

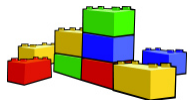
Chapitre 11, exercice 1

Instructions pour employer la série ELF5 du répertoire CH11EX01

Le répertoire CH11EX01 comporte un exercice de base destiné à tous les apprenants et un exercice avancé réservé aux seuls apprenants de la version avancée du cours.

Exercice de base (Pour tous les utilisateurs du cours)

Préalable Le chapitre 11 du cours de base doit avoir été suivi jusqu'à la page 50.



Objectif Le but est de modéliser la série au moyen d'un modèle ARIMA avec une intervention. La méthode de Box et Jenkins est employée à cette fin.



Données Il s'agit de l'évolution des cours de l'action Elf-Aquitaine à la bourse de Paris à raison d'une cotation tous les cinq jours entre le 19 juin 1988 et le 17 octobre 1989, soit 69 données. La variable est notée ELF5. La série a déjà été employée dans l'exercice 3 du chapitre 10.



Structure de l'exercice

L'exercice comporte une seule partie :

- Dans la partie 1, le but de l'exercice est de compléter la spécification du modèle en ajoutant une intervention, d'estimer le paramètre du modèle proposé et d'interpréter le modèle résultant.

Partie 1 Nous commençons par rappeler les résultats obtenus dans le chapitre 10 à propos de cette série. L'objectif est ensuite de revoir la spécification en ajoutant une intervention au modèle, de procéder à l'estimation du modèle, d'interpréter les résultats et de simplifier le modèle obtenu.

1.1 INTRODUCTION

- ⇒ Afin de lancer le logiciel, suivez les instructions données en annexe de l'introduction du cours.
- ⇒ Choisissez le répertoire de données approprié sur votre disque (pas sur le CD-ROM): menu File ⇒ Open. Choisissez DATA puis CHAP11 puis CH11EX01.

