séries temporelles courtes. A l'aide d'une chronique économique nous montrons alors que la qualité de nos prévisions est comparable à celle fournie par la démarche de BOX et JENKINS.

The estimation of the parameters of the ARMA processes considerably developed during the last ten years. Computation methods used are generally complex. The MARQUARDT's algorithm used by BOX and JENKINS is an example of it. In this paper, we propose the greatest slope's method for that estimation. The computation algorithm is simple and in spite of some disadvantages offers certain advantages: fitting coefficients of the ARMA model to the changes in the structure of time-series, a prevision of short time-series. Through an economic time-series taken as example, we then show that the fitness of our previsions are comparable to BOX and JENKINS' method.

Die Estimation der Parameter des Prozessus ARMA hat sich im Lauf der letzten zehn Jahre sehr stark entwickelt. Die Rechenmethoden, die dazu verwendet werden sind im allgemeinen komplex. Der Algorithmus von MARQUARDT, angewendet von BOX und JENKINS ist dafür ein Beispiel. Wir schlagen in diesem Artikel vor die Verwendung der Technik des größten Gefälles um diese Estimation zu verwirklichen.

Der Algorithmus der Berechnung ist einfach und hat trotz einiger Nachteile bestimmte Vorteile : Anpassung der Koeffizienten des Modells ARMA an die langsamen Veränderungen der Struktur der Zeitserien, Prevision von kurzen Zeitserien. Mit Hilfe einer ökonomischen Zeitserie zeigen wir, daß die Qualität unserer Voraussagen kann mit der, die von BOX und JENKINS geliefert wird, verglichen werden.

Les modèles économétriques présentent généralement un attrait beaucoup plus grand pour l'économiste que les modèles issus de simples ajustements statistiques. En effet, l'élaboration des premiers fait appel à une théorie économique sous-jacente, c'est-à-dire une explication logique, en termes d'enchaînements causaux, de la formation d'une ou plusieurs variables. A contrario, les méthodes d'ajustements statistiques, parmi lesquelles la plus développée actuellement est celle de BOX et JENKINS [4], ne font appel à aucune hypothèse à priori sur la manière dont sont engendrées les observations. En fait, on suppose, que celles-ci appartiennent à une classe de processus linéaires

Journal de la Société de Statistique de Paris, tome 122, n° 3, 1981.

stationnaires. Cette démarche est certes un handicap pour l'économiste qui souhaite expliquer mais aussi prévoir. Cependant ces méthodes conduisent à des variances résiduelles plus faibles que celles fournies par les modèles économétriques. Toutefois les deux approches sont plus complémentaires qu'opposées car les modèles d'ajustement univariés permettent une connaissance plus précise de la forme des résidus et les modèles d'ajustements dynamiques, une spécification plus rigoureuse des équations dans les constructions économétriques.

Qu'il s'agisse des méthodes univariées ou dynamiques la méthodologie de BOX et JENKINS consiste à parcourir trois étapes : identification, estimation-test et prévision, en itérant les deux premières étapes jusqu'à l'obtention d'un modèle correct. Nous proposons dans ces pages une variante de la phase d'estimation du cas univarié qui permet l'obtention de prévisions à partir de séries temporelles courtes.

En effet, l'estimation, au sens des moindres carrés, des paramètres du modèle identifié est réalisée le plus souvent au moyen de l'algorithme de MARQUARDT [18]. Celui-ci est une variante de la méthode de GRADIENT, mais au lieu de se déplacer d'une longueur constante dans la direction du gradient on utilise une nouvelle direction et la longueur du déplacement varie en fonction de l'angle formé par les directions de deux itérations successives [29]. Lorsque la structure de la série temporelle ne se modifie pas fondamentalement au cours du temps (série stationnaire par exemple) les paramètres obtenus permettent de bonnes prévisions. Néanmoins cette méthode est assez sophistiquée et peut poser des problèmes de compréhension. De plus l'expression analytique des composantes du vecteur des dérivées partielles étant inconnue la méthode numérique à laquelle on recourt (définition de la dérivée en un point) nécessite un nombre assez important d'observations.

Nous utilisons la méthode de la plus grande pente (Steepest descent) pour estimer les paramètres du modèle. Elle a l'avantage d'une grande simplicité et sa compréhension ne nécessite pas un investissement mathématique approfondi. De plus l'approximation du vecteur des dérivées partielles personnalise l'outil et confère à l'estimation et à la prévision une nature spécifique. En effet, les coefficients du modèle sont modifiés au fur et à mesure que l'on parcourt la période étudiée et leurs valeurs finales correspondent aux meilleurs paramètres du modèle représentant le groupe d'observations de l'extrémité de la série. En raison de cette caractéristique la méthode est particulièrement adaptée aux prévisions à court terme à partir de petits échantillons d'observations.

Utilisée dans les travaux de MAKRIDAKIS et WHEELWRIGHT [13, 14, 15, 16, 17] sur le « filtre adaptatif », elle a conduit à des résultats très satisfaisants. Le « filtre adaptatif généralisé » [17, CH. 9] nécessite peu d'informations, à priori, sur la série temporelle. Ainsi, on peut, par exemple, estimer les paramètres d'un processus non stationnaire en moyenne. Toutefois, la forme théorique des modèles du « filtre adaptatif généralisé » relève de la classe des processus ARMA (p, q) linéaires (1). Dans le cas non stationnaire l'identification de ces processus n'est pas aisée, leurs ordres p et q sont élevés et difficiles à déterminer. Le nombre d'itérations nécessaire pour faire converger l'algorithme de la plus grande pente vers l'optimum de la fonction critère s'accroît alors, ainsi que le temps de calcul de chaque itération.

Si l'on choisit de filtrer la série temporelle pour la rendre stationnaire on identifie des modèles intégrant un nombre de paramètres moins élevé et permettant une utilisation rapide, donc plus performante, de la chaîne de programmes informatique que nous avons élaborée [27]. En utilisant, en outre, des séries courtes (ou une extrémité de série) on évite le risque de modifications importantes de la structure de la chronique ce qui permet d'utiliser certains tests de sélection.

Cet article analyse l'utilisation de l'algorithme de la plus grande pente comme méthode d'estimation des paramètres des modèles ARMA stationnaires. Nous l'appliquons par suite aux 57 observations de la consommation de ciment en Languedoc-Roussillon. Les résultats obtenus, comparés à ceux issus du « Filtre adaptatif généralisé » et de la méthode de BOX et JENKINS, illustrent l'efficacité de notre procédure pour la prévision à court terme d'une série temporelle.

---

1. Du terme anglo-saxon : « autoregressive-moving average process ».

I. L'ESTIMATION DES PARAMÈTRES DU MODÈLE ARMA (p, q) :

1) La classe des modèles ARMA retenue dans l'étude

Le théorème de WOLD [33, p. 89] est à l'origine des processus aléatoires linéaires et stationnaires de la classe ARMA (p, q). Si $\widetilde{X}_t$ est un processus aléatoire indéterminable, centré et stationnaire d'ordre deux, le modèle ARMA (p, q) vérifie la relation :

(1) $$\varnothing_p(B)\,\widetilde{X}_t = \theta_q(B)\,Z_t$$

où $$B^i\,\widetilde{X}_t = \widetilde{X}_{t-i};\ B^j Z_t = Z_{t-j}$$

$$\varnothing_p(B) = 1 - \sum_{i=1}^{p} \varnothing_i B^i$$

est le polynôme d'opérateurs de la partie autorégressive du modèle

$$\theta_q(B) = 1 - \sum_{j=1}^{q} \varnothing_j B^j$$

est le polynôme d'opérateurs de la partie moyenne mobile du modèle et Z_t un processus purement aléatoire de moyenne nulle. Les séries temporelles économiques sont rarement stationnaires. En général, elles présentent une tendance qui est le plus souvent monotone, des saisonnalités d'amplitudes variables et des fluctuations résiduelles de variabilité croissante dans le temps. On doit alors utiliser des transformations linéaires de la série temporelle qui conduisent à la définition des modèles ARIMA (1) suivants :

$$\varnothing_p(B)(1-B)^d\,\widetilde{X}_t = \theta_q(B)\,Z_t$$

et $$\varnothing_P(B^s)(1-B^s)^D\,\widetilde{X}_t = \theta_Q(B^s)\,Z_t$$

pour le cas saisonnier où s est la période (en unité de temps) de la saisonnalité

$$B^{is}\,\widetilde{X}_t = \widetilde{X}_{t-is};\ B^{js} Z_t = Z_{t-js}$$

$$\varnothing_P(B^s) = 1 - \sum_{i=1}^{p} \varnothing_i B^{is}$$

est le polynôme d'opérateurs de la partie autorégressive saisonnière du modèle

$$\theta_Q(B^s) = 1 - \sum_{j=1}^{Q} \theta_j B^{js}$$

est le polynôme d'opérateurs de la partie moyenne mobile saisonnière du modèle.

