Initiation à l'analyse des séries temporelles et à la prévision

Guy Mélard
ECARES CP114 et Institut de Recherche en Statistique
Université Libre de Bruxelles
Avenue Franklin Roosevelt 50
B-1050 Bruxelles
e-mail : gmelard@ulb.ac.be

Résumé

Nous présentons l'analyse des séries temporelles qui est employée dans de nombreuses sciences et techniques. Nous mettons notamment l'accent sur les méthodes de prévision telles qu'elles sont utilisées, en particulier pour la prévision des ventes dans les entreprises. Nous insistons sur les aspects statistiques et économétriques de ces méthodes et sur leurs limites respectives. Nous abordons également l'analyse spectrale. Outre quelques illustrations simples des méthodes, nous donnons des indications sur les logiciels employés, les cours disponibles et quelques aspects de calcul.

Abstract

Time Series Analysis is applied in numerous sciences and techniques. It is introduced here with an accent of forecasting methods as they are used, in particular by companies interested in sales forecasting. We focus on statistical and econometric aspects of these methods and their limitations. Spectral analysis is also considered. Beside some simple illustrations of these methods, we provide indications on software being used, on available courses and on some computational aspects.

1 Introduction

1.1 Objectifs de cet article

Cet article a pour but de présenter l'analyse des séries temporelles. L'analyse des séries temporelles est un domaine de la statistique, de l'économétrie et des sciences de l'ingénieur qui est très employé dans de nombreuses sciences et techniques. On pourrait même dire qu'elle n'est pas assez employée, compte tenu de ses possibilités. Ici, nous mettons notamment l'accent sur les méthodes de prévision telles qu'elles sont utilisées, en particulier pour la prévision des ventes dans les entreprises, mais nous traitons également d'autres exemples. Nous insistons sur les aspects statistiques et économétriques des méthodes purement temporelles et sur leurs limites respectives. Nous abordons également l'analyse spectrale. Outre quelques illustrations simples des méthodes, nous donnons des indications sur les cours disponibles, les logiciels employés et quelques aspects de calcul.

1.2 Description générale des methodes couvertes

L'analyse des séries temporelles, et plus particulièrement la prévision à court et moyen terme, a connu des développements importants depuis trente ans. La diffusion de logiciels spécialisés la met à la portée de toutes les organisations. La prévision est fondamentale dans la mesure où elle est à la base de l'action. La prise de décision doit en effet toujours reposer sur des prévisions. C'est ainsi qu'une entreprise commerciale s'intéresse aux prévisions des ventes futures pour faire face à la demande, gérer sa production et ses stocks, mais aussi orienter sa politique commerciale (prix, marketing, produits, etc.). Il s'agit ici de prévision à court terme. De même, on essaie de prévoir le rendement d'un investissement, la pénétration d'un marché ou l'effet du passage aux 35 heures. Il

s'agit alors de prévisions à moyen terme. Enfin, on peut envisager des prévisions à long terme comme la prévision des besoins en services publics (hôpitaux, écoles, etc.). Nous nous limitons ici à la prévision à court et moyen terme, renonçant dès lors aux modèles de population, aux méthodes qualitatives et technologiques de prévision et aux grands modèles économétriques. Ces deux dernières approches sont d'ailleurs très coûteuses en temps humain et ne s'avèrent pas meilleures en général.

La plupart des méthodes que nous étudierons sont relatives à la prévision de séries chronologiques ("*time series*"), qu'on appelle aussi séries temporelles ou chroniques.

Il est possible d'exposer les méthodes de prévision à plusieurs niveaux, le ton simple et l'approche générale de Granger (1980), la présentation très complète, basée sur des exemples, de Makridakis, Wheelwright et Hyndman (1997), la synthèse de Coutrot et Droesbeke (1990), Droesbeke *et al.* (1990), Bourbonnais et Terraza (2004), ou celui des ouvrages plus avancés sur la théorie statistique et économétrique comme Gourieroux et Monfort (1990), Lütkepohl (1993) ou Hamilton (1994). Notre texte est largement inspiré de Mélard (1990a). En revanche, la plupart des exemples traités sont neufs et certains seront incorporés dans la nouvelle édition de ce dernier livre, en préparation.

