

# REVUE DE STATISTIQUE APPLIQUÉE

F. BORGARD

## **Sur l'interaction entre les outliers et les données homogènes lors de la modélisation d'un processus ARMA ou ARIMA**

*Revue de statistique appliquée*, tome 38, n° 2 (1990), p. 5-20

[http://www.numdam.org/item?id=RSA\\_1990\\_\\_38\\_2\\_5\\_0](http://www.numdam.org/item?id=RSA_1990__38_2_5_0)

© Société française de statistique, 1990, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) 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/>

# SUR L'INTERACTION ENTRE LES OUTLIERS ET LES DONNÉES HOMOGENÈS LORS DE LA MODÉLISATION D'UN PROCESSUS ARMA OU ARIMA

F. BORGARD

*Université Pierre et Marie Curie, ISUP*

## RÉSUMÉ

On étudie ici des séries temporelles  $X_t$  qui ne diffèrent d'un processus ARMA ou ARIMA  $Z_t$  que par quelques valeurs outliers et l'on montre, avec l'appui d'exemples, que pour détecter les outliers, l'analyse des valeurs résiduelles doit se faire avec la modélisation retenue.

*Mots-Clés : Processus ARMA - outlier - Modélisation*

## ABSTRACT

This article is concerned with  $X_t$  time serie that differ from an  $Z_t$  ARMA or ARIMA process only by a few outliers, and it shows, with examples, that for outliers detection the analysis of the residual values should be effected using the model identification.

*Key words : Autoregressive moving average model - outlier - Model identification.*

## I. Introduction

La modélisation par un modèle ARMA, ARIMA ou SARIMA d'une série temporelle présentant des outliers par rapport à la modélisation de cette même série sans outlier conduit à retenir :

- soit un modèle totalement différent proposant des liaisons temporelles qui ne peuvent s'expliquer par la nature de la série
- soit le même type de modèle mais avec des estimateurs différents des paramètres.

Nous nous proposons ici de montrer que pour détecter les outliers l'analyse des valeurs résiduelles doit se faire en liaison avec la modélisation retenue. En effet, un outlier relié par une liaison temporelle d'ordre  $k$  à une donnée homogène de la série met en évidence la non cohérence de cette donnée avec celle de l'outlier.

## II. Processus ARMA, ARIMA, SARIMA. Modèles Outliers au niveau de la modélisation ARMA

### II.1. Rappels sur les processus ARMA, ARIMA et SARIMA

D'après la théorie classique des processus ARMA, ARIMA et SARIMA (Box et Jenkins (1976), Gouriéroux et Monfort (1983))  $Z_t$  est dit suivre un processus autoregressif moyenne mobile ARMA  $(p, q)$  si  $(Z_t)_{t \in \mathbb{Z}}$  est un processus faiblement stationnaire qui peut s'écrire :

$$\begin{aligned} \Phi(B)Z_t &= \theta(B)\epsilon_t \\ \text{où} \quad \theta(B) &= 1 - \varphi_1 B \dots - \varphi_p B^p \\ \text{et} \quad \theta(B) &= 1 - \theta_1 B \dots - \theta_q B^q \end{aligned}$$

sont deux polynômes en  $B$ ,  $B$  étant l'opérateur retard défini par  $B(Z_t) = (Z_{t-1})$ .  $\epsilon_t$  un processus bruit blanc c'est-à-dire une suite de variables aléatoires non corrélées d'espérance mathématique nulle et de variance  $\sigma^2$

Les polynômes  $\Phi(B)$  et  $\theta(B)$  ont leurs racines en dehors du cercle unité et sont supposés ne pas avoir de facteur en commun.

$(Z_t)$  est un processus ARMA  $(p, d, q)$  si le processus  $(1 - B)^d$  est un processus ARMA  $(p, q)$  et plus généralement  $Z_t$  est un processus SARIMA  $(p, d, q; P, D, Q)$  si le processus  $(1 - B)^d (1 - B^S)^D$  est un processus ARMA tel que :

$$\Phi(B) \Phi_1(B^S)(1 - B)^d(1 - B^S)^D Z_t = \theta(B)\theta_1(B^S) \epsilon_t$$

avec  $\Phi(B)$  polynôme en  $B$  de degré  $p$   
avec  $\Phi_1(B^S)$  polynôme en  $B^S$  de degré  $P$   
avec  $\theta(B)$  polynôme en  $B$  de degré  $q$   
avec  $\theta_1(B^S)$  polynôme en  $B^S$  de degré  $Q$

Les polynômes  $\Phi(B)$  et  $\theta(B)$  n'ont pas de facteur en commun ainsi que les polynômes  $\Phi_1(B^S)$  et  $\theta_1(B^S)$ . Les quatre polynômes ont leurs racines à l'extérieur du cercle unité.

### II.2. Modèles d'outliers au niveau de la modélisation ARMA

Si on se réfère à Tsay (1986) on peut considérer, au niveau des processus ARMA, deux modèles d'outliers : l'un entraînant une modification ponctuelle, l'autre une modification durable des valeurs résiduelles par la modélisation de la série sans outlier :

Soit  $Z_t$  la série sans outlier qui suit un modèle ARMA :

$\Phi(B)Z_t = \theta(B)\epsilon_t$  et  $X_t$  la série avec outliers.

On peut considérer dans le cas d'un seul outlier que  $X_t$  et  $Z_t$  sont liés l'un à l'autre soit par la relation :

$$X_t = Z_t + \theta(B)\Phi^{-1}(B)\omega\xi_t(T) \quad (1)$$

soit par la relation :

$$X_t = Z_t + \omega\xi_t(T) \quad (2)$$

où  $\omega$  est une constante réelle qui représente l'amplitude de la perturbation et où  $\xi_t(T)$  est une fonction indicatrice indiquant que l'outlier se produit à l'instant  $T$ .

(1) représente le modèle I.O (Innovation outlier) qui introduit la perturbation directement au niveau de l'innovation.

(2) représente le modèle A.O (Additive outlier) qui introduit la perturbation directement à la valeur de la série stationnaire.