En combinant ces deux modèles ARIMA on réalise la forme la plus générale des modèles ARIMA (p, d, q) × (P, D, Q) multiplicatifs utilisée par BOX et JENKINS :

(2) $$\varnothing_p(B)\,\varnothing_P(B^s)(1-B)^d(1-B^s)^D\,\widetilde{X}_t = \theta_q(B)\,\theta_Q(B^s)\,Z_t$$

Toutefois, comme la procédure d'estimation employée ultérieurement n'est pas adaptée à la forme multiplicative de cette relation nous utilisons, dans notre application, le modèle additif :

(3) $$(\varnothing_p(B) + \varnothing_P(B^s) - 1)(1-B)^d(1-B^s)^D\,\widetilde{X}_t = (\theta_Q(B^s) + \theta_q(B) - 1)\,Z_t$$

C'est le processus ARMA (p, q) stationnaire :

(4) $$\varnothing(B)\,\widetilde{\widetilde{X}}_t = \theta(B)\,Z_t$$

---

1. Du terme anglo-saxon : « autoregressive integrated moving average process ».

où $\varnothing(B)$ et $\theta(B)$ sont des polynômes d'opérateurs de degrés au plus égaux à sP (noté p) et sQ (noté q) respectivement, et, $\tilde{\tilde{X}}_t$ le processus stationnaire $(1-B)^d(1-B^s)^D \tilde{X}_t$ que l'on notera $\tilde{X}_t$ dans la suite de l'exposé.

Il s'agit de trouver les valeurs $\varnothing_i$ (i = 1, ... p) et θj (j = 1, ... q) qui minimisent l'erreur entre les valeurs observées et les valeurs calculées par le modèle (4).

Plus exactement on utilise le critère des moindres carrés. En écrivant :

$$\tilde{X}_t = \varnothing_1 \tilde{X}_{t-1} + \dots + \varnothing_p \tilde{X}_{t-p} + Z_t - \theta_1 Z_{t-1} - \dots - \theta_q Z_{t-q}$$

alors

$$Z_t = \tilde{X}_t - \left(\sum_{i=1}^{p} \varnothing_i \tilde{X}_{t-i} - \sum_{j=1}^{q} \theta_j Z_{t-j}\right)$$

où Z_t est appelée l'erreur de prévision.

Le problème à résoudre est alors le suivant :

$$\text{(5)} \qquad \text{Min}\left(\sum_{t=k}^{n} Z_t^2\right) = \text{Min}\,(\varnothing(\varnothing_i, \theta_j)) = \text{Min}\ F(\nu_i)$$

où n est le nombre d'observations de la série temporelle

$k = \max(p, q) + 1$

$F(\nu_i)$ la fonction objectif (ou critère)

et ν_i est un vecteur de $p+q$ composantes.

2) Estimation des paramètres par l'algorithme de la plus grande pente

a) *La méthode de la plus grande pente*

La relation (5) traduit la recherche de l'optimum sans contrainte d'un modèle non linéaire de plusieurs variables indépendantes $\varnothing_i$ (i = 1, ..., p) et θj (j = 1, ..., q).

Deux classes de méthodes sont utilisées couramment pour résoudre ce type de problèmes [29] :

— les méthodes directes qui n'utilisent pas le calcul des dérivées partielles de la fonction objectif. La technique de ROSENBROCK, qui généralise et perfectionne l'approche de l'optimum par dichotomie dans le cas d'une variable, est l'une des plus performantes parmi ces méthodes.

— les méthodes qui ont recours aux notions classiques de géométrie analytique et qui utilisent les dérivées partielles de la fonction objectif comme celles de MARQUARDT et FLETCHER-POWELL par exemple. Elles ont toutes pour origine la méthode du GRADIENT.

L'algorithme de la plus grande pente est ainsi le plus simple de cette seconde classe de méthodes d'optimisation sans contraintes. Son application est aisée mais il ne converge que très lentement lorsqu'on approche du minimum de la fonction critère, ce qui explique l'emploi de versions plus sophistiquées qui accélèrent la convergence.

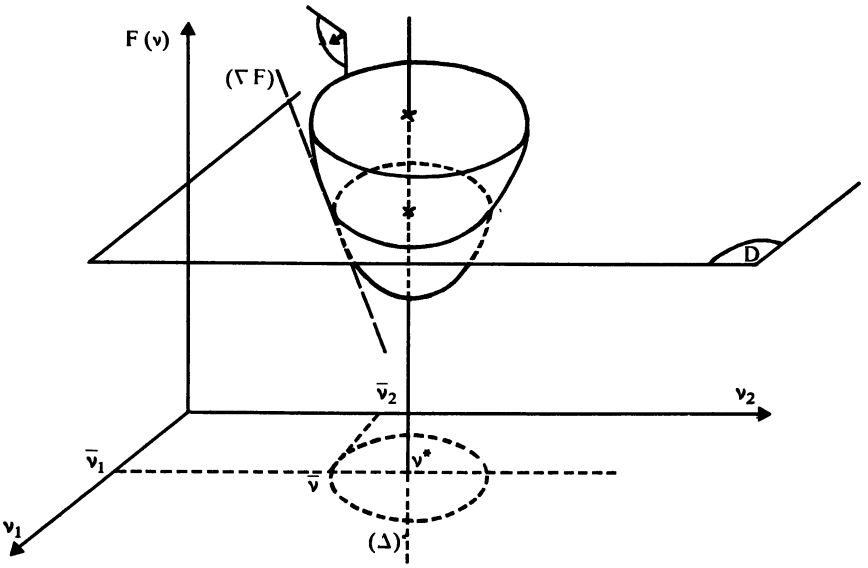
La fonction critère multidimensionnelle $F(\nu)$ est une hypersurface de réponse à $(p+q+1)$ dimensions, que nous supposerons unimodale. Les contours de la surface peuvent être quelconque; s'ils présentent des arêtes, la méthode ne convergera que très lentement vers le minimum.

Soit un point de l'hypersurface défini par le vecteur ν_1 et la valeur $F(\nu_1)$ de la fonction objectif. Soit une variation de $F(\nu)$, mesurée par $\Delta F(\nu) = F(\nu_1) - F(\nu_2)$, lorsqu'on se déplace en un point $(\nu_2, F(\nu_2))$ de l'hypersurface, de telle sorte que $\varepsilon = \nu_1 - \nu_2$ est petit.

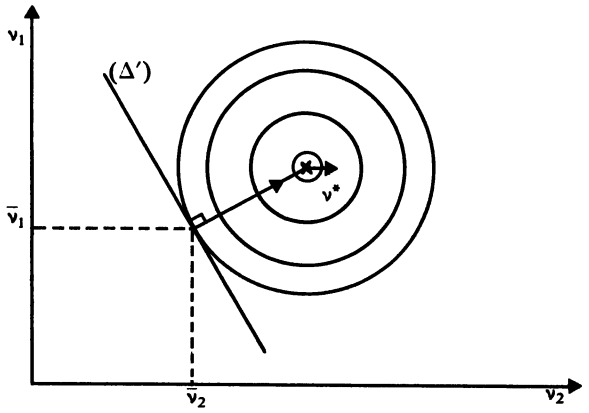
Cette variation est maximale, en valeur absolue, quand sa direction est celle du vecteur gradient de la fonction $F(\nu_1)$. Le gradient étant le vecteur des dérivées premières de la fonction critère par rapport aux variables indépendantes, notons le :

$$\nabla F(\nu) = \delta F(\nu)/\delta\nu$$

Ce théorème de la variation maximale de la fonction objectif fonde la méthode de la plus grande pente que nous allons expliciter géométriquement dans un cas simple. Supposons que le vecteur v ait deux composantes et que la représentation dans l'espace à trois dimensions de la fonction critère soit une paraboloïde.



Les projections sur le plan (v_1, v_2) des contours de la surface de réponse, intersections du volume en cuvette avec les plans (D) orthogonaux à l'axe (Δ) donnent des courbes de niveau qui peuvent être, dans un cas idéal, des cercles concentriques.