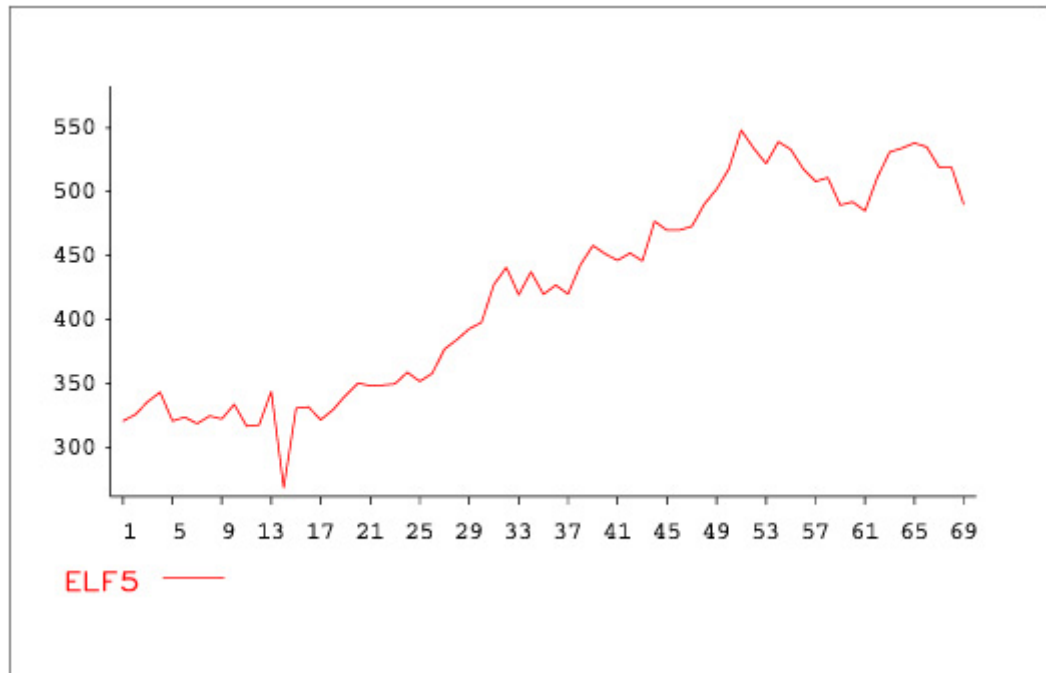


Remarque

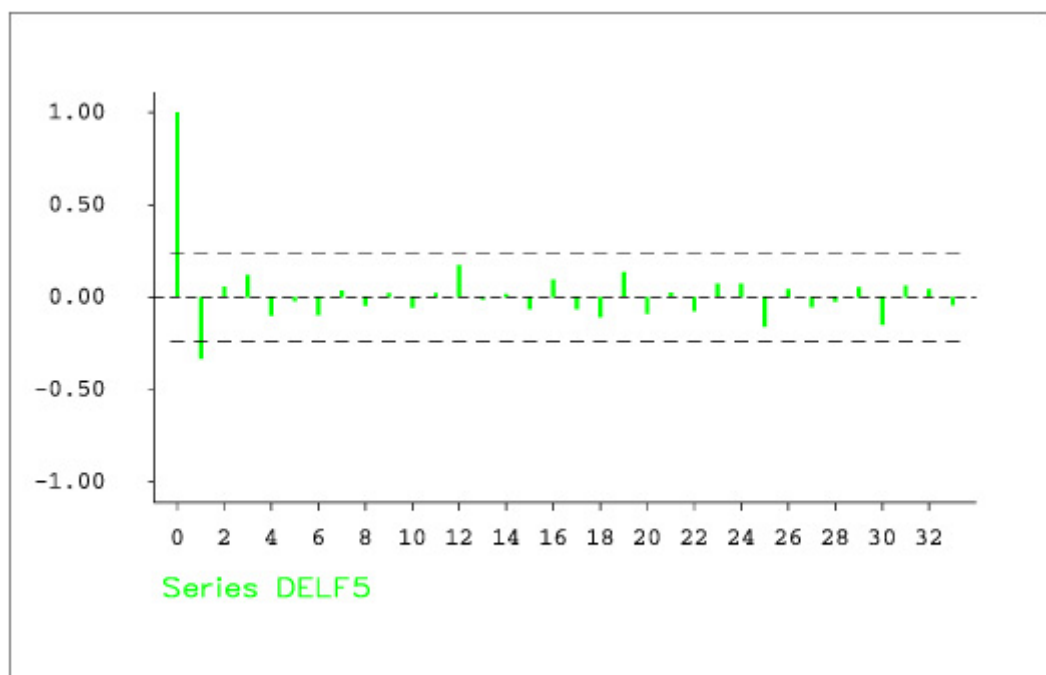
Des instructions plus détaillées ont été données sous forme de remarques lors de la partie 1 de l'exercice 1 du chapitre 10. Prière de s'y référer.

- ⇒ Chargez le problème déjà préparé : ELF5. Vous devez alors voir dans le bas de l'écran que la variable dépendante est ELF5, que les données ne sont pas datées et que l'échantillon d'estimation est 1 – 69 et que les prévisions seront calculées jusqu'en 71.

Rappelons le graphique de la série. Dans l'exercice 3 du chapitre 10, nous avons vu que la série ELF5 n'est pas compatible avec un processus stationnaire. Dans la partie 2 de l'exercice, la série a dû subir une différence ordinaire pour la rendre stationnaire.



Dans la partie 2 de l'exercice, nous avons commencé par examiner les autocorrélations de la série en différence $DEL5 = \nabla ELF5$ que voici.



Cette série ne peut pas être considérée comme produite par un bruit blanc. En considérant que les autocorrélations sont tronquées au-delà du retard 1, nous avons proposé un modèle moyenne mobile d'ordre 1 ou MA(1) pour la série en différence $\nabla ELF5_t = e_t - \theta e_{t-1}$. Nous avons ensuite estimé le paramètre θ pour trouver le modèle suivant : $\nabla ELF5_t = (1 - 0,33B) e_t$.

Cela signifie que les différences contiennent de l'information susceptible

d'améliorer la prévision du cours. Par conséquent, la *promenade aléatoire* ne constitue pas un modèle plausible de la série des cours d'Elf-Aquitaine à raison d'une cotation tous les cinq jours. On se serait attendu au contraire. En effet, sous l'hypothèse d'un marché efficient — c'est-à-dire un marché avec un nombre suffisamment grand d'acteurs, où l'information circule bien et où les coûts de transaction sont négligeables, la *promenade aléatoire* est la règle générale.

Dans la partie 4 de l'exercice, nous avons examiné les résidus. Il était apparent qu'il y avait une valeur atypique au temps 14, un résidu valant $-70,53$, soit près de quatre fois l'écart-type résiduel égal à $0,17$ et de ce fait significatif à moins de $0,01\%$. Ceci correspond au trou du 18 septembre 1988 qui est visible dans le graphique de la série en fonction du temps. La cause de ce trou est une répartition des dividendes.

Grâce à un diagramme de dispersion, nous avons aussi remarqué que l'autocorrélation de retard 1 constatée dans la série est provoquée par cette donnée aberrante.

Enfin, dans la partie 5, nous avons préparé une série appelée ELF5C qui ne possède pas la donnée aberrante remarquée précédemment. La valeur numérique de l'observation au temps 14 avait été simplement modifiée de manière manuelle, en prenant la moyenne des deux observations voisines.

Nous avons remarqué que, comme par miracle, l'autocorrélation de retard 1 de la série en différence disparaissait. L'idée du présent exercice est d'effectuer cette correction autrement.