Pour plusieurs outliers, on a plus généralement :

$$X_t = \sum_{j \in J} \omega_j V_j(B) \xi_t(T_j) + Z_t$$

avec :

$$\begin{aligned} V_j(B) &= 1 \text{ pour un outlier A.O.} \\ V_j(B) &= \theta(B)\Phi^{-1}(B) \text{ pour un outlier I.O.} \end{aligned}$$

Dans le cas d'un outlier, si on pose  $e_t = \Phi(B)\theta^{-1}(B)X_t$

On a :

$$\begin{aligned} \text{soit } e_t &= \omega\xi_t(T) + \epsilon_t \text{ modèle I.O.} \\ \text{soit } e_t &= \omega\Phi(B)\theta^{-1}(B)\xi_t(T) + \epsilon_t \text{ modèle A.O.} \end{aligned}$$

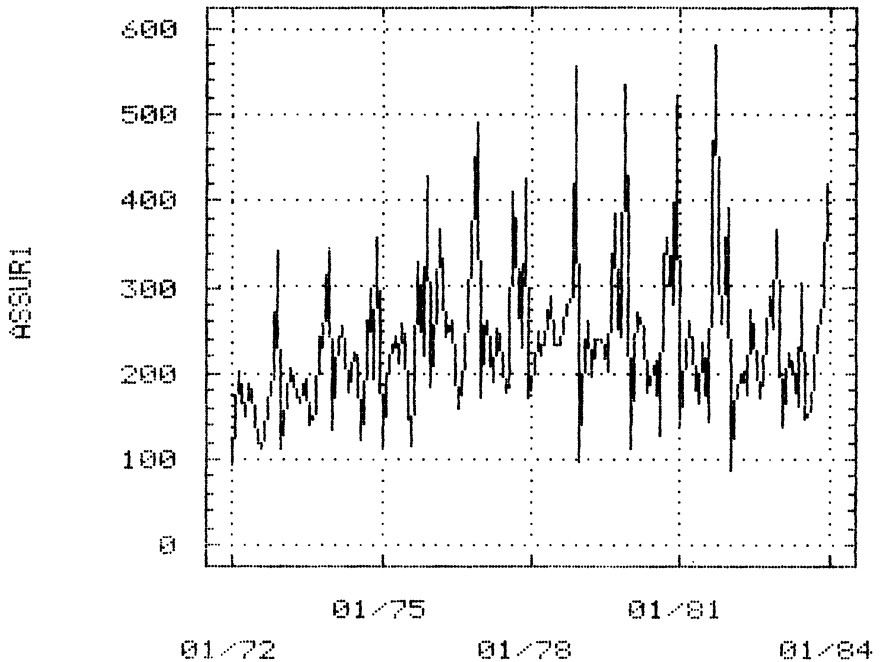
Dans le premier cas, l'amplitude de l'outlier n'est perçue qu'à l'instant  $T$ . Dans le deuxième cas, il est encore perceptible  $h$  instants après  $T$  avec la pondération  $\pi_h$  (avec  $\pi_h$  coefficient de  $B^h$  dans la division de  $\Phi(B)$  par  $\theta(B)$ ).

### III. Etude de la détection des outliers à partir des valeurs résiduelles

#### *III.1. Exemple de modélisations différentes pour la série avec outliers et pour la série sans outlier – et étude de l'effet sur les valeurs résiduelles d'une modélisation retenant une forte liaison temporelle d'ordre 12*

Nous allons étudier la série "produit Assurance Vie" sur des données mensuelles de janvier 1972 à décembre 1983 (Source : Journées d'étude en Statistique ASU 1984)

Le graphe de cette série (figure 1) montre une très légère tendance à la hausse et une saisonnalité d'ordre 12 marquée. Le tableau des valeurs mensuelles (figure 2) montre que les valeurs les plus importantes se situent en décembre, les valeurs les plus faibles en janvier. Ce qui est en accord avec le système de prime perçue par l'assureur.



**FIGURE 1**  
*Produit Assurance Vie. Données mensuelles du 01/72 au 12/83*

Un examen attentif des données montre que pour les mois de septembre quatre valeurs ne sont pas en cohérence avec le reste de la série. Ce sont :

- pour septembre 1977 la valeur 411
- pour septembre 1979 la valeur 385
- pour septembre 1980 la valeur 325
- pour septembre 1981 la valeur 581

Cette dernière valeur est la plus importante de la série.

- La valeur 411 est encadrée par les valeurs 191 et 291
- La valeur 385 est encadrée par les valeurs 203 et 258
- La valeur 325 est encadrée par les valeurs 129 et 356
- La valeur 581 est encadrée par les valeurs 143 et 322

La validation des données de ce fichier a montré que, de fait, ces quatre valeurs sont erronées. Au produit "Assurance vie" a été ajouté la vente d'un autre produit et les valeurs correctes sont :

| ANNEE | JAN. | FEV. | MARS | AVRIL | MAI | JUIN | JUIL. | AOUT | SEPT. | OCT | NOV. | DEC. |
|-------|------|------|------|-------|-----|------|-------|------|-------|-----|------|------|
| 1972  | 94   | 154  | 201  | 151   | 186 | 178  | 122   | 113  | 128   | 166 | 202  | 342  |
| 1973  | 111  | 171  | 206  | 186   | 165 | 185  | 193   | 141  | 155   | 244 | 231  | 346  |
| 1974  | 133  | 207  | 255  | 226   | 178 | 223  | 218   | 121  | 157   | 299 | 193  | 356  |
| 1975  | 114  | 180  | 218  | 243   | 212 | 257  | 234   | 114  | 187   | 329 | 221  | 429  |
| 1976  | 185  | 279  | 367  | 298   | 244 | 261  | 208   | 159  | 220   | 278 | 343  | 492  |
| 1977  | 172  | 251  | 261  | 190   | 251 | 243  | 179   | 191  | 411   | 291 | 231  | 427  |
| 1978  | 170  | 204  | 250  | 221   | 262 | 290  | 232   | 233  | 254   | 279 | 288  | 557  |
| 1979  | 96   | 182  | 260  | 195   | 240 | 240  | 227   | 203  | 385   | 258 | 243  | 534  |
| 1980  | 113  | 225  | 270  | 252   | 179 | 193  | 216   | 129  | 325   | 356 | 279  | 523  |
| 1981  | 139  | 224  | 262  | 225   | 166 | 236  | 198   | 143  | 581   | 322 | 257  | 391  |
| 1982  | 87   | 160  | 199  | 203   | 174 | 274  | 213   | 171  | 200   | 289 | 253  | 367  |
| 1983  | 137  | 190  | 225  | 186   | 163 | 304  | 146   | 158  | 203   | 245 | 286  | 421  |

FIGURE 2  
Tableau des valeurs mensuelles de la série ASSUR1  
"Produit Assurance Vie"

pour septembre 1977 229 et non 411  
pour septembre 1979 226 et non 385  
pour septembre 1980 222 et non 325  
pour septembre 1981 209 et non 581

Les deux séries ASSUR1 (avec les quatre outliers) et ASSUR2 (avec les données corrigées) deviennent stationnaires après une différenciation en  $1 - B$  et une différenciation en  $1 - B^{12}$ .