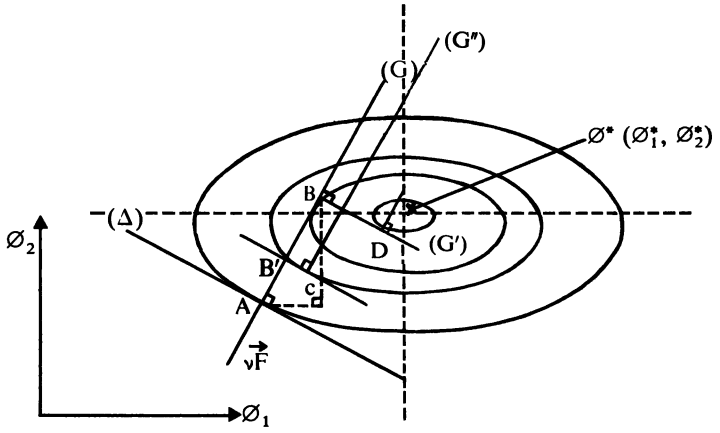
Si on choisit un point $\bar{v}$ ($\bar{v}_1$, $\bar{v}_2$) d'une courbe de niveau et que l'on se déplace dans la direction du gradient, qui est, par définition perpendiculaire à la tangente (Δ') passant par $\bar{v}$ on atteint immédiatement le minimum v^* recherché. Il en est de même si les lignes de niveau de la surface de réponse sont elliptiques et que le point de départ se situe sur l'un des deux axes. Dans le cas d'un point quelconque d'une ellipse ou d'un contour présentant une arête il faut un nombre infini de trajets en zigzags pour atteindre le minimum [32, p. 136].

La méthode de la plus grande pente consiste donc à choisir une valeur initiale $\bar{v}$ du vecteur des variables indépendantes puis à améliorer la première approximation F ($\bar{v}$) réalisée en choisis-

sant un point de l'hypersurface dont les coordonnées sont les composantes de $\bar{\bar{v}}$ de sorte que la diminution $\Delta F(v) = F(\bar{\bar{v}}) - F(\bar{v})$ soit maximale en valeur absolue. La direction dans laquelle s'effectue cette réduction est celle du vecteur gradient. On suppose donc que F(v) est une fonction différenciable convexe ((1)).

b) *Application à l'estimation des paramètres du modèle ARMA (p, q)*

Supposons à nouveau que le vecteur v ait deux composantes. On veut par exemple estimer les paramètres d'un processus autorégressif du second ordre noté AR (2). Supposons de plus que les courbes de niveaux de la surface de réponse sont des ellipses concentriques.



Choisissons un point $A(Ø^0_1, Ø^0_2)$ quelconque d'une courbe de niveau. Il appartient au volume de révolution de l'espace à trois dimensions $(F(Ø_1, Ø_2), Ø_1, Ø_2)$. Le vecteur gradient est représenté dans $(Ø_1, Ø_2)$ par sa projection (G). Elle est perpendiculaire, par définition, à la tangente (Δ) en A de l'ellipse. Ce vecteur des dérivées a pour composantes les vecteurs :

$$\overrightarrow{AC} = \frac{\delta F(Ø_1, Ø_2)}{\delta Ø_1} \quad \text{et} \quad \overrightarrow{CB} = \frac{\delta F(Ø_1, Ø_2)}{\delta Ø_2}$$

Le déplacement dans la direction du gradient, conduit au point $B(Ø^1_1, Ø^1_2)$ lorsque la variation de la fonction critère est maximale. Dans ces conditions on redéfinit la nouvelle direction (G') du gradient et le déplacement au point $D(Ø^2_1, Ø^1_2)$. Le trajet en zigzags amène, après plusieurs itérations, au point $Ø^*(Ø^*_1, Ø^*_2)$ qui représente dans cette projection les coordonnées du minimum de la fonction critère.

L'équation paramétrique de la droite (G) perpendiculaire à la droite (Δ) est :

$$\overrightarrow{AB} = \lambda' \overrightarrow{\nabla F}$$

Pour chacune des deux composantes du vecteur $\overrightarrow{\nabla F}$ on a alors :

$$Ø^1_1 = Ø^0_1 + \lambda' \overrightarrow{AC} = Ø^0_1 + \lambda' \delta F(Ø_1, Ø_2)/\delta Ø_1$$

et

$$Ø^1_2 = Ø^0_2 + \lambda' \overrightarrow{CB} = Ø^0_2 + \lambda' \delta F(Ø_1, Ø_2)/\delta Ø_2$$

---

1. Une fonction F(v) est convexe si pour tout $\bar{v}$ et pour tout $\bar{\bar{v}}$ de coordonnées réelles et si pour tout α_1, α_2 dans [0, 1] tels que $\alpha_1 + \alpha_2 = 1$ on a :

$$F(\alpha_1 \bar{\bar{v}} + \alpha_2 \bar{v}) \leqslant \alpha_1 F(\bar{\bar{v}}) + \alpha_2 F(\bar{v})$$

ou encore si et seulement si

$$\frac{\delta_2 F(v)}{\delta v^2} \geqslant 0$$

λ' est un nombre réel de signe positif si la direction de la plus grande pente est ascendante et négatif si elle est descendante. Son rôle est fondamental car le déplacement correct d'un point à un autre des courbes de niveau dépend de sa valeur. Un λ' trop important peut déboucher sur une oscillation ou une divergence explosive, un λ' trop faible conduira à un nombre considérable d'itérations. Ainsi, dans ce dernier cas, le déplacement du point A peut atteindre le point B' et le nouveau calcul s'effectuer dans la direction (G'').

Si on calcule la différentielle de la fonction critère $F(\varnothing_1, \varnothing_2)$:

$$\Delta F(\varnothing_1, \varnothing_2) = \frac{\delta F(\varnothing_1, \varnothing_2)}{\delta \varnothing_1} \Delta\varnothing_1 + \frac{\delta F(\varnothing_1, \varnothing_2)}{\delta \varnothing_2} \Delta\varnothing_2$$

en posant

$$\Delta\varnothing_1 = \lambda' \frac{\delta F(\varnothing_1, \varnothing_2)}{\delta \varnothing_1}$$

et

$$\Delta\varnothing_2 = \lambda' \frac{\delta F(\varnothing_1, \varnothing_2)}{\delta \varnothing_2}$$

alors :

$$\Delta F(\varnothing_1, \varnothing_2) = \lambda' \left\{ \left(\frac{\delta F(\varnothing_1, \varnothing_2)}{\delta \varnothing_1} \right)^2 + \left(\frac{\delta F(\varnothing_1, \varnothing_2)}{\delta \varnothing_2} \right)^2 \right\} \tag{6}$$

Comme la fonction critère est convexe la variation $\Delta F(\varnothing_1, \varnothing_2)$ est négative et λ' l'est alors aussi. En écrivant $\lambda' = -\lambda$ où λ est un nombre réel positif le calcul des paramètres du processus AR(2) qui minimisent la fonction critère $F(\varnothing_1, \varnothing_2)$ est à l'itération k :

$$\begin{aligned} \varnothing_1^k &= \varnothing_1^{k-1} - \lambda \frac{\delta F(\varnothing_1, \varnothing_2)}{\delta \varnothing_1} \\ \varnothing_2^k &= \varnothing_2^{k-1} - \lambda \frac{\delta F(\varnothing_1, \varnothing_2)}{\delta \varnothing_2} \end{aligned} \tag{7}$$

De façon très générale pour une fonction objectif $F(v)$ différentiable convexe, où v est le vecteur à $p+q$ composantes, on aura avec un nombre fini d'itérations (dépendant de la précision recherchée) :

$$v_i^{k+1} = v_i^k - \lambda \nabla F(v_i^k) \tag{8}$$

Dans la mesure où les composantes du gradient $\nabla F(v_i)$ sont connues, le vecteur v_i, pour une itération donnée, dépend uniquement du paramètre λ. Il en résulte que $F(v_i) = F_i(\lambda)$ et la variation optimum de la fonction $F_i(\lambda)$, à chaque itération, est obtenue en résolvant l'équation $dF_i(\lambda)/d\lambda = 0$ au moyen de la méthode de RAPHSON-NEWTON par exemple. Dans ce cas on aboutit au vecteur v_i^* recherché après un nombre K d'itérations, défini par la précision désirée de la variation relative de la fonction critère. Le vecteur v_i^* peut être considéré comme le « vecteur moyen » qui convient le mieux pour réduire la fonction critère en tenant compte de toutes les observations de la série temporelle. Cette façon de procéder nécessite donc un calcul numérique des composantes du vecteur gradient et une détermination du paramètre.

Au contraire, la méthode du filtre adaptatif généralisée, utilise une approximation à un instant donné du vecteur gradient. Il en résulte que le paramètre λ n'est pas calculé par une procédure automatique; c'est une donnée fournie par l'utilisateur.

Rappelons que ν_i est un vecteur à $(p+q)$ dimensions qui regroupent les p composantes du vecteur $\varnothing_i$ et les q composantes du vecteur θj.