1.2 UNE VARIABLE DE RÉGRESSION POUR LA CORRECTION D'UNE DONNÉE

La démarche que nous allons employer a déjà été utilisée dans le chapitre 7 relatif à la régression linéaire multiple. Elle a été rappelée dans le cours, aussi n'est-il pas nécessaire de la détailler ici. Elle consiste à employer une variable binaire égale à 1 au temps où l'événement atypique s'est produit et égale à 0 ailleurs. Dans ce cas-ci, on définit une variable — qu'on appelle $I14$ — comme suit : $I14_t = 1$ si $t = 14$, et $I14_t = 0$ si $t \neq 14$. L'impact de l'événement sera quantifié par un paramètre qu'il faudra estimer. Notons b_{14} ce paramètre. En négligeant pour le moment l'autocorrélation, un modèle de régression pour la variable ∇ELF5_t serait le suivant : $\nabla \text{ELF5}_t = b_{14} I14_t + e_t$.

En tenant compte de l'autocorrélation, on aura le modèle suivant :

$$\nabla \text{ELF5}_t = b_{14} I14_t + e_t - \theta e_{t-1},$$

ou

$$\nabla \text{ELF5}_t - b_{14} I14_t = e_t - \theta e_{t-1}.$$

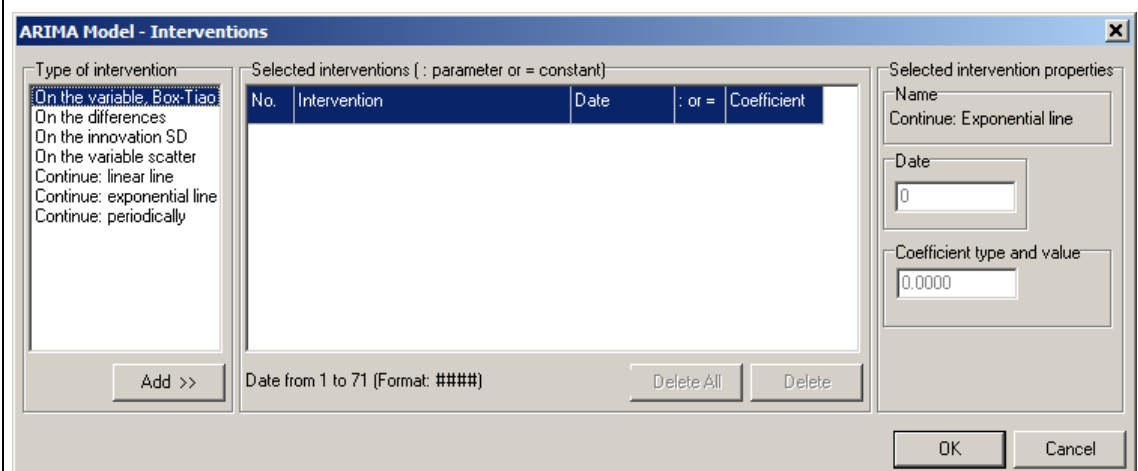
Il y a deux paramètres à estimer : b_{14} et θ . Nous verrons dans la partie A de l'exercice avancé comment l'estimation peut être effectuée. Nous nous contentons ici d'observer les résultats, qui sont obtenus par des méthodes numériques itératives.

Nous employons à cette fin Time Series Expert. Dans les autres exercices, nous emploierons aussi le logiciel Demetra qui sera utilisé pour la désaisonnalisation dans les chapitres 12 et 13.

1.3 ESTIMATION DU MODÈLE

Nous allons procéder à l'estimation des deux paramètres b_{14} et θ .

- ⇒ Pour accéder à la boîte de dialogue pour l'estimation : menu Methods ⇒ Arima model ⇒ Estimation. Dans la fenêtre de dialogue Arima model – Estimation, vous avez 0 en face de Differences et 0 en face de Seasonal differences.
- ⇒ Utilisez les triangles pointés vers le haut et vers le bas sur la ligne Differences pour faire apparaître 1.
- ⇒ Cliquez Add Parameters pour ouvrir la fenêtre de dialogue. La zone Type apparaît. Cliquez sur la ligne Ordinary moving average. Cliquez Add. La fenêtre montre le type de paramètre sélectionné, Ordinary moving average, l'ordre 1 et la valeur initiale du coefficient 0.0000. Cliquez OK pour quitter cette fenêtre.
- ⇒ Cliquez sur Add Interventions pour ouvrir la fenêtre de dialogue. La zone Type apparaît.



⇒ Cliquez sur la ligne On the variable : Box-Tiao (I). Cliquez Add. Cliquez dans la zone date où 0 est affiché. Tapez 14. Cliquez OK. Cliquez dans la zone Coefficient type and value et choisissez Parameter. Cliquez OK.

L'intervention sur la variable, à la date 14, est validée.