Mais leurs fonctions d'autocorrélation sont totalement différentes et ne peuvent conduire à une même modélisation. ASSUR1 présente une fonction d'autocorrélation (figure 3) avec deux pics significatifs pour  $k = 1$  et  $k = 12$ ; celui en 12 étant très important. Tandis que la fonction d'autocorrélation de ASSUR2 (figure 4) n'a aucun pic significatif en 12.

Une première modélisation de ASSUR1 donne le modèle :

$$(1 - \varphi_{12}B^{12})(1 - B)(1 - B^{12}) \text{ ASSUR1} = \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2}$$

avec  $\varphi_{12} = -0,48$   $\theta_1 = 0,81$   $\theta_2 = 0,13$

avec un test d'adéquation non satisfaisant et des valeurs résiduelles ayant une fonction d'autocorrélation et une fonction d'autocorrélation partielle présentant un pic nettement significatif en 9. Remodélisant ASSUR1 en écrivant que les  $\epsilon_t$  suivent un modèle autorégressif d'ordre 9 ( $(1 - \varphi_9B^9)\epsilon_t = a_t$  bruit blanc avec  $\varphi_9 = 0,28$ ) on arrive à une bonne adéquation : les valeurs résiduelles tant du point de vue de leur fonction d'autocorrélation que de leur fonction d'autocorrélation partielle, ou de leur périodogramme (figures 5, 6, et 7) peuvent être considérées comme les