Le vecteur gradient $\nabla F(\nu_i)$ peut donc s'écrire $\nabla F(\varnothing_i)$ pour les p premières valeurs et $\nabla F(\theta_j)$ pour les q suivantes. De plus en tenant compte de la relation (5) on peut représenter ces vecteurs par

$$\nabla F(\varnothing_i) = \delta\left(\sum_{i=1}^{n} Z_t^2\right)/\delta\varnothing_i$$

et

$$\nabla F(\theta_j) = \delta\left(\sum_{i=1}^{n} Z_t^2\right)/\delta\theta_j$$

Soit, pour t fixé, l'approximation [12, p. 40] :

$$\nabla F(\varnothing_i) = \delta Z_t^2/\delta\varnothing_i = 2\,Z_t\,\delta Z_t/\delta\varnothing_i$$

et

$$\nabla F(\theta_j) = \delta Z_t^2/\delta\theta_j = 2\,Z_t\,\delta Z_t/\delta\theta_j$$

et comme

$$Z_t = \widetilde{X}_t - \left(\sum_{i=1}^{p} \varnothing_i \widetilde{X}_{t-i} - \sum_{j=1}^{q} \theta_j Z_{t-j}\right)$$

$$\delta Z_t/\delta\varnothing_i = -\,\widetilde{X}_{t-i}$$

$$\delta Z_t/\delta\varnothing_j = +\,Z_{t-j}$$

En utilisant la relation (8) on transforme les paramètres du processus ARMA (p, q) de telle sorte qu'on minimise, au sens des moindres carrés, les résidus obtenus à partir du modèle (5) :

$$\varnothing_{i,t}^{(k'+1)} = \varnothing_{i,t}^{(k')} + 2\lambda Z_t \widetilde{X}_{t-i} \tag{9}$$

$$\theta_{j,t}^{(k'+1)} = \theta_{j,t}^{(k')} - 2\lambda Z_t Z_{t-j}$$

où

$$i = 1, \ldots, p$$
$$j = 1, \ldots, q$$
$$t = T, \ldots, n \text{ et } T = \max(p+1, q+1)$$

Lorsqu'on parcourt les n observations il y a $k' = (n - T + 1)$ transformations des paramètres. Lorsque l'ensemble des paramètres a été modifié par la relation (9) et que l'on a parcouru toutes les observations de la série temporelle (ce qui correspond à une itération) on répète le processus jusqu'à ce que la réduction relative de l'erreur au carré devienne inférieure à un niveau fixé à priori.

Après κ itérations le modèle construit avec les vecteurs $\varnothing_{i,\,n}$ et $\theta_{j,\,n}$ est celui qui convient le mieux pour décrire les dernières observations de la série. Cette façon de procéder autorise ainsi l'emploi de séries temporelles contenant peu d'observations et recommande l'utilisation de toute procédure (stationnarisation, filtrage...) stabilisant l'évolution dans le temps de la chronique. De ce fait le modèle conduit à des écarts résiduels acceptables et certains tests de sélection associés à la méthode de BOX et JENKINS peuvent être utilisés valablement.

c) *Convergence de l'algorithme*

Dans l'optique précédente d'une révision des coefficients du modèle, le paramètre λ ne peut être calculé au cours des k' transformations de chacune des k itérations. On attribue à ce paramètre une valeur comprise dans un intervalle qui garantit (condition suffisante) la convergence de l'algorithme de la plus grande pente. Cette valeur (cf. annexe) est telle que :

$$o < \lambda < \frac{1}{\sum_{i=1}^{p} \tilde{X}_{t-i}^2 + \sum_{j=1}^{q} Z_{t-j}^2}$$

Compte tenu de la démarche utilisée pour aboutir à ce résultat nous effectuons nos calculs en standardisant les observations de la façon suivante :

$$\tilde{X}_i^* = \frac{\tilde{X}_i}{\left(\sum_{j=t-1}^{t-p} \tilde{X}_j^2\right)^{1/2}} \text{ et } Z_i^* = \frac{Z_i}{\left(\sum_{k=t-1}^{t-q} X_k^2\right)^{1/2}}$$

où $i = t-1, ..., t-p$ et t_o l'origine des temps est égale au Max $(p, q) + 1$.

Les valeurs $\tilde{X}_i^*$ et Z_i^* sont bornées supérieurement par un et les valeurs

$$\tilde{X}_i^{*2}, Z_i^{*2}, \sum_{i=t-1}^{t-p} \tilde{X}_i^{*2}, \sum_{k-t=1}^{t-q} Z_k^{*2}$$

par p et q.

Ainsi la standardisation des observations, en plus d'accélérer la convergence de l'algorithme, forme une plage de variation de λ répondant à la simple inéquation :

$$o < \lambda < \frac{1}{p+q}$$

II. Prévision de la consommation de ciment en Languedoc-Roussillon

La consommation de ciment des cinq départements de l'Aude, du Gard, de l'Hérault, des Pyrénées-Orientales et de la Lozère est un indicateur conjoncturel de l'activité du Bâtiment pour la région Languedoc-Roussillon.

La série trimestrielle que nous utilisons débute en (Mars-Avril-Mai) 1964 et s'achève au même trimestre de 1978 (1). L'identification, l'estimation et enfin la prévision au deuxième trimestre 1979 sont réalisées au moyen de la chaîne de programme Auter [27].

a) *L'identification*

Les représentations graphiques de la série origine de 57 observations et de la série filtrée par une moyenne mobile de longueur quatre met en exergue l'existence d'une composante linéaire croissante (graphique 1) (2).

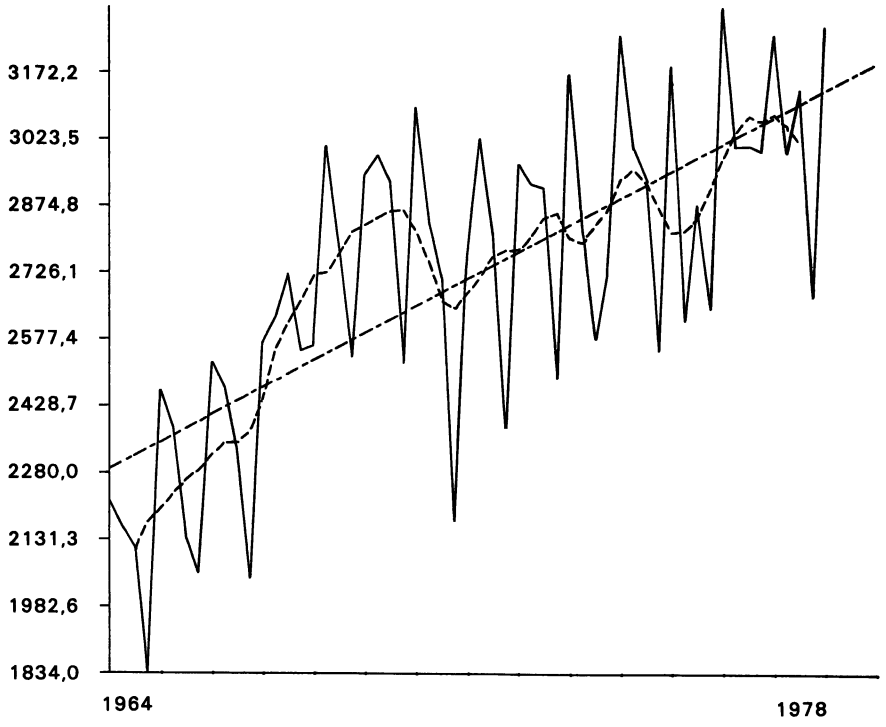
---

1. Ces données ont été relevées dans la Revue « Économie Méridionale » publiée par le Centre Régional de Productivité de la Faculté de Droit et des Sciences Économiques de Montpellier. L'unité est le millier de tonnes.

---

2. Les caractéristiques statistiques de cette tendance, estimée par la méthode des moindres carrés ordinaires, sont les suivantes :

$$T_t = 15{,}095\ t + 2\ 274{,}988 \qquad r = 0{,}705$$
$$(0{,}2759)\ (69{,}6923)$$



GRAPHIQUE 1

Les pics de période quatre observations, dénotent la présence d'un mouvement saisonnier. Le tableau ci-dessous montre, entre 1965 et 1977, que les pointes enregistrées graphiquement se situent, en moyenne, au second trimestre (Mars-Avril-Mai) de chaque année.