Les méthodes de prévision sont souvent subdivisées en catégories. On distingue notamment les courbes de croissance, les moyennes mobiles, la décomposition saisonnière, le lissage exponentiel, la régression multiple, la méthode de Box et Jenkins, pour se limiter à ce qui est couvert ici. Cette classification est parfois poussée à l'extrême, comme dans la compétition de Makridakis *et al.* (1984) qui a mis en présence une quinzaine de méthodes sur 1001 séries chronologiques. C'est l'ensemble d'information utilisé qui permet surtout de distinguer les méthodes de prévision. Nous en parlerons dans la première partie de l'exposé.

Il est fondamental, à notre avis, d'envisager les modèles sous-jacents aux méthodes de prévision et d'envisager d'estimer les paramètres de ces méthodes en employant une approche statistique. La modélisation est l'objet de notre deuxième partie qui sera illustrée sur des exemples. Nous insistons surtout sur la prévision probabiliste (la détermination de la distribution de probabilité de la valeur future, y compris le résumé qu'on appelle intervalle de prévision).

Les illustrations sont réalisées en employant notamment le logiciel TSE, Time Series Expert, que nous avons développé en collaboration principalement avec Jean-Michel Pasteels. On peut en trouver une version de démonstration à l'adresse

`ftp.ulb.ac.be/pub/packages/tse`

La plupart des logiciels de prévision, comme les logiciels statistiques et économétriques, permettent rapidement de réaliser ces analyses sur micro-ordinateur mais les concepts de base des méthodes les plus avancées restent complexes à aborder. Nous espérons que cet exposé situera ces méthodes dans le contexte des méthodes de prévision. Il faut insister que si les outils ont été développés il y a plus de trente ans, beaucoup de progrès ont été réalisés (et le sont encore de nos jours) et que les apports pluridisciplinaires de tous ceux, chercheurs opérationnels, statisticiens, économètres et ingénieurs, qui participent au développement de la méthodologie, ont été intégrés.