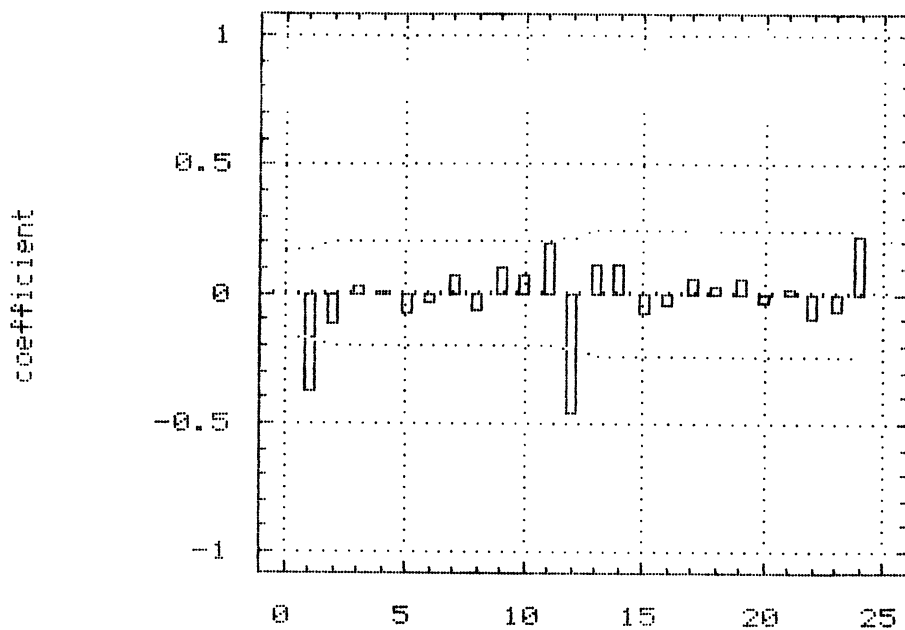


FIGURE 3

*Fonction d'autocorrélation de  $(1-B)(1-B*12)$  ASSUR1*

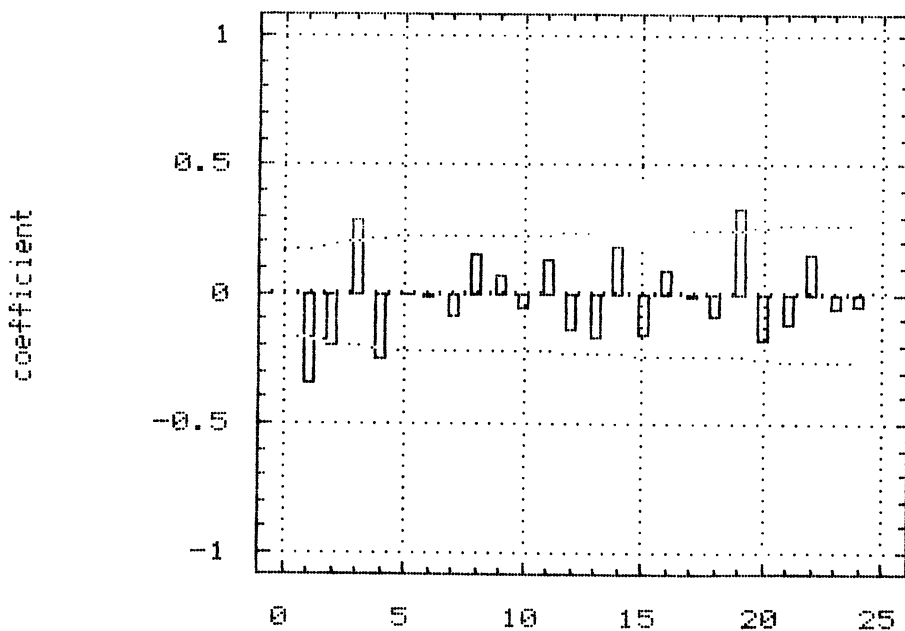
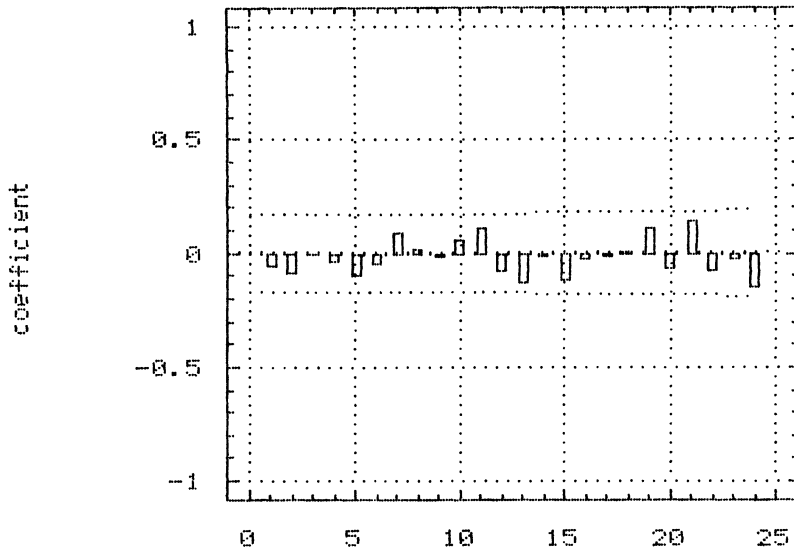
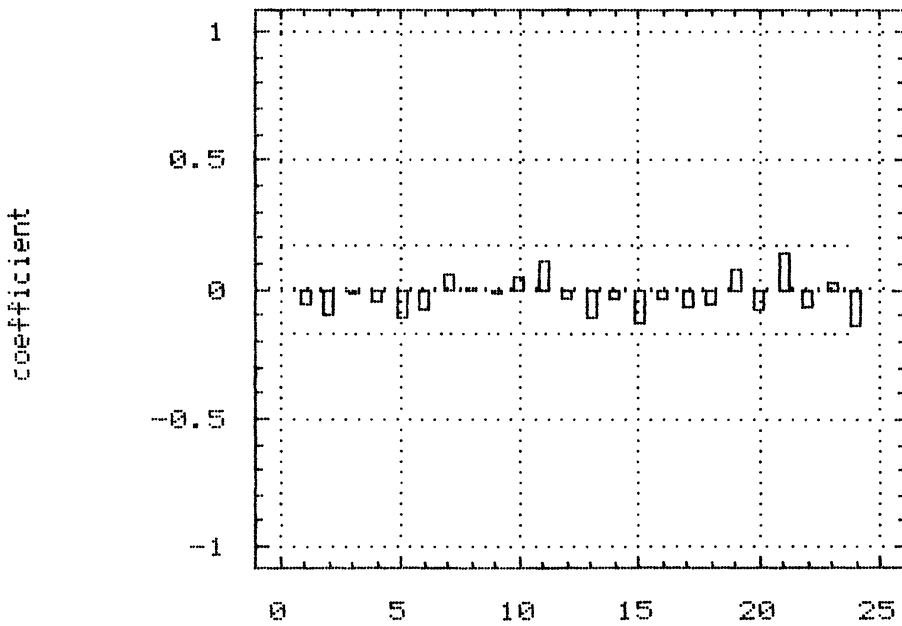


FIGURE 4

*Fonction d'autocorrélation de  $(1-B)(1-B*12)$  ASSUR2*



**FIGURE 5**  
*Fonction d'autocorrélation des résidus*  
*(deuxième modélisation de ASSURI)*



**FIGURE 6**  
*Fonction d'autocorrélation partielle des résidus*  
*(deuxième modélisation de ASSURI)*



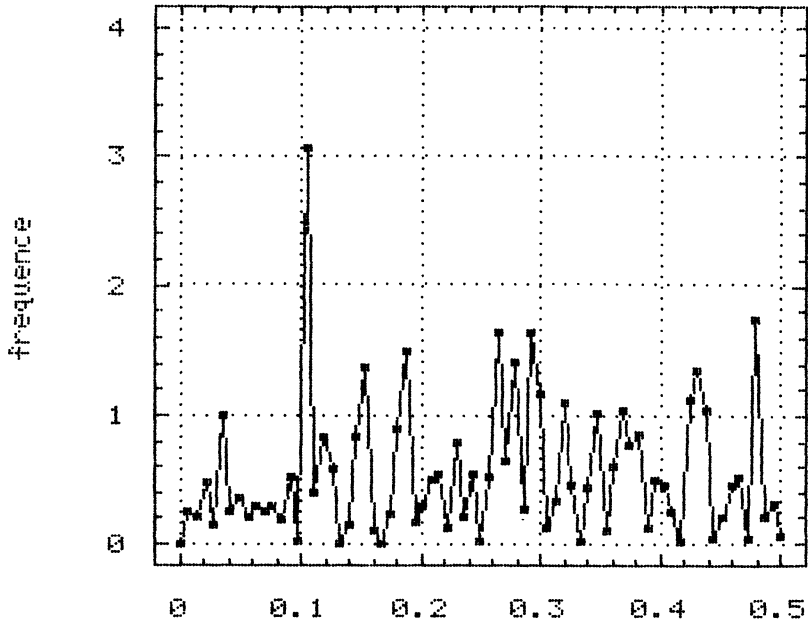


FIGURE 7  
Périodogramme des résidus (X10000)  
(deuxième modélisation de ASSUR1)

réalisations d'un bruit blanc. Mais la relation temporelle d'ordre 9 introduite dans la modélisation n'a pas d'interprétation économique. Si on recherche les valeurs résiduelles dont l'amplitude dépasse  $3,5\sigma^*$  ( $\sigma$  étant l'écart-type des résidus) on a pour les deux modèles, deux valeurs :

1er modèle :

|                |         |  |
|----------------|---------|--|
| Septembre 1981 | 230,5   | l'écart-type des résidus<br>étant 58,4 |
| Septembre 1982 | - 263,2 |  |

2ème modèle :

|                |         |  |
|----------------|---------|--|
| Septembre 1981 | 237,2   | l'écart-type des résidus<br>étant 56,0 |
| Septembre 1982 | - 222,6 |  |

On constate que les deux modèles qui l'un et l'autre possèdent une forte liaison d'ordre 12 (correspondant au terme  $(1 - \varphi_{12}B^{12})(1 - B^{12})$ ) ont bien intégré les valeurs outliers de septembre 1977, septembre 1979, Septembre 1980 et ont mal intégré l'outlier de septembre 1981 ainsi que la valeur de septembre 1982.