Trimestres	1	2	3	4
Moyennes série Origine	2421,9	2930,4	2824,6	2760,2
Moyennes série des différences pre mières	−295,7	508,5	−105,8	−64,4

L'existence de la composante saisonnière est confirmée par l'analyse de variance réalisée sur la série des différences premières : la statistique de FISHER prend la valeur 27,3 pour $\nu_1 = 3$ et $\nu_2 = 3$ degrés de liberté. L'importance des pics, sur les douze années prises en compte par le tableau de BUYS-BALLOT, est relativement constante. Ce fait est confirmé par le coefficient de régression (−0,07) non significativement différent de zéro, de la pente de la droite ajustée à partir des écarts-types et des moyennes de chaque année. Le modèle représentant la série temporelle est ainsi de type additif et son estimation peut être, en conséquence, valablement réalisée par l'algorithme de la plus grande pente. Cette investigation rapide des principales composantes de la chronique permet, avant l'identification du processus, de choisir le filtre adéquat pour stationnariser et/ou réduire le nombre de paramètres du modèle ARMA. Nous analysons, pour cet exemple, la série non transformée (notée (a)) et sa filtrée (notée (b)) par la relation :

$$Y_t = (1 - B^4)X_t$$

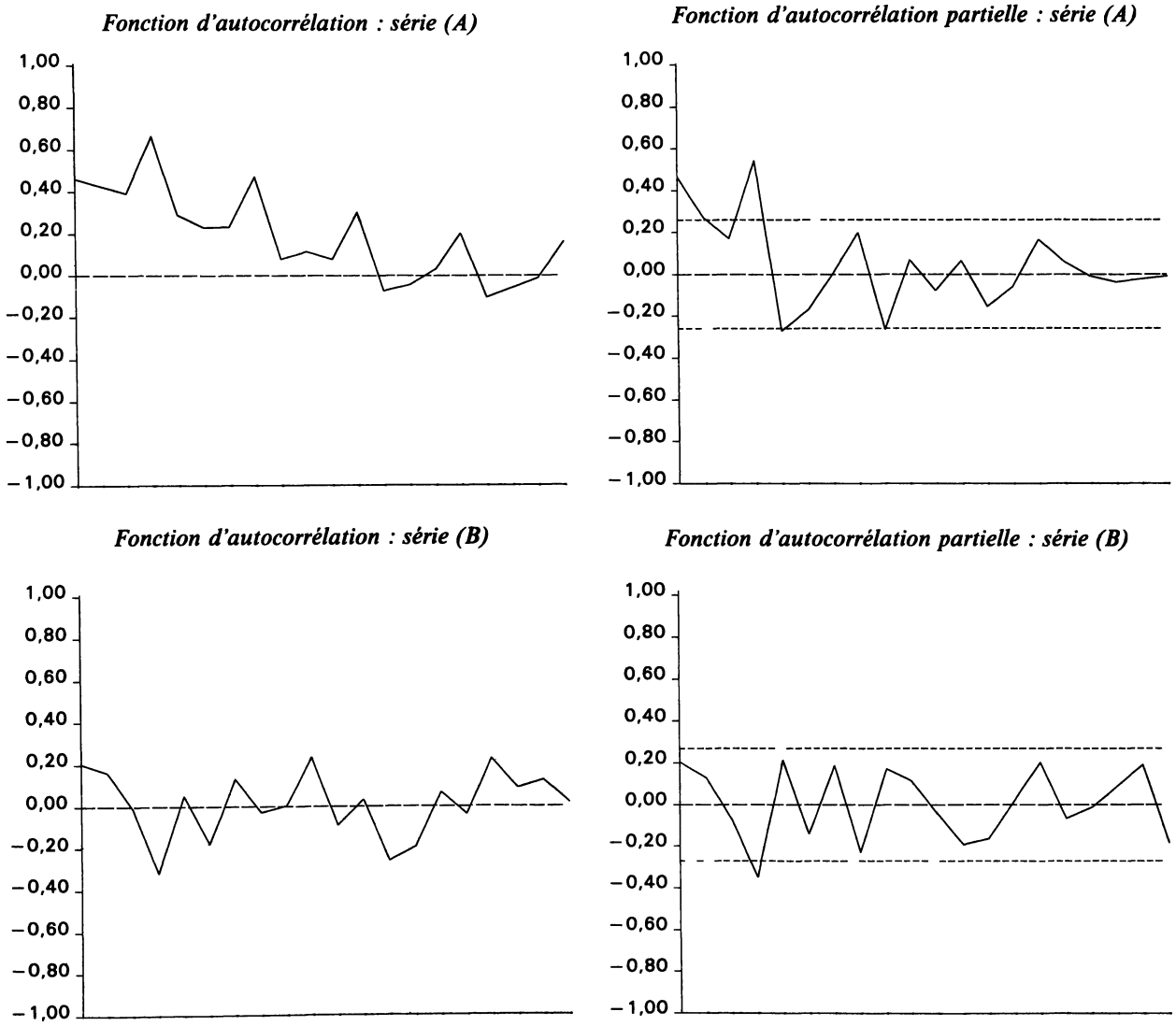
Ce filtre a la propriété de réduire la composante saisonnière et de tendance.

L'identification du processus générateur s'effectue au moyen des fonctions d'autocorrélations et d'autocorrélations partielles [4, ch. 3], [23]. Le graphique 2 représente l'évolution de chacune d'entre elles en fonction de vingt retards.

Nous avons alors sélectionné :

pour la série origine (a) : le modèle (1) : AR(4)
pour la série filtrée (b) : le modèle (2) : MA(4)
le modèle (3) : ARMA(1, 4)
le modèle (4) : ARMA(0, 1) + (1, 0) saisonnier.

GRAPHIQUE 2



b) *L'estimation*

L'estimation préliminaire des paramètres des modèles (hormis le (4)) permet une convergence plus rapide de l'algorithme de la plus grande pente (le critère d'arrêt étant une variation relative des carrés des résidus inférieure à 10^{-6}).

Le tableau synoptique, ci-dessous, résume pour chacun des modèles le nombre total de paramètres estimés (NT), la valeur de la constante d'apprentissage λ, le nombre d'itérations (ITER) nécessaire à la convergence de l'algorithme, la moyenne ($\bar{Z}$) et la variance (V) des résidus, la valeur du χ^2 avec le nombre de degrés de liberté pour l'adéquation globale du modèle, la statistique AIC [23] qui permet de sélectionner le modèle le plus performant.

	NT	λ	ITER	$\bar{Z}$	V	$\chi^2(\nu)$	AIC
(1)	4	0,245	3	63,2	34 665	18,9 (10)	607,8
(2)	4	0,01	13	−7,9	32 368	11,5 (9)	603,9
(3)	5	0,10	28	−6,9	31 711	14,8 (8)	606,5
(4)	2	0,02	43	−5,8	37 219	15,4 (11)	613,9

Remarquons la convergence plus lente du modèle (4), en introduisant comme valeurs liminaires des paramètres 0,5, par comparaison avec les autres modèles pour lesquels les valeurs initiales des coefficients ont été calculées [4]. Les résidus sont peu ou pas du tout autocorrélés pour l'ensemble des modèles. Le test du χ^2 au seuil 0,05 (sauf pour (1) où $\alpha = 0{,}02$) confirme l'adéquation générale des constructions aux observations de la série. Néanmoins comme la valeur de la statistique AIC est minimale pour le modèle (2) on peut présumer de la validité de ce dernier.

c) *La prévision*

Mais c'est surtout au niveau de la prévision que l'on juge la performance du modèle. En raison de la procédure d'estimation utilisée on peut retenir comme critère de performance le modèle qui engendre les plus faibles résidus calculés en fin de série ou sur prévision dans la mesure où l'information est disponible.

Le tableau 1 décrit les écarts (Z_t) et les écarts relatifs (dZ_t, en pourcentage) entre la série calculée et la chronique initiale pour les huit derniers trimestres. Les caractéristiques de ces résidus confirment que le modèle (2) fournit la meilleure description de l'évolution dans le temps de la consommation de ciment. le tableau 2 qui retrace les mêmes informations pour les trois premiers points de la prévision classe le modèle (4) avant le (2) pour la performance en prévision. Les graphiques 3 illustrent en trait plein la série origine et en trait pointillé la série calculée par chacun des deux modèles recolorés. La ligne verticale limite les observations utilisées pour estimer les paramètres des différents processus. La visualisation des différents graphiques confirme les conclusions précédentes.

L'étude des caractéristiques de ces résidus montre de plus que la stationnarisation préalable de la chronique améliore sensiblement les performances prévisionnelles des modèles (1). Notre procédure paraît ainsi plus appropriée pour effectuer les prévisions à court terme des séries temporelles que le « filtre adaptatif » de MAKRIDAKIS et WHEELWRIGHT (résultats du modèle (1)).