Remarques



1. En déplaçant le curseur à droite avec la flèche vers la droite, on peut modifier la valeur initiale. Dans ce cas-ci, il ne faut rien modifier. Outre un paramètre, le programme permet aussi de spécifier une constante, et aussi + ou – une autre intervention préalablement définie.
2. Après avoir entré une première intervention, il est possible d'en entrer une autre.
3. Nous n'emploierons ici que l'intervention sur la variable qui suffit pour corriger une donnée aberrante ou effet AO — additive outlier. Dans un autre exercice, nous produirons d'autres effets en employant d'autres types d'interventions.
4. Il est aussi possible de supprimer une intervention dans la liste. Se positionner sur la ligne appropriée et cliquez Delete.
5. Les interventions choisies restent actives durant toute une session. Elles sont sauvegardées avec le problème.

⇒ Pressez une seconde fois Enter pour revenir à la fenêtre de dialogue Arima model – Estimation puis cliquez OK pour lancer le programme. Nous allons consulter le fichier de sortie.

Après l'estimation, le fichier de sortie apparaît. Nous le reprenons en an-



nexe de cette partie, pour la commodité.

Remarque

Le format est identique à celui discuté dans l'exercice 1 du chapitre 10 à l'exception des rubriques relatives aux interventions.

Ces fichiers ont la structure suivante :

1. Un en-tête reprenant notamment le nom et l'intitulé de la série, sa longueur, et pour les données datées, les dates de début et de fin.
2. Les paramètres et leurs valeurs initiales, ici au nombre de deux appelés MA1 et KI14.
3. La méthode d'estimation utilisée, basée sur la méthode du maximum de vraisemblance exacte.
4. La description du modèle : on mentionne ici l'intervention, la différence, une constante additive et le modèle ARMA
5. Les itérations de l'estimation non linéaire : on voit bien ici l'évolution de l'estimation et que la convergence (à la précision relative de $0,001 = 10^{-3}$) a été atteinte après 6 itérations.
6. La matrice de corrélation : elle peut avoir de l'intérêt puisqu'il y a deux paramètres.
7. Les valeurs finales des paramètres, ici des deux paramètres θ et b_{14} , notés MA1 et KI14, avec les estimations respectives 0,064 et $-67,7$, les erreurs-types associées 0,132 et 10,1, les statistiques de Student correspondantes $0,064/0,132 = 0,5$ et $-67,7/10,1 = -6,7$, la première comprise entre $-1,96$ et $1,96$, donc non significative à 5 %, et la seconde inférieure à $-1,96$, donc significative à 5 %. Les intervalles de confiance à 95 %, le premier qui contient 0 et le second qui ne contient pas 0, confirment ces conclusions.
8. L'estimation de la constante, sous l'intitulé MEAN, valant ici 2,49.
9. Les mesures de synthèse, notamment l'écart-type résiduel sous l'intitulé 'Standard deviation', ici 13,58.
10. L'analyse des résidus, similaire à ce qui a été discuté au chapitre 10.
11. Les résultats d'ajustement et de prévision, notamment les intervalles de prévision d'horizon 1 et d'origine variable, similaire à ce qui a été discuté au chapitre 10.

Il est recommandé de sauvegarder le fichier, sous le nom ELF5INT.

- ⇒ Dans la fenêtre de sortie, clic droit, choisissez Save as.
- ⇒ Tapez ELF5INT. Cliquez Save.
- ⇒ Descendez dans le fichier pour repérer les différentes rubriques.

? Vérifiez les résultats qui ont été cités.



1.3.1 Votre réponse

? Ecrivez l'équation du modèle estimé. Interprétez les résultats de l'estimation.



1.3.2 Votre réponse

On recommande de sauvegarder le problème pour le charger de nouveau si l'exercice est interrompu.

- ⇒ Sauvegardez le problème modifié : menu File ⇒ Save as. Tapez ELF5INT et cliquez Save.

1.4 VALIDATION DU MODÈLE

Il convient de consulter les points suivants de l'analyse des résidus. Si l'exercice est interrompu, il faut accéder à la sortie de la façon suivante.



⇒ Pour accéder au fichier de résultats : menu Reports ⇒ Statistic report. Tapez ELF5INT et cliquez Open. Le fichier s'ouvre. Descendez sur la ligne "=== Residual Analysis".

?

Notez la valeur de l'écart-type résiduel.

1.4.1 Votre réponse



⇒ Pour visualiser le graphique des résidus en fonction du temps : menu Graphics ⇒ Series et spécifiez le nom Resid.

?

Comparez chacun des résidus avec l'écart-type résiduel. Peut-on accepter la normalité des erreurs ?

1.4.2 Votre réponse



⇒ Pour visualiser les autocorrélations de la série résiduelle : menu Graphics ⇒ Residual autocorrelations and partials.

?

Que peut-on dire des autocorrélations résiduelles ?

1.4.3 Votre réponse

Nous allons reprendre l'étude d'un modèle simplifié.

1.5 ESTIMATION D'UN SECOND MODÈLE

Ce modèle ne va pas comporter le paramètre θ , seulement le paramètre b_{14} .

?