\* L'expérience montre qu'avec des données homogènes on peut avoir des résidus dont l'amplitude est comprise entre 2 et  $3\sigma$ .

Ils donnent pour ces deux mois le même ordre de grandeur de la valeur résiduelle (positif pour septembre 1981, négatif pour septembre 1982).

La valeur de septembre 1982 (200) qui est cohérente avec celle d'août 1982 (171) et celle d'octobre 1982 (289) apparaît d'après l'étude des valeurs résiduelles comme une valeur trop faible qu'il faut corriger en hausse de 220 à 260 points pour la rendre cohérente avec le reste de la série !

Ceci est dû au fait que la valeur outlier de septembre 1981 crée au niveau de la modélisation ARMA de  $(1 - B)(1 - B^{12})X_t$  deux valeurs outliers dont l'amplitude perturbatrice est de même grandeur, mais de signes opposés pour septembre 1981 et septembre 1982. En effet  $(1 - B)(1 - B^{12})X_t$  fait intervenir pour septembre 1981 la différence des valeurs de septembre 1981 avec septembre 1980 et pour septembre 1982 la différence des valeurs de septembre 1982 avec septembre 1981.

On peut, de plus, remarquer que parmi les outliers, seule, la valeur la plus importante (581) a été détectée par l'étude des valeurs résiduelles.

### ***III.2. Effet sur une liaison temporelle d'ordre 1 de deux valeurs outliers consécutives de perturbation opposée - Reconnaissance des outliers***

Etude d'un taux de variation de consommations mensuelles (Source : GAZ DE FRANCE).

Cette série comporte 49 valeurs (figure 8). Elle représente le cas classique dit "d'horloge" avec une valeur sous estimée (ici la valeur en 29) suivie immédiatement d'une valeur apparemment surestimée (ici la valeur 30). Si on corrige les valeurs de la série, on a le même type de modèle pour la série initiale et la série corrigée :

$$(1 - \varphi_{12}B^{12})X_t = \theta_0 + \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - \theta_3\epsilon_{t-3}$$

Pour ces deux séries la modélisation est très satisfaisante. Mais pour la série avec les outliers,  $\theta_1$  n'est pas significativement différent de zéro.

Lorsque l'on étudie les résidus de la série avec les outliers, ils présentent un écart-type de 5,5 et le seul résidu important est en 29 et vaut -18,24 (la valeur absolue est donc située entre 3 et  $3,5\sigma$ ); la valeur correspondante de la série peut donc être considérée comme faible mais presque cohérente avec l'ensemble de la série. En 30, la valeur résiduelle n'est que de 9,71. La liaison d'ordre 1 dans la modélisation n'existant pas, ce résidu est calculé sans faire intervenir la valeur en 29 et fait apparaître cette valeur en 30 comme une valeur cohérente de la série ce qui est contraire à la comparaison graphique des valeurs en 29 et 30.

### ***III.3. Etude du rôle des liaisons temporelles d'ordre 1 et d'ordre 12 pour la détection de quatre outliers intégrés à la série du modèle "Air-line" de Box et Jenkins***

Cette série que nous avons notée AIRI1 (Source Box et Jenkins) comporte 144 valeurs mensuelles (figures 9 et 10) et est le modèle classique du modèle

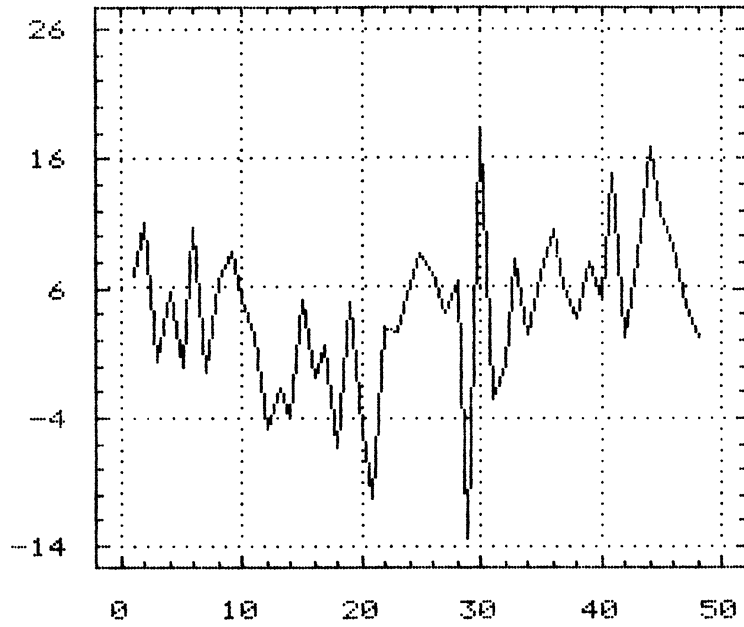


FIGURE 8  
( $\times 100$ ) *Taux de variation de consommation mensuelle*

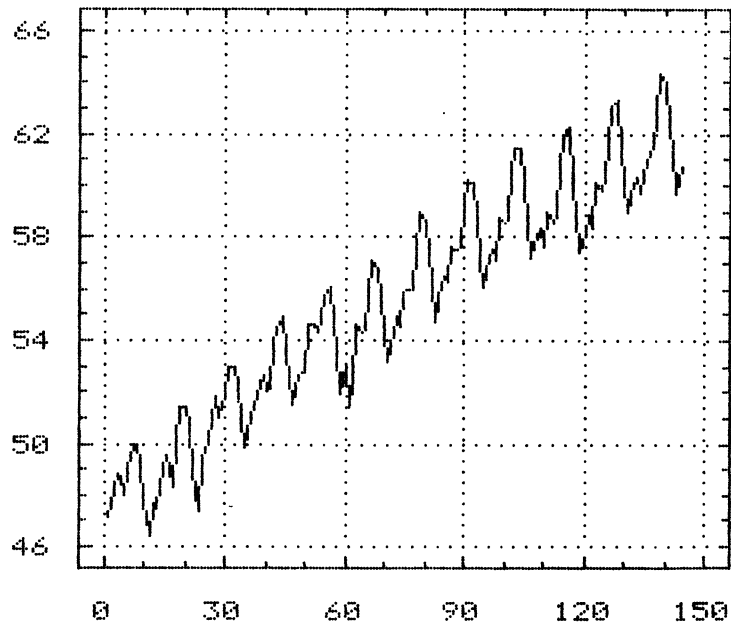


FIGURE 9  
*Série Air - line*

Variable: CHRONO.AIR1 (length = 144)