La confrontation de nos calculs avec ceux issus de l'estimation des paramètres au moyen de l'algorithme de MARQUARDT (méthode de BOX et JENKINS) est plus délicate en raison d'une part de la saisonnalité additive de la chronique et d'autre part du nombre limité des observations. Ainsi, seul le modèle (3) a pu être estimé (2); toute autre sélection conduit à une non convergence de

1. L'exemple unique proposé dans l'article pourrait être un cas particulier. Nous avons pu néanmoins vérifier cette conclusion lors de l'élaboration de prévisions à court terme de chroniques économiques par notre procédure : cf : « le point sur la conjoncture » publié trimestriellement (en collaboration avec M. RODRIGUEZ) : supplément à l'Économie Méridionale. CRPEE — Faculté de Droit de Montpellier.

2. Le modèle s'écrit :

$$Y_t = \underset{(0{,}141)}{0{,}205}\, Y_{t-1} + Z_t - \underset{(0{,}084)}{0{,}094}\, Z_{t-} - \underset{(0{,}069)}{0{,}208}\, Z_{t\,2} - \underset{(0{,}0798)}{0{,}272}\, Z_{t-3} + \underset{(0{,}0637)}{0{,}661}\, Z_{t-4} + \underset{(28{,}276)}{55{,}486}$$

$X^2_C = 7{,}24$ $\nu = 9$; AIC = 571,996.

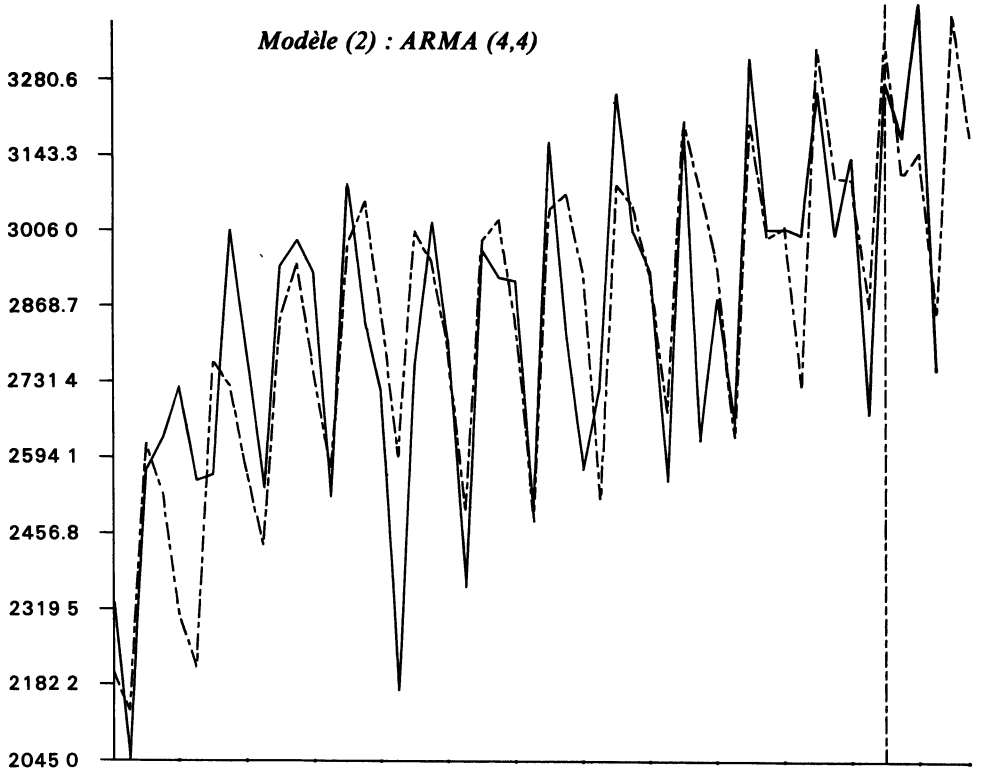
l'algorithme de MARQUARDT. En comparant les écarts résiduels engendrés par ce modèle (tableaux 1 et 2) avec ceux de son homologue estimé par l'algorithme de la plus grande pente on constate à nouveau une meilleure performance de ce dernier.

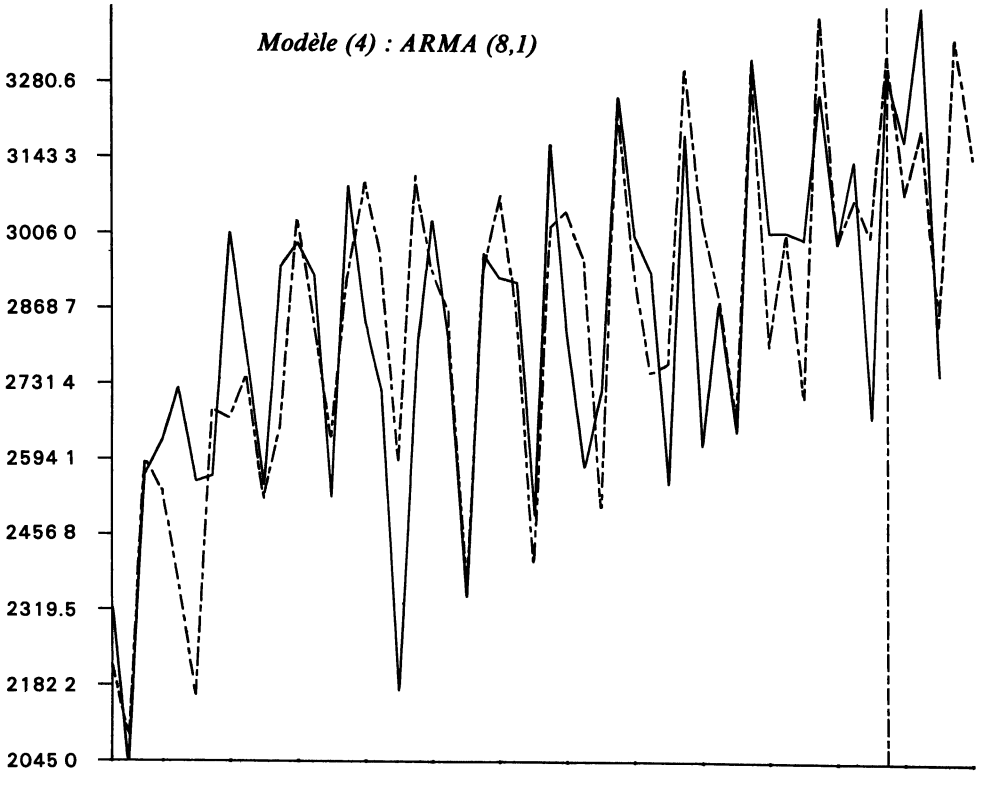
TABLEAU 1

t	Modèle 1 (filtre adaptatif)		Modèle 2		Modèle 3		Modèle 4		Modèle 3 (méthode BOX et JENKINS)	
	Z_t	dZ_t (%)	Z_t	dZ_t (%)	Z_t	dZ_t (%)	Z_t	dZ_t (%)	Z_t	dZ_t (%)
50	347,8	11,6	13,8	0,5	85,6	2,8	204,6	6,8	−136,1	−4,5
51	86,1	2,9	− 0,9	−0,03	62,7	2,1	9,4	0,3	−213,1	−7,1
52	305,9	10,2	270,1	9,0	237,6	7,9	283,2	9,5	203,6	6,8
53	35,8	1,1	− 79,9	−2,5	−183,4	−5,0	−145,4	− 4,5	195,6	6,0
54	0,8	0,03	−110,3	−3,7	− 93,1	−3,1	3,9	0,1	−151,2	− 5,1
55	130,7	4,2	39,6	1,3	94,5	3,0	70,0	2,2	−148,1	− 4,7
56	−295,1	−11,0	−196,4	−7,3	−261,7	−9,8	−332,8	−12,5	−277,8	−10,4
57	106,3	3,2	79,1	−2,4	93,7	−2,9	− 54,1	− 1,7	256,0	7,8
		$\overline{dZ}_t = 2{,}76$ $V = 47{,}97$ $\Sigma\lvert dZ_t\rvert = 44{,}2$		$\overline{dZ}_t = 0{,}65$ $V = 22{,}69$ $\Sigma\lvert dZ_t\rvert = 26{,}4$		$\overline{dZ}_t = 0{,}69$ $V = 32{,}25$ $\Sigma\lvert dZ_t\rvert = 37{,}2$		$\overline{dZ}_t = 0{,}045$ $V = 45{,}66$ $\Sigma\lvert dZ_t\rvert = 37{,}6$		$\overline{dZ}_t = -1{,}4$ $V = 50{,}6$ $\Sigma\lvert dZ_t\rvert = 52{,}4$