Pourquoi supprimer le paramètre θ alors qu'il était significatif dans l'exercice 3 du chapitre 10.

1.5.1 Votre réponse



Nous passons par la boîte de dialogue des modèles ARIMA, en procédant comme suit.

- ⇒ Pour accéder à la boîte de dialogue pour l'estimation : menu Methods ⇒ Arima model ⇒ Estimation. Dans la fenêtre de dialogue Arima model – Estimation, vous avez 1 en face de Differences et 0 en face de Seasonal differences.
- ⇒ Cliquez Add Parameters pour ouvrir la fenêtre de dialogue. La fenêtre montre le type de paramètre sélectionné, Ordinary moving average, l'ordre 1 et la valeur initiale du coefficient 0.0000. Pour le supprimer, cliquez Delete. Cliquez OK.
- ⇒ Cliquez Add Interventions pour ouvrir la fenêtre de dialogue. La fenêtre montre l'intervention sur la variable au temps 14. NE PAS LA SUPPRIMER. Cliquez OK.
- ⇒ Cliquez OK pour lancer le programme. Nous allons consulter le fichier de sortie.

?

Commentez vos observations relatives au nombre d'itérations, la valeur finale du paramètre, les autocorrélations résiduelles.

1.5.2 Votre réponse

- ⇒ Pour visualiser les autocorrélations de la série résiduelle : menu



Graphics \Rightarrow Residual autocorrelations and partials.

?

Vérifiez quels sont les retards pour lesquels les autocorrélations sont significatives au seuil de 5%.

1.5.3 Votre réponse



\Rightarrow Pour visualiser les autocorrélations partielles de la série résiduelle, pressez Enter. Pour revenir ensuite aux autocorrélations, pressez encore une fois Enter.

?

Vérifiez quels sont les retards pour lesquels les autocorrélations partielles sont significatives au seuil de 5%.

1.5.4 Votre réponse

?

Pourrait-on encore améliorer le modèle? Quelles conclusions tirez-vous de ces résultats ?

1.5.5 Votre réponse



SYNTHÈSE

Nous avons conclu l'exercice 3 du chapitre 10 par le fait que la série ELF5 n'est pas compatible avec une *promenade aléatoire*, essentiellement à cause d'une seule donnée qui apparaît comme aberrante. Nous avons déjà remarqué que l'autocorrélation de retard 1 constatée dans la série est provoquée par cette donnée aberrante. La série des résidus montre clairement un résidu égal à -68 au point 14, soit près de quatre fois l'écart-type résiduel égal à $0,17$. Ceci correspond au trou du 18 septembre 1988 qui est visible dans le graphique de la série en fonction du temps.



L'analyse d'intervention qui a été réalisée confirme ceci. En introduisant une variable de régression de type binaire, on corrige la donnée aberrante. Ce n'est pas un hasard si l'estimation du coefficient de régression correspondant vaut -68 , proche des -71 du résidu du premier modèle.