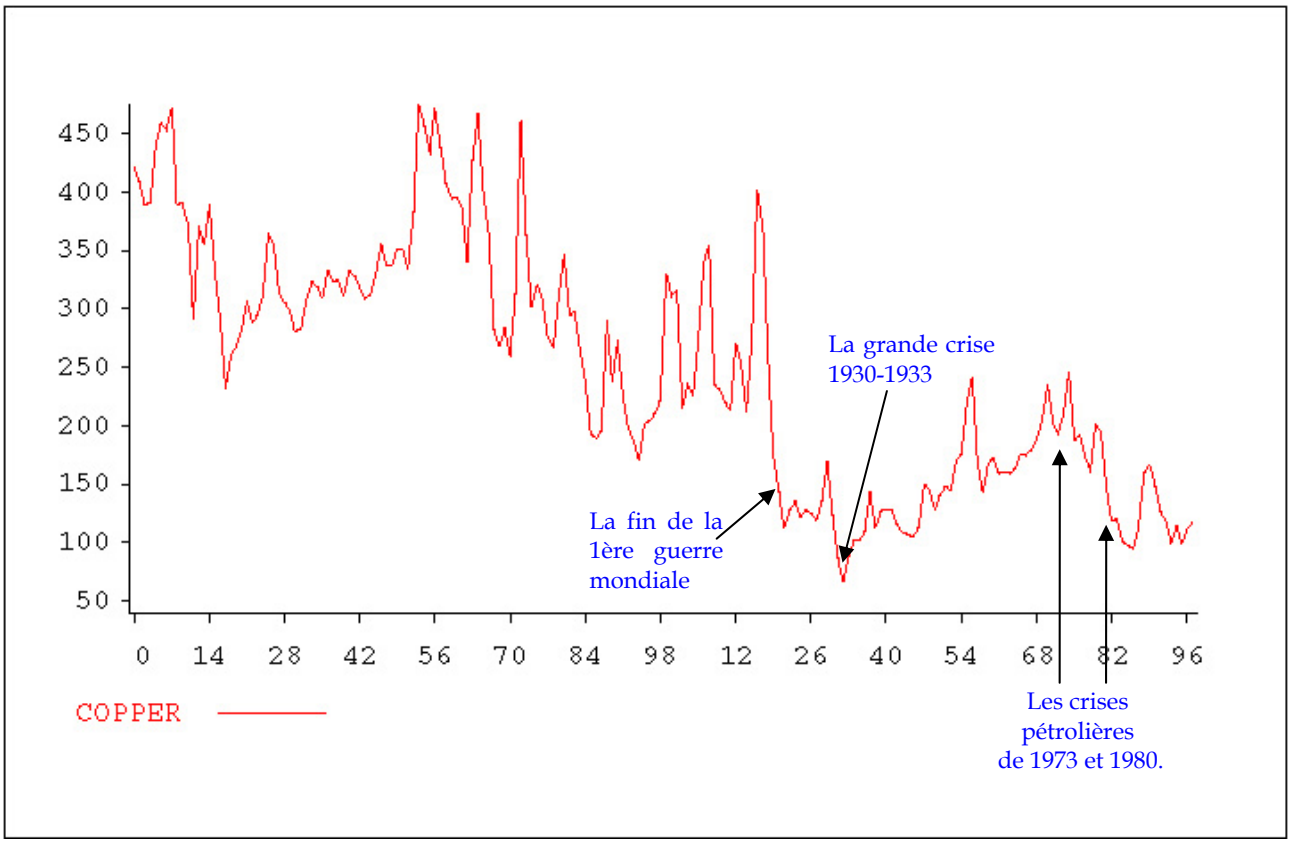
1.3 Quelques exemples

Nous comptons illustrer les méthodes à l'aide de quelques exemples assez variés. La discussion de ces exemples permettra de montrer l'application des différentes méthodes.

Exemple 1. CU, les prix du cuivre (1800-1997)

Les données (Martino, 1983) sont annuelles. Le graphe annoté est présenté dans la figure 1. On peut localiser la fin de la 1ère guerre mondiale, la grande crise des années '30, les crises pétrolières de 1973 et 1980.

Figure 1. Les données du prix du cuivre



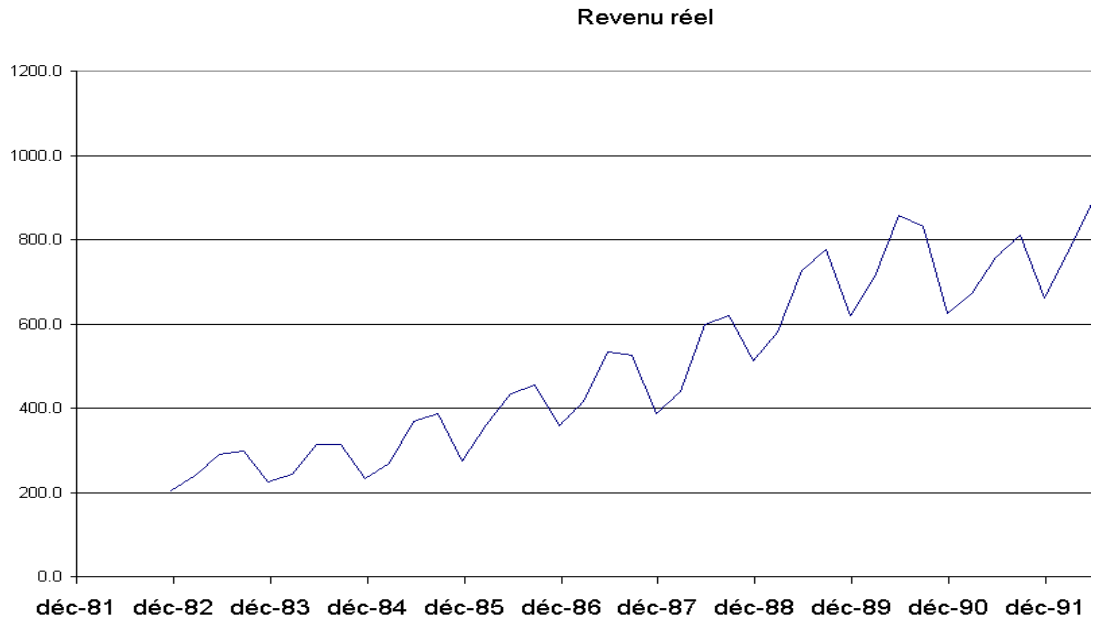
Exemple 2. PV15MIN, les retards au décollage à un aéroport (janvier 1994 - mai 1998)

Les données et le contexte proviennent d'un journal belge (La Dernière Heure, 28 juillet 1998, page 6). Il s'agit des pourcentages de vols enregistrant plus de 15 minutes de retard au décollage (appendice A, tableau A.4, partie B1 Original Series) à l'aéroport de Bruxelles-National. Elles sont illustrées dans la figure 18. Ces retards ont eu tendance à s'accentuer, en partie à cause des faibles temps de rotation entre deux vols ou pour des problèmes techniques, mais la plupart sont imputables à une congestion du trafic aérien, principalement à cause des capacités insuffisantes des aéroports pour accueillir le trafic estival.