TABLEAU 2

t	Modèle 1 (filtre adaptatif)		Modèle 2		Modèle 3		Modèle 4		Modèle 3 (méthode BOX et JENKINS)	
	Z_t	dZ_t (%)	Z_t	dZ_t (%)	Z_t	dZ_t (%)	Z_t	dZ_t (%)	Z_t	dZ_t (%)
58	231	7,3	69	2,3	162	5,1	95	3,0	95,9	3,0
59	3041	8,9	273	8,0	329	9,6	223	6,5	131,6	3,9
60	44	1,6	−105	−3,8	− 86	−3,1	− 85	−3,1	−258,2	−9,4
		$\overline{dZ}_t = 5{,}9$ $\Sigma\lvert dZ_t\rvert = 17{,}8$		$\overline{dZ}_t = 2{,}2$ $\Sigma\lvert dZ_t\rvert = 14{,}1$		$\overline{dZ}_t = 3{,}9$ $\Sigma\lvert dZ_t\rvert = 12{,}6$		$\overline{dZ}_t = 2{,}1$ $\Sigma\lvert dZ_t\rvert = 12{,}6$		$\overline{dZ}_t = -0{,}83$ $\Sigma\lvert dZ_t\rvert = 16{,}3$


Modèle (2) : ARMA (4,4)
3280.6
3143.3
3006 0
2868.7
2731 4
2594 1
2456.8
2319 5
2182 2
2045 0


Modèle (4) : ARMA (8,1)
3280.6
3143 3
3006 0
2868 7
2731 4
2594 1
2456 8
2319.5
2182 2
2045 0

En retenant, en définitive, le modèle (2) qui engendre les écarts résiduels les moins importants, la consommation de ciment en Languedoc-Roussillon peut être représentée valablement, tout au moins pour les dernières observations, par le modèle recoloré ARMA (4, 4) :

$$X_t = X_{t-4} + Z_t - 0{,}0812\, Z_{t-1} - 0{,}144\, Z_{t-2} - 0{,}135\, Z_{t-3} + 0{,}644\, Z_{t-4} + 70{,}77$$

Les valeurs prévisionnelles avec leur intervalle de confiance (au seuil $\alpha = 0{,}05$) sont consignées dans le tableau ci-après :

Observations N°	Borne inférieure	Prévision	Borne supérieure
58	2 752,7	3 105,3	3 457,9
59	2 788,9	3 145,3	3 501,6
60	2 500,0	2 859,5	3 218,9
61	3 017,9	3 397,7	3 778,5
62	2 794,2	3 176,1	3 557,9

⁂

L'estimation des paramètres d'un modèle ARMA par l'algorithme de la plus grande pente a un aspect pédagogique certain. La présentation méthodologique de l'algorithme ne nécessite pas de développements mathématiques complexes. En raison de la simplicité des résultats auxquels on parvient il est facilement programmable et peut s'intégrer aisément dans un logiciel de la méthode de BOX et JENKINS.

Cet algorithme est aussi une méthode performante pour déterminer les paramètres du processus générateur stationnaire d'une chronique contenant peu d'observations. La seconde partie de cet article a montré qu'une telle approche débouchait sur des écarts résiduels prévisionnels relativement faibles. La méthode améliore sensiblement les résultats du « filtre adaptatif » de MAKRIDAKIS et WHEELWRIGHT et se révèle même compétitive avec celle de BOX et JENKINS. La révision des paramètres lorsqu'on parcourt la chronique confère son originalité à l'instrument qui nécessite de ce fait peu d'informations sur le cheminement temporel de la chronique. Il est ainsi possible dans le cas de séries longues de ne retenir, pour l'estimation, que les observations de l'extrémité de la chronique. Cette possibilité permet alors de limiter les risques de modifications structurelles de la chronique mais aussi de réduire notablement le temps de calcul de chaque itération.

Lorsque les composantes de la série ne connaissent pas de modifications fondamentales pour la période d'étude certains tests de sélection associés à la méthode de BOX et JENKINS restent valables pour les séries courtes. Il s'agit, par exemple, des tests de la fonction d'autocorrélation, du χ^2, du minimum de la statistique AIC. La stationnarisation et l'élimination de composantes par filtrage facilitent l'identification du processus générateur et conduisent parfois à des modèles à faible nombre de paramètres. Lorsqu'on combine un tel modèle avec l'utilisation d'une série courte, l'algorithme converge encore plus rapidement vers le minimum de la fonction critère.

Ce dernier aspect de la méthode est important dans la mesure où le choix arbitraire de la constante d'apprentissage λ nécessite plusieurs essais. Sur ce point, notre propre expérience ne débouche pas sur un critère valable de résolution et l'efficacité de la méthode est limitée par le choix de ce paramètre. Ceci d'autant plus que les résultats de l'estimation sont très sensibles à la valeur de λ. On a pu tout au plus constater que λ tend vers $1/p$ lorsque le processus générateur est autorégressif et non stationnaire et vers une valeur très faible lorsqu'il est autorégressif ou de moyenne mobile et stationnaire.

Mais la détermination non automatique de λ n'est pas le seul inconvénient du recours à l'algorithme de la plus grande pente. Ce dernier nécessite, de par sa construction, une saisonnalité additive de la chronique. Bien que l'emploi des logarithmes des observations puisse être, dans certains cas, un moyen de tourner la difficulté, cette contrainte limite sérieusement le domaine d'application de la méthode.

Néanmoins si les précautions d'utilisation sont respectées on peut réaliser, avec un temps de calcul modeste, des prévisions que bon nombre de méthodes ne sont pas à même d'engendrer à partir de petits échantillons.

⁂

Annexe

Appelons ΔZ_t^k la diminution de l'erreur entre deux itérations successives ($t = 1, \ldots n$). On peut écrire

$$\Delta Z_t^k = Z_t^{k-1} - Z_t^k = \mu Z_t^{k-1}$$

où μ est le facteur de réduction

$$Z_t^{k-1} = \widetilde{X}_t - \sum_{i-1}^{p} \varnothing_i^{k-1} \widetilde{X}_{t-i} + \sum_{j=1}^{q} \theta_j^{k-1} Z_{t-j}^{k-1}$$

$$Z_t^k = \widetilde{X}_t - \sum_{i-1}^{p} \varnothing_i^k \widetilde{X}_{t-i} + \sum_{j=1}^{q} \theta_j^k Z_{t-j}^k$$

$$\Delta Z_t^k = \sum_{i=1}^{p} (\varnothing_i^k - \varnothing_i^{k-1}) \widetilde{X}_{t-i} + \sum_{j=1}^{q} \theta_j^{k-1} Z_{t-j}^{k-1} - \sum_{j=1}^{q} \theta^k Z_{t-j}^k$$

Or d'après la relation (9) et en supposant que les valeurs de Z_{t-j}^k et Z_{t-j}^{k-1} sont proches d'une itération à l'autre :

$$\Delta Z_t^k = 2\lambda Z_t^{k-1} \sum_{i=1}^{p} \widetilde{X}_{t-i}^2 + 2\lambda Z_t^{k-1} \sum_{j=1}^{q} (Z_{t-j}^{k-1})^2$$

$$\Delta Z_t^k = 2\lambda Z_t^{k-1} \left(\sum_{i=1}^{p} \widetilde{X}_{t-i}^2 + \sum_{j=1}^{q} (Z_{t-j}^{k-1})^2 \right) \tag{10}$$

$$\lambda = \frac{\mu Z_t^{k-1}}{2 Z_t^{k-1} \left(\sum_{i=1}^{p} \widetilde{X}_{t-i}^2 + \sum_{j=1}^{q} (Z_{t-j}^{k-1})^2 \right)} = \frac{\mu}{2 \left(\sum_{i=1}^{p} \widetilde{X}_{t-i}^2 + \sum_{j=1}^{q} (Z_{t-j}^{k-1})^2 \right)}$$

Cette relation généralise l'équation (6) dans le cadre particulier de l'approximation, à un instant donné, du vecteur gradient. Elle montre de par la présence du coefficient μ l'inefficacité d'un calcul automatique du paramètre λ. Pour cette raison on précise le domaine de variation de λ. L'écriture matricielle du problème à résoudre (équation (5)) est la suivante :

$$\nabla F = -2A + 2\underset{\sim}{\varnothing}' B \tag{11}$$

Dans laquelle :

∇F est le vecteur colonne du gradient

A le vecteur de composantes $\widetilde{X}_t \widetilde{X}_{t-i}$ ($i = 1, \ldots, p$) et $(-\widetilde{X}_t Z_{t-j})$($j = 1, \ldots, q$)

$\underset{\sim}{\varnothing}'$ le vecteur ligne des paramètres $\varnothing_i$ et θj