Annexe

Résultats de l'estimation d'un modèle MA(1) avec intervention pour la série VELF5

```

ESTIM,READ,AUXIN=1,KEY=ELF5,RESIT=1,RESIN=Resid,FITF=1,FITL=1,FITN=Fit,/
TAPEF=1,NFORC=Forc,DIFF=1,CHECK=FAP$X,LEAD= 2,KNOW
RANDOM NOISE GENERATOR: INITIAL SEED IS "?UN@vh?zx" ( 1423, 5609, 3801)
////////////////////////////////////
ANSECH-PC 2.3c, AUTHOR:G.MELARD 23/09/01 14:56:38. PROBLEM( 1):ELF5
TITLE: "cCOURS D'ELF-AQUITAINE 19/6/88-17/10/89"
SERIES READ FROM DISK, NAME IS ELF5.DB, LENGTH 69
////////////////////////////////////
===KNOWLEDGE ABOUT INTERVENTIONS (BOX-TIAO)
DIRECTIVE TYPE DATE STEP NATURE PARAM/VALUE COMMENTS
I 14 BOX-TIAO 14 VALUE KI 14:
1 DIRECTIVE(S), 1 PARAMETER(S), 0 CONSTANT(S).
2 PARAMETERS WITH STARTING VALUES :
1 MA 1 .00000
2 KI 14 .00000
=== ESTIMATION BY MAXIMIZATION OF THE EXACT (LOG) LIKELIHOOD
(FAST ALGORITHM WITH TOLERANCE 1.0E-05)
=== MODEL DESCRIPTION FORM DEGREE/ORD PARAMETERS NUMBER
- BOX-TIAO INTERVENTION SEE ABOVE 1 KIddd 1
- DIFFERENCE REGULAR 1
- ADDITIVE CONSTANT AUTOMATIC
- ARMA MODEL MOVING AVERAGE POLYNOMIAL REGULAR 1 MA nn 1
*** WARNING-THE INFORMATION MATRIX WILL BE COMPUTED FROM 1ST ORDER DERIVATIVES
*** WARNING-THE INFORMATION MATRIX WILL BE COMPUTED FOR A GAUSSIAN MODEL
NON LINEAR ESTIMATION:
ITER SUM OF SQ MA 1 KI 14
0 2142.4E+01 .000 .000
1 1202.1E+01 1.822E-02 -66.9
2 1199.3E+01 6.714E-02 -67.4
3 1199.3E+01 6.353E-02 -67.7
4 1199.3E+01 6.394E-02 -67.7
5 1199.3E+01 6.390E-02 -67.7
6 1199.3E+01 6.391E-02 -67.7
=ITERATION STOPS - RELATIVE CHANGE IN EACH COEFFICIENT LESS THAN 1.00000E-03
CORRELATION MATRIX
MA 1 KI 14
MA 1 1.00
KI 14 .21 1.00
FINAL VALUES OF THE PARAMETERS WITH 95% CONFIDENCE LIMITS
NAME VALUE STD ERROR T-VALUE LOWER UPPER
1 MA 1 6.39055E-02 .13239 .5 -.20 .33
2 KI 14 -67.679 10.132 -6.7 -88. -47.
ESTIMATION HAS TAKEN .1 SEC. FOR 19 EVALUATIONS OF S.S. (MEAN TIME=, .003)
N.B. QUICK RECURSIONS USED FROM TIME 5
THE FOLLOWING PARAMETERS WERE ESTIMATED SEPARATELY
MEAN 2.4853
THE FOLLOWING CONSTANTS WERE INVOLVED IN THE LEAST SQUARES ESTIMATION METHOD
ARMA .99997
=== SUMMARY MEASURES <V>
SUM OF SQUARES : COMPUTED = 11993.2 ADJUSTED = 11992.5
VARIANCE ESTIMATES : BIASED = 176.360 UNBIASED = 184.500
TOTAL NUMBER OF PARAMETERS = 3 STANDARD DEVIATION = 13.5831
INFORMATION CRITERIA : AIC = 560.840 SBIC = 569.908
=== RESIDUAL ANALYSIS WITH 68 RESIDUALS, BEGINNING AT TIME 2===
MEAN = .317595E-01 ,T-STATISTIC = .02 (FOR TESTING ZERO MEAN)
=OUTLIERS <R(OR)S>
1 - 5 % 31: 27.20 44: 27.88 51: 28.42
69: -31.72
=SIGNIFICANT AUTOCORRELATIONS (USING BARTLETT LIMITS) <A(OR)S>

```

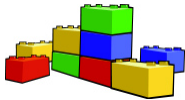
[illegible]



Exercice avancé

(Pour les utilisateurs de la version avancée du cours)

Préalable Le chapitre 11 du cours de base doit avoir été suivi jusqu'à la page 135.



Objectif



Le but est de visualiser la corrélation avec retard sur les résidus et de représenter les autocorrélations d'une méthode qui a été traitée dans la partie avancée des exercices.

Données Les mêmes données que celles utilisées pour l'exercice de base.



Structure de l'exercice

L'exercice avancé une seule partie :

- Dans la partie A, le but de l'exercice est d'expliquer les différentes approches d'estimation des paramètres d'un modèle ARIMA avec interventions ou de modèles de régression à erreurs ARIMA, sur l'exemple d'une intervention unique sur un modèle simple, de type moyenne mobile d'ordre 1 ou autorégressive d'ordre 1.

Partie A Le but de l'exercice est d'expliquer les différentes approches d'estimation des paramètres d'un modèle ARIMA avec interventions ou de modèles de régression à erreurs ARIMA, sur l'exemple d'une intervention unique sur un modèle simple, de type moyenne mobile d'ordre 1 ou de type autorégressif d'ordre 1.

Nous présentons deux approches possibles.

A.a L'APPROCHE DES MODÈLES NON LINÉAIRES

Considérons d'abord un processus moyenne mobile, d'ordre 1. Il dépend d'un paramètre θ . L'équation est $y_t = e_t - \theta e_{t-1}$, où les e_t constituent un processus bruit blanc de moyenne 0 et de variance σ^2 . Nous avons vu, dans la présentation du chapitre 9, qu'un tel modèle est non linéaire par rapport au paramètre θ . C'est cette propriété qui nous empêche de déterminer l'estimation au sens des moindres carrés. Nous avons trouvé comme remède l'estimation des moindres carrés non linéaires qui consiste à chercher numériquement la valeur de θ qui minimise la somme des carrés des erreurs $S(\theta)$, définie par

$$S(\theta) = \sum_{t=2}^T e_t^2$$

où l'erreur e_t s'obtient par la récurrence $e_t = y_t + \theta e_{t-1}$, pour $t = 2, \dots, T$, et e_1 est fixé, par exemple à 0.