|       |      |       |      |       |      |       |      |       |      |        |      |        |      |        |      |
|-------|------|-------|------|-------|------|-------|------|-------|------|--------|------|--------|------|--------|------|
| ( 1)  | 4718 | ( 19) | 5136 | ( 37) | 5142 | ( 55) | 5576 | ( 73) | 5489 | ( 91)  | 6023 | ( 109) | 5829 | ( 127) | 6306 |
| ( 2)  | 4771 | ( 20) | 5136 | ( 38) | 5193 | ( 56) | 5606 | ( 74) | 5451 | ( 92)  | 6004 | ( 110) | 5762 | ( 128) | 6326 |
| ( 3)  | 4883 | ( 21) | 5063 | ( 39) | 5263 | ( 57) | 5468 | ( 75) | 5587 | ( 93)  | 5872 | ( 111) | 5892 | ( 129) | 6138 |
| ( 4)  | 4860 | ( 22) | 4890 | ( 40) | 5199 | ( 58) | 5352 | ( 76) | 5595 | ( 94)  | 5724 | ( 112) | 5852 | ( 130) | 6009 |
| ( 5)  | 4796 | ( 23) | 4736 | ( 41) | 5209 | ( 59) | 5193 | ( 77) | 5598 | ( 95)  | 5602 | ( 113) | 5894 | ( 131) | 5892 |
| ( 6)  | 4905 | ( 24) | 4942 | ( 42) | 5384 | ( 60) | 5303 | ( 78) | 5753 | ( 96)  | 5724 | ( 114) | 6075 | ( 132) | 6004 |
| ( 7)  | 4997 | ( 25) | 4977 | ( 43) | 5438 | ( 61) | 5138 | ( 79) | 5897 | ( 97)  | 5753 | ( 115) | 6196 | ( 133) | 6033 |
| ( 8)  | 4997 | ( 26) | 5011 | ( 44) | 5489 | ( 62) | 5236 | ( 80) | 5849 | ( 98)  | 5707 | ( 116) | 6225 | ( 134) | 5969 |
| ( 9)  | 4913 | ( 27) | 5182 | ( 45) | 5342 | ( 63) | 5460 | ( 81) | 5743 | ( 99)  | 5875 | ( 117) | 6001 | ( 135) | 5038 |
| ( 10) | 4779 | ( 28) | 5094 | ( 46) | 5252 | ( 64) | 5425 | ( 82) | 5613 | ( 100) | 5852 | ( 118) | 5883 | ( 136) | 6133 |
| ( 11) | 4644 | ( 29) | 5147 | ( 47) | 5147 | ( 65) | 5455 | ( 83) | 5468 | ( 101) | 5872 | ( 119) | 5737 | ( 137) | 6157 |
| ( 12) | 4771 | ( 30) | 5182 | ( 48) | 5268 | ( 66) | 5576 | ( 84) | 5628 | ( 102) | 6045 | ( 120) | 5820 | ( 138) | 6282 |
| ( 13) | 4745 | ( 31) | 5293 | ( 49) | 5278 | ( 67) | 5710 | ( 85) | 5649 | ( 103) | 6146 | ( 121) | 5886 | ( 139) | 6433 |
| ( 14) | 4836 | ( 32) | 5293 | ( 50) | 5278 | ( 68) | 5680 | ( 86) | 5624 | ( 104) | 6146 | ( 122) | 5835 | ( 140) | 6407 |
| ( 15) | 4949 | ( 33) | 5215 | ( 51) | 5464 | ( 69) | 5557 | ( 87) | 5759 | ( 105) | 6001 | ( 123) | 6006 | ( 141) | 6230 |
| ( 16) | 4905 | ( 34) | 5088 | ( 52) | 5460 | ( 70) | 5434 | ( 88) | 5746 | ( 106) | 5849 | ( 124) | 5981 | ( 142) | 6133 |
| ( 17) | 4828 | ( 35) | 4984 | ( 53) | 5434 | ( 71) | 5313 | ( 89) | 5762 | ( 107) | 5720 | ( 125) | 6040 | ( 143) | 5966 |
| ( 18) | 5004 | ( 36) | 5112 | ( 54) | 5493 | ( 72) | 5434 | ( 90) | 5924 | ( 108) | 5817 | ( 126) | 6157 | ( 144) | 6068 |

FIGURE 10

## SARIMA multiplicatif

$$(1 - B)(1 - B^{12})X_t = (1 - \theta_1 B)(1 - \theta_{12} B^{12})\epsilon_t$$

$$\text{avec } \theta_1 = 0,35 \quad \theta_{12} = 0,64.$$

Nous avons changé quatre valeurs de manière à modifier les liaisons d'ordre 1 et d'ordre 12.

Les valeurs en 107 et en 108 ont été changées en 6000 et 6200 (remplaçant 5720 et 5817) tandis que les valeurs en 119 et en 120 ont été changées en 5400 et 5200 (remplaçant 5737 et 5820)

Pour cette série modifiée notée AIRM4, on a une très bonne adéquation par le modèle :

$$(1 - \varphi_{12} B^{12})(1 - B)(1 - B^{12}) \text{ AIRM4} = (1 - \theta_1 B - \theta_2 B^2)(1 - \theta_{12} B^{12})\epsilon_t$$

$$\text{avec } \varphi_{12} = -0,38 \quad \theta_1 = 0,28 \quad \theta_2 = 0,36 \quad \theta_{12} = 0,74$$

(voir en figure 11, 12 et 13 respectivement la fonction d'autocorrélation, la fonction d'autocorrélation partielle, le périodogramme des résidus).