B la matrice carrée de dimension $p+q$:

$$\begin{array}{c} \\ 1 \\ \vdots \\ k \\ \vdots \\ p \\ 1 \\ \vdots \\ n \\ \vdots \\ q \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & \dots & 1 & \dots & p & 1 & \dots & m \dots q \end{array} \\ \left[\begin{array}{ccc|c} & & & \\ & \widetilde{X}_{t-k}\,\widetilde{X}_{t-1} & & \widetilde{X}_{t-k}\,Z_{t-m} \\ & & & \\ \hline & & & \\ & \widetilde{X}_{t-n}\,Z_{t-1} & & Z_{t-n}\,Z_{t-m} \\ & & & \end{array}\right] \end{array} = \left[\begin{array}{c|c} C & D \\ \hline F & E \end{array}\right]$$

Pour obtenir le minimum de la fonction critère Z_t^2 il faut que :

$$\nabla F = O$$

$$\text{ou que : } \quad \vec{\varphi}\, B = A$$

La résolution de ce système nécessite une matrice B symétrique et définie positive. Cette condition est toujours vérifiée si l'on utilise l'espérance mathématique de la fonction critère pour effectuer ces calculs. Si les observations sont de plus centrées et réduites B est alors une matrice des corrélations.

La relation (11) nous fournit l'expression du vecteur gradient à l'itération (k + 1).

$$\frac{\nabla F^{(k+1)}}{2} = \vec{\varphi}^{(k+1)}\, B - A$$

La modification des composantes du vecteur $\vec{\varphi}$ à chaque itération peut s'écrire :

$$\vec{\varphi}^{(k+1)} = \vec{\varphi}^{(k)} - \lambda \nabla F^{(k)}$$

d'où

$$\frac{\nabla F^{(k+1)}}{2} = \left[\vec{\varphi}^{(k)} - \lambda \nabla F^{(k)}\right] B - A$$

$$\frac{\nabla F^{(k+1)}}{2} = \frac{\nabla F^{(k)}}{2} - \lambda \nabla F^{(k)}\, B$$

$$\frac{\nabla F^{(k+1)}}{2} = [I - 2\lambda B]\, \frac{\nabla F^{(k)}}{2}$$

Soit encore :

$$\frac{\nabla F^{(k)}}{2} = [I - \lambda' B]\, \frac{\nabla F^{(0)}}{2}$$

Le processus itératif est convergent lorsque $\nabla F^{(k)}$ tend vers zéro. On peut démontrer [8, p. 115] que cette condition est réalisée lorsque :

$$0 < \lambda' < \frac{2}{\rho_{max}}$$

où ρ_{max} est la plus grande valeur propre de la matrice B. Les matrices E et F étant nulles une borne supérieure de ρ_{max} est la quantité :

$$\sum_{i=1}^{p} E\left[\widetilde{X}_{t-i}^2\right] + \sum_{j=1}^{q} E\left[Z_{t-j}^2\right]$$

En tenant compte de la relation (10) (calculée à partir de l'espérance mathématique de ΔZ_t^k)on réalise la condition suffisante de convergence de l'algorithme de la plus grande pente :

$$0 < \mu \leqslant 2$$

La plage de variation de λ est alors :

$$0 < \lambda < \frac{1}{\sum_{i=1}^{p} \tilde{X}_{t-i}^2 + \sum_{j=1}^{q} Z_t^{2-j}}$$

BIBLIOGRAPHIE

[1] ANDERSON T.-W. — The statistical analysis of time series John Wiley 1971.

[2] BEREZIN et ZHIDKOV. — Computing Methods — the solution of sets of linear algebric equations — Pergamon press — Copyright 1965.

[3] BOSSIER F. et MELARD G. — Une application économique des modèles ARIMA non stationnaires — Publications économétriques. Vol. XI fasc. 1.

[4] BOX G.E.P. et JENKINS G.-M. — Time series Analysis : Forecasting and control Holden — Day. 1976.

[5] CORMARY Y., GUILBOT A. — Processus d'optimisation en quatre étapes appliqué à la recherche des paramètres des modèles déterministes. Laboratoire d'Hydrologie Mathématique — Université des Sciences et Techniques du Languedoc — Montpellier. Novembre 1971.

[6] DAVID M. — Statistique des modèles de régression à erreurs liées : cas des estimateurs de BOX et JENKINS — Thèse Université de Rennes — 1975.

[7] DENNIS — Mathematical programming and Electrical Networks Copyright 1959.

[8] DURAND E. — Solutions numériques des équations algébriques — Tome II — Masson 1961.

[9] HIMMELBLAU D.-M. — Applied Non linear Programming. Mc Graw Hill New-York 1972.

[10] HUYBERECHTS S. — Orientations récentes en analyse chronologique — Cahiers du Centre d'études de Recherche Opérationnelle, 19, n° 3/4 — 1977.

[11] INDHEHAGOPIAN J.-P. — Les modèles ARIMA et la démarche BOX et JENKINS. Application à un exemple CERESSEC — Juillet 1978.

[12] LAMY M. et VERHELLE R. — Formulation mathématique de l'enquête conjoncturelle relative aux prévisions de l'effectif ouvrier dans le secteur sidérurgique Belge. Approche BOX et JENKINS. C.R.E.D.E.L. Université de Liège.

[13] MAKRIDAKIS S. et WHEELWRIGHT S. — Forecasting with adaptative filtering — R.A.I.R.O. (7e Année, V-1, 1973).

[14] MAKRIDAKIS S. et WHEELWRIGHT S. — Choix et valeur des méthodes de prévision. Éditions d'organisation — Paris. 1974.

[15] MAKRIDAKIS S. et WHEELWRIGHT S. — Adaptative filtering : An Integrated Autoregressive/Moving Average Filter for Time Series Forecasting. Opl. Res. Q. Pergamon Press 1977. Vol. 28, 2.

[16] MAKRIDAKIS S. et WHEELWRIGHT S. — Interactive Forecasting Holden — Day — San Francisco. 1978.

[17] MAKRIDAKIS S. et WHEELWRIGHT S. — Forecasting : Methods and Applications John Wiley 1978.

[18] MARQUARDT D.-W. — An algorithm for least squares estimation of non linear parameters. Journ. Soc. Indust. Appl. math., 11, n° 2, Juin 1963, pp. 431-441.

[19] MELARD G. — Processus purement indéterminables à paramètre discret — Approches fréquentielle et temporelle. Thèse Université Libre de Bruxelles — 1975.

[20] MELARD G. et DROESBEKE F. — Comparaison de prévisions obtenues par modèles de BOX et JENKINS avec celles obtenues par un modèle économétrique trimestriel de la Belgique. Colloque de l'ARAE, Lyon — 1976.

[21] MELARD G. — Prévision de Ventes — Colloque international d'économétrie Appliquée — Strasbourg 1977.

[22] MONFORT A. — Approche de BOX et JENKINS et approche économétrique des séries temporelles. Annales de l'INSEE n° 32 — 1978.

[23] OZAKI T. — On the Order Determination of ARIMA Models. Applied Statistics, 26, n° 3, pp. 290-301 — 1977.

[24] POUZET — Analyse numérique — chapitre 4 — Laboratoire de calcul de la faculté des Sciences de Lille — 1964.

[25] ROY R. — Sur le comportement des autocovariances et des autocorrélations d'une série non stationnaire — Université de MONTRÉAL — Faculté des Arts et des Sciences — Juin 1978.

[26] TERRAZA Michel — Les filtres des différences d'ordre p et des différences de moyennes mobiles simples dans l'analyse spectrale. Journal de la société de statistique de Paris — 1979 — n° 2.

[27] TERRAZA Michel — Utilisation de la chaîne AUTER pour la prévision à court terme d'une série temporelle — Laboratoire d'Économétrie de la Faculté de Droit et des Sciences Économiques de Montpellier — Novembre 1978 — Document interne — non publié.

[28] VIGNAU J.-P. — Économie Mathématique. Cours de DES — Faculté de Droit et des Sciences Économiques de Montpellier — 1972.

[29] VIGNES J. — Méthodes numériques d'optimalisation d'une fonction de plusieurs variables. Chimie Industrie — Génie chimique — Vol. 97 — n° 8 — Avril 1967.

[30] WALSH G. — Methods of optimization — John Wiley — New-York 1975.

[31] WILDE et BEIGHTLER — Foundations of optimization — Prentice Hall 1967.

[32] WILDE D.-J. — Méthodes de recherche d'un optimum — Dunod 1966.

[33] WOLD H. — A study in the analysis of stationary time series — Almqvist — Wiksell — Stockholm — 1954.