Considérons maintenant la présence d'une variable explicative x . Supposons le modèle de régression linéaire $y_t = b_0 + b_1 x_t + u_t$, où u_t est un terme d'erreur, comme e_t , sauf que nous allons supposer que les erreurs u_t sont de moyennes nulles et de variance constante mais qu'elles ne sont pas indépendantes. Elles sont par exemple déterminées par un processus moyenne mobile d'ordre 1, décrit par l'équation suivante $u_t = e_t - \theta e_{t-1}$, où les e_t constituent un processus bruit blanc de moyenne 0 et de variance σ^2 . On peut encore employer l'estimation des moindres carrés non linéaires qui consiste à chercher numériquement la valeur des paramètres θ , b_0 et b_1 qui minimise la somme des carrés des erreurs $S(\theta, b_0, b_1)$, définie par

$$S(\theta, b_0, b_1) = \sum_{t=2}^T e_t^2$$

où l'erreur e_t s'obtient par la récurrence $e_t = y_t - (b_0 + b_1 x_t) + \theta e_{t-1}$, pour $t = 2, \dots, T$, et e_1 est fixé, par exemple à 0.

C'est la méthode utilisée par Time Series Expert.



Remarques

1. Nous avons mentionné au chapitre 10 que l'estimation des paramètres doit peut s'effectuer, de préférence, par la méthode du maximum de vraisemblance exacte, *exact maximum likelihood*, qui nécessite de supposer la distribution des innovations, n'est simple dans le contexte des modèles ARMA que dans le cas d'une distribution normale mais ne requiert pas de données de départ ni surtout d'erreur de départ. On la recommande fortement pour des séries courtes, de longueur $T < 50$, même si les calculs sont plus complexes que celui de la somme des carrés des erreurs. La méthode n'est pas trop sensible à la supposition de normalité des innovations. La même remarque s'applique également ici.
2. La méthode précédente fonctionne pour tout modèle ARMA(p, q) mais le nombre de paramètres à estimer peut être élevé : $p + q + k$, où k est le nombre de coefficients de régression.

A.b L'APPROCHE DES MODÈLES LINÉAIRES

Le cas d'un processus autorégressif d'ordre 1 est plus facile. Il dépend d'un paramètre ϕ . L'équation du modèle est $u_t = \phi u_{t-1} + e_t$, où les e_t constituent un processus bruit blanc de moyenne 0 et de variance σ^2 . Nous avons vu, dans la présentation du chapitre 9, qu'un tel modèle est linéaire par rapport au paramètre ϕ . On peut donc ici déterminer l'estimation au sens des moindres carrés, qui consiste en l'occurrence à employer comme estimation l'autocorrélation de retard 1.

Considérons maintenant la présence d'une variable explicative x . Supposons le modèle de régression linéaire $y_t = b_0 + b_1 x_t + u_t$, où u_t est un terme d'erreur, comme e_t , sauf que nous allons supposer que les erreurs u_t , sont de moyennes nulles et de variance constante mais qu'elles ne sont pas indépendantes.

Supposons pour commencer que les erreurs u_t consistent en un processus AR(1) au lieu d'un processus bruit blanc. Cela signifie que $u_t = \phi u_{t-1} + e_t$, où les e_t constituent un processus bruit blanc de moyenne 0 et de variance σ^2 . Nous allons montrer qu'il est alors possible d'obtenir une estimation pour les coefficients de régression b_0 et b_1 . En effet, écrivons l'équation du modèle de régression au temps $t - 1$: $y_{t-1} = b_0 + b_1 x_{t-1} + u_{t-1}$, d'où en multipliant cette équation par ϕ : $\phi y_{t-1} = \phi b_0 + \phi b_1 x_{t-1} + \phi u_{t-1}$, et en soustrayant membre à membre : $y_t - \phi y_{t-1} = b_0 - \phi b_0 + b_1 (x_t - \phi x_{t-1}) + u_t - \phi u_{t-1}$, ce qui peut s'écrire :

$$y_t - \phi y_{t-1} = b_0 (1 - \phi) + b_1 (x_t - \phi x_{t-1}) + e_t.$$

Pour un ϕ donné, on peut calculer :

$$Y_t = y_t - \phi y_{t-1}, \quad U_t = 1 - \phi, \quad \text{et} \quad X_t = x_t - \phi x_{t-1}.$$

On a donc une régression de la forme

$$Y_t = b_0 U_t + b_1 X_t + e_t,$$

un modèle de régression linéaire multiple à deux variables explicatives et sans constante, où cette fois les erreurs e_t sont indépendantes. Pour un ϕ donné, on peut donc estimer b_0 et b_1 par la méthode des moindres carrés, de manière exacte. Comme ϕ n'est pas connu, il faut développer une méthode itérative, par exemple de la manière suivante.