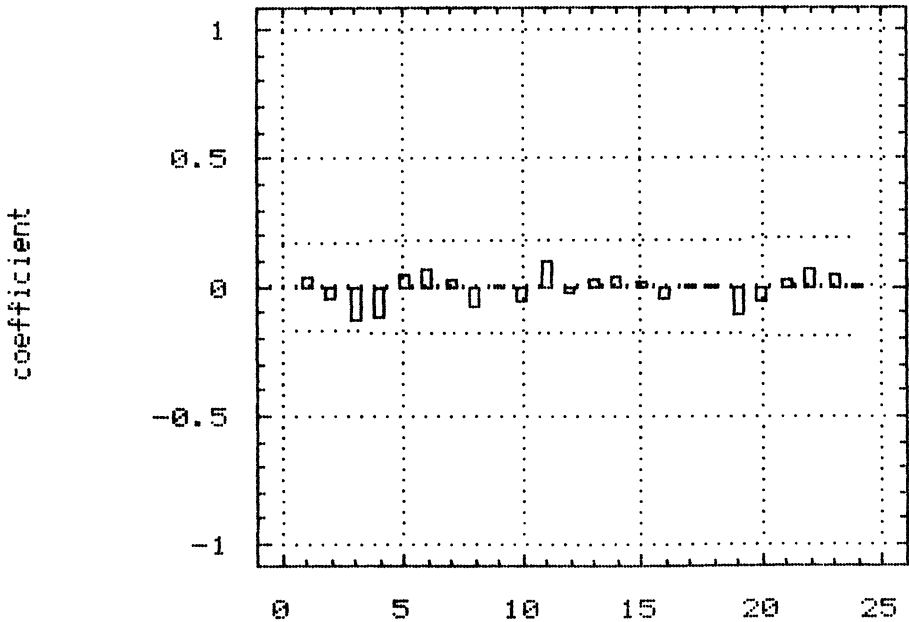


FIGURE 11

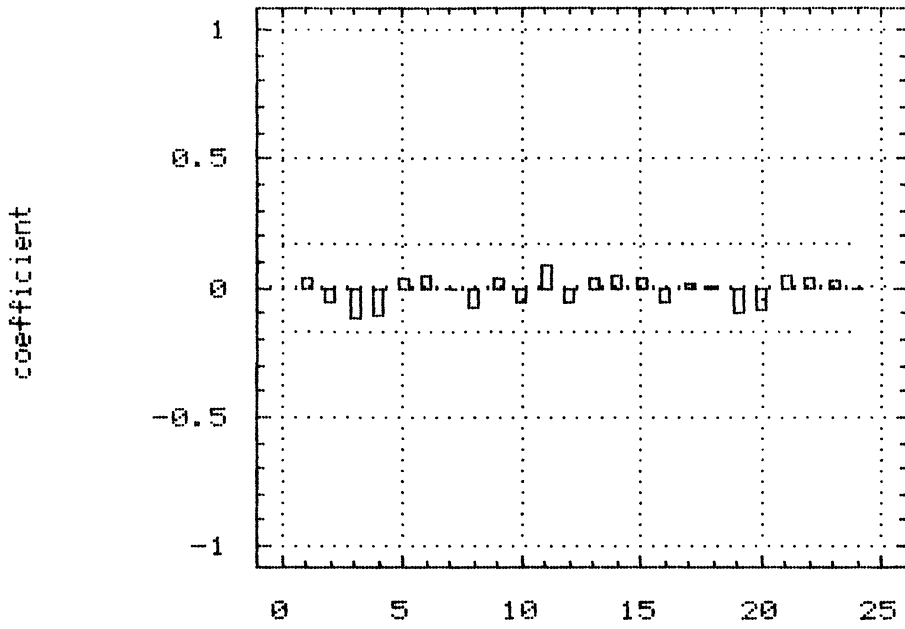


FIGURE 12

(X 10000)

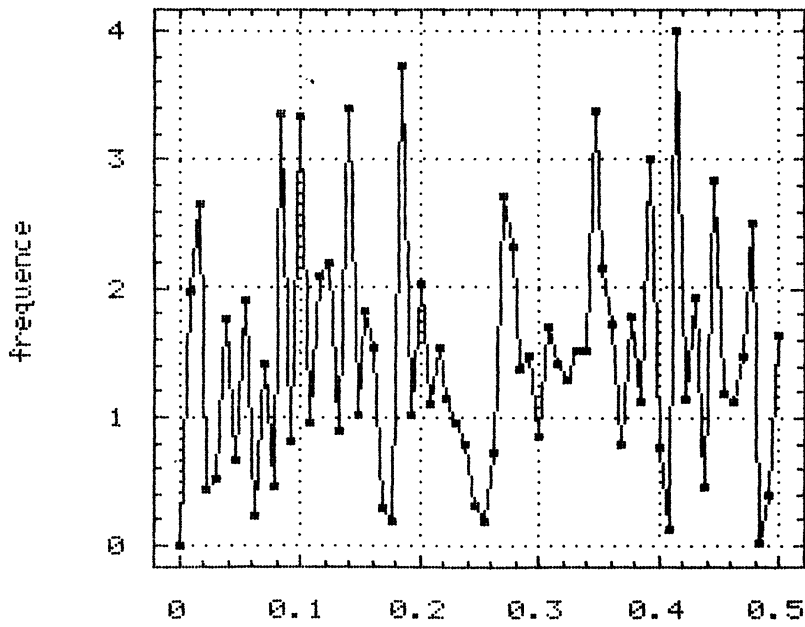


FIGURE 13

On peut d'abord rechercher si l'utilisation du modèle SARIMA de AIR1 :

$$(1 - B)(1 - B^{12})X_t = (1 - \theta_1 B)(1 - \theta_{12} B^{12})\epsilon_t$$

permet de détecter les outliers.

Lorsque l'on prend cette modélisation, les valeurs estimées de  $\theta_1$  et  $\theta_{12}$  sont :  $\theta_1 = 0,33$  et  $\theta_{12} = 0,91$

La liaison d'ordre 12, par rapport à la modélisation de la série AIR1 ( $\theta_1 = 0,35, \theta_{12} = 0,64$ ), est nettement affaiblie. La différenciation  $1 - B^{12}$  est presque compensée par le terme  $(1 - 0,91B^{12})$  ce qui entraîne que seules les valeurs résiduelles en 119, 120 et en 121 ont une valeur absolue dépassant  $3,5\sigma$ .