Exemple 3. DISNEY, les revenus de Walt Disney Company (1982-1991)

Il s'agit de revenus trimestriels, en millions de dollars, d'après les rapports de la société [basé sur Levin et Rubin, 1998, pp. 910-913]. Les données sont présentées dans la figure 2.

Figure 2. Les données des revenus de Walt Disney Company



Exemple 4. ICECREAM, les ventes d'un glacier aux Etats-Unis

Les données (Kadiyala, 1970) sont relatives à 30 périodes de 4 semaines commençant du 18 mars 1984 au 15 juin 1986. Les variables utilisées sont:

IC	les ventes de crème glacée par habitant (en pintes)
PRICE	les prix de la crème glacée (en dollars par pinte)
INCOME	le revenu moyen hebdomadaire des ménages (en dollars)
TEMP	la température extérieure moyenne (en degrés Fahrenheit)
LAGTEMP	la variable TEMP retardée d'une période
DATE	une variable de temps prenant les valeurs entières de 1 à 30.

Les données sont présentées dans la figure 29. Après quelques manipulations de date dans un tableur, nous avons pu repérer le numéro de la période de 4 semaines dans laquelle tombent les principaux jours fériés aux Etats-Unis :

Independance Day (4 juillet)	4, 17, 30
Labour Day (1er lundi de septembre)	7, 20
Thanksgiving (4e jeudi de novembre)	10, 22
Memorial Day (dernier lundi de mai)	3, 16, 29.

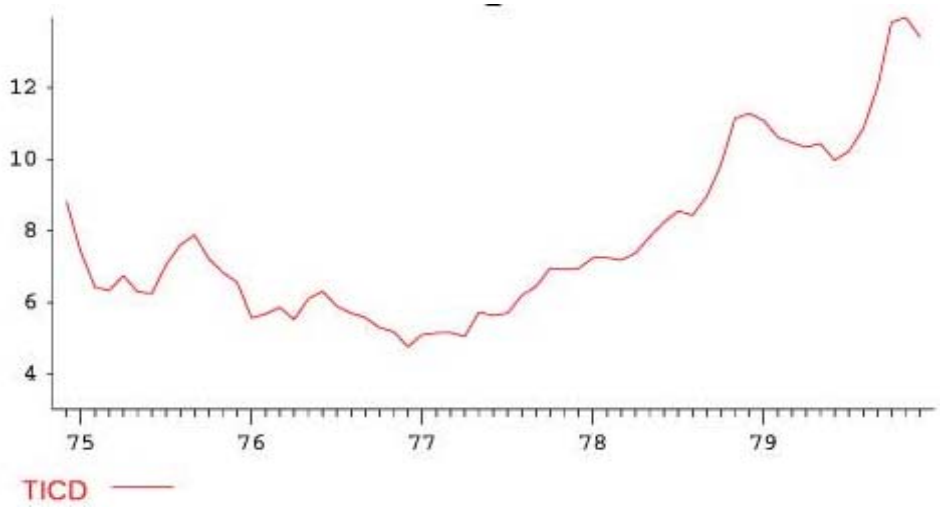
Exemple 5. PIB, le produit intérieur brut de l'Italie

La série montrée dans la figure 42 est trimestrielle et couvre la période du 1er trimestre 1980 au 4e trimestre 1991, mais nous réserverons les deux dernières années de données à la comparaison avec les prévisions ex post.

Exemple 6. TICD, les taux d'intérêt des certificats de dépôt aux Etats-Unis

La série (figure 3) est mensuelle et sont relatives à la période entre avril 1975 et décembre 1979. Les paramètres seront estimés sur la période qui va jusqu'en décembre 1978. La série est utilisée dans Pindyck et Rubinfeld (1976) et les données sont tirées d'un exemple d'une version de démonstration du logiciel SORITEC Sampler.