- 1) On effectue une régression linéaire des y_t en fonction des x_t , par la méthode des moindres carrés ;
- 2) On répète jusqu'à convergence les étapes suivantes :
 - a) On estime ϕ comme autocorrélation de retard 1 des résidus e_t ;
 - b) Pour cette valeur de ϕ , on calcule les variables $Y_t = y_t - \phi y_{t-1}$, $U_t = 1 - \phi$, et $X_t = x_t - \phi x_{t-1}$;
 - c) On effectue la régression $Y_t = b_0 U_t + b_1 X_t + e_t$, par la méthode des moindres carrés .

Dans la littérature économétrique, cette approche porte les noms de Hildreth-Lu.

Cette approche n'est pas utilisable si l'on remplace le processus autorégressif par un processus moyenne mobile. Néanmoins, on peut décomposer le problème en deux problèmes séparés : un problème de régression et un problème d'estimation de paramètres ARMA, comme nous allons le voir.

Commençons par traiter un problème de régression ordinaire, par exemple $y_t = b_0 + b_1 x_t + e_t$, où les erreurs e_t , sont de moyennes nulles et de variance constante. Nous avons vu que la méthode des moindres carrés consiste ici à écrire, en supposant T données :

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_T \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_T \end{pmatrix}.$$

et l'estimateur recherché est $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$. Nous avons vérifié ceci dans

l'exercice 1 du chapitre 7, partie E.

Considérons maintenant des erreurs qui ne sont pas indépendantes mais qui constituent un processus moyenne mobile, d'ordre 1. Il dépend d'un paramètre θ . L'équation du processus est $y_t = e_t - \theta e_{t-1}$, où les e_t constituent un processus bruit blanc de moyenne 0 et de variance σ^2 . Nous avons vu, dans la présentation du chapitre 9, que la variance et les autocovariances d'un tel processus sont faciles à obtenir. La variance vaut $(1 + \theta^2) \sigma^2$, l'autocovariance de retard 1 vaut $-\theta \sigma^2$, et les autocovariances de retard supérieur à 1 sont nulles. On peut introduire la *matrice d'autocovariance* d'un processus (nous l'avons fait à propos de la définition des autocorrélations partielles). Commençons par le cas d'un processus bruit blanc de variance σ^2 . La *matrice d'autocovariance* des variables aléatoires e_1, e_2, \dots, e_T est une matrice de dimensions $T \times T$ contenant des zéros partout sauf sur la diagonale, qui contient σ^2 :

$$\begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix}.$$

De même, la *matrice d'autocovariance* des variables y_1, y_2, \dots, y_T est une matrice de dimensions $T \times T$ qui a la forme suivante :

$$\Sigma = \begin{pmatrix} (1+\theta^2)\sigma^2 & -\theta\sigma^2 & \dots & 0 \\ -\theta\sigma^2 & (1+\theta^2)\sigma^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & (1+\theta^2)\sigma^2 \end{pmatrix}.$$

Dans ce cas, et en supposant θ et σ^2 connus — nous lèverons cette supposition plus loin — on peut obtenir l'estimateur des moindres carrés généralisés, *generalised least squares* ou GLS, des paramètres, ici b_0 et b_1 . Cet estimateur s'écrit :

$$\hat{\mathbf{b}} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \mathbf{X}'\Sigma^{-1}\mathbf{y}.$$

En fait σ^2 se simplifie dans cette expression donc sa connaissance ne joue aucun rôle. En revanche, θ ne se simplifie pas. Pour l'estimer et estimer également les deux autres paramètres b_0 et b_1 , on peut employer la méthode itérative suivante :

- 1) On effectue une régression linéaire des y_t en fonction des x_t , selon l'équation $y_t = b_0 + b_1 x_t + u_t$, par la méthode des moindres carrés ;
- 2) On répète jusqu'à convergence les étapes suivantes :
 - a) On estime θ comme paramètre moyenne mobile à partir des résidus

u_t , par la méthode du chapitre 9, donc en minimisant la somme des carrés des erreurs $S(\theta)$, définie par

$$S(\theta) = \sum_{t=2}^T e_t^2$$

où l'erreur e_t s'obtient par la récurrence $e_t = y_t - \theta e_{t-1}$, pour $t = 2, \dots, T$, et e_1 est fixé, par exemple à 0 ;

b) Pour cette valeur de θ , on calcule la matrice de covariance Σ (au facteur σ^2 près) et on estime les deux autres paramètres b_0 et b_1 , par la méthode des moindres carrés généralisés ;

c) On détermine les résidus u_t par différence $y_t - (b_0 + b_1 x_t) = u_t$.

C'est la méthode utilisée par l'approche RegARIMA du module X-12-ARIMA de Demetra. Elle a l'avantage que les paramètres ARMA sont les seuls qui sont l'objet de l'estimation non linéaire.

SYNTHÈSE

Nous avons présenté de manière succincte les deux méthodes qui peuvent être employées pour estimer les paramètres d'un modèle de régression à erreurs ARMA.

[Retour au chapitre 11](#)