La plus grande valeur absolue résiduelle se trouve en 121 (576,5), la liaison d'ordre 1 jouant un rôle plus important que la liaison d'ordre 12, la valeur de la série en 121 est surtout comparée à la valeur outlier en 120 et, par rapport à cet outlier, elle est trop importante.

Les valeurs outliers en 119 et en 120 sont détectées comme étant trop faibles (-381,8 -463,8). Elles sont trop faibles et par rapport à la valeur en 118 et par rapport aux valeurs outliers en 107 et en 108. On voit qu'ici les deux liaisons d'ordre 1 et d'ordre 12 interviennent. Si on étudie ensuite les valeurs résiduelles dont la valeur absolue est supérieures à  $2\sigma$ , s'ajoutent aux valeurs précédentes, les valeurs en 107 (qui est un outlier) en 109 et en 110. La plus importante en valeur absolue se produisant en 109 (-306,1), puis en 107 (272,3) et enfin en 110 (-200,1). En effet, la valeur en 109 (5829) apparaît comme trop faible devant l'outlier en 108 (6200). De même l'outlier en 107 (6000) apparaît comme trop grand devant la valeur en 106 (5849) et en 95 (5602). Mais l'outlier en 108 (6200) apparaît comme cohérent avec l'outlier en 107 (6000).

Si l'on modifie maintenant la valeur outlier en 120 de AIRM4 en l'augmentant, lui donnant la valeur 5600 et non 5200, la série obtenue AIRM5 par rapport à la modélisation  $(1 - B)(1 - B^{12})X_t = (1 - \theta_1 B)(1 - \theta_{12} B^{12})\epsilon_t$  a quatre valeurs résiduelles qui en valeur absolue sont supérieures à  $3,5\sigma$  : parmi elles, on trouve les deux valeurs outliers 107 et 119, mais comme précédemment les valeurs outliers en 108 et en 120 qui toutes deux sont cohérentes respectivement avec les valeurs en 107 et en 119 ne sont pas décelées. Tandis que les valeurs en 109 et en 121 importantes par rapport aux valeurs en 108 et en 120 ont une valeur résiduelle de -340,5 et de 334,2.

Si on essaie ensuite de déterminer pour AIRM4 avec le modèle  $(1 - B)(1 - B^{12})X_t = (1 - \theta_1 B)(1 - \theta_{12} B^{12})\epsilon_t$ , si les valeurs importantes résiduelles correspondent bien à un modèle "Additive-outlier" on a une réponse négative. Les valeurs en 121 et 119 sont plus significatives pour un modèle "Innovation-outlier" que pour un modèle "Additive-outlier". En effet, si on considère les valeurs résiduelles situées  $h$  instants après 121 ou 119 ( $h \geq 0$ ), la somme de ces valeurs résiduelles pondérées par le coefficient de  $B^h$  dans  $[(1 - \theta_1 B)(1 - \theta_{12} B^{12})]^{-1}$  et normée par la racine carrée de la somme des carrés de ces coefficients est à comparer à  $3,5\sigma$  pour un modèle A.O. tandis que pour un modèle I.O. c'est la valeur résiduelle elle-même située en 121 ou en 119 qui est à comparer à  $3,5\sigma$ .

pour le modèle A.O.  
 en 121 486,2  
 en 119 -354,9

et :

pour le modèle I.O.  
 en 121 576,5  
 en 119 -381,8

On voit que celles correspondant au modèle I.O. sont plus importantes en valeur absolue que celles correspondant au modèle A.O.

Si l'on reprend l'étude des valeurs résiduelles de AIRM4 avec le modèle bien adapté :

$$(1 - \varphi_{12}B^{12})(1 - B)(1 - B^{12})X_t = (1 - \theta_1B - \theta_2B^2)(1 - \theta_{12}B^{12})\epsilon_t$$

On retrouve globalement les mêmes résultats que pour le modèle

$$(1 - B)(1 - B^{12})X_t = (1 - \theta_1B)(1 - \theta_{12}B^{12})\epsilon_t$$

En 107, 109, 119, 120 et 121 les valeurs absolues résiduelles sont supérieures à  $2\sigma$ .

En 119, 120 et 121 les valeurs absolues résiduelles sont supérieures à  $3,5\sigma$ .

On note toutefois les différences suivantes : la valeur résiduelle en 120 (-423) devient plus importante en valeur absolue que celle en 121 (400) et la valeur résiduelle en 61 (-190) prend la place de la valeur résiduelle en 110.

On constate donc que le modèle :

$$(1 - \varphi_{12}B^{12})(1 - B)(1 - B^{12})X_t = (1 - \theta_1B - \theta_2B^2)(1 - \theta_{12}B^{12})\epsilon_t$$

a donné plus de cohérence à l'ensemble de la série mais qu'il laisse aux valeurs qui apparaissent outlier dans le modèle

$$(1 - B)(1 - B^{12})X_t = (1 - \theta_1B)(1 - \theta_{12}B^{12})\epsilon_t$$

leur apparence d'outlier.

#### IV. Conclusion

Ces trois études de séries illustrent l'effet des outliers sur les liaisons temporelles perçues lors de la modélisation :

- modification totale du modèle pour la série "Assurance Vie"
- disparition de la liaison d'ordre 1 pour la série "taux de variation de consommation de gaz"
- atténuation de la liaison d'ordre 12 pour la série "Air-line"



et montrent que les valeurs résiduelles n'indiquent pas directement les valeurs outliers mais seulement la cohérence ou la non cohérence entre les valeurs de la série d'après le schéma de liaisons temporelles retenues par la modélisation.

### Bibliographie

- ABRAHAM, B., and BOX, G.E.P. (1979) "Bayesian Analysis of some outlier problems in time series" *Biometrika*, 66, 229-236.
- BOX, G.E.P., and JENKINS, G.M. (1976) Time series analysis : forecasting and control, *San Francisco : Holden-Day*.
- FOX, A.J. (1972), "Outliers in time series", *Journal of the Royal Statistical Society, Ser. B. 34*, 350-363.
- GOURIEROUX, C. et MONFORT, A. (1983) Cours de séries temporelles, *Economica Paris*.
- TSAY, R.S. (1936) "Time series model specification in the presence of outliers" *Journal of the American Statistical Association, Vol. 81, N° 393*, 132-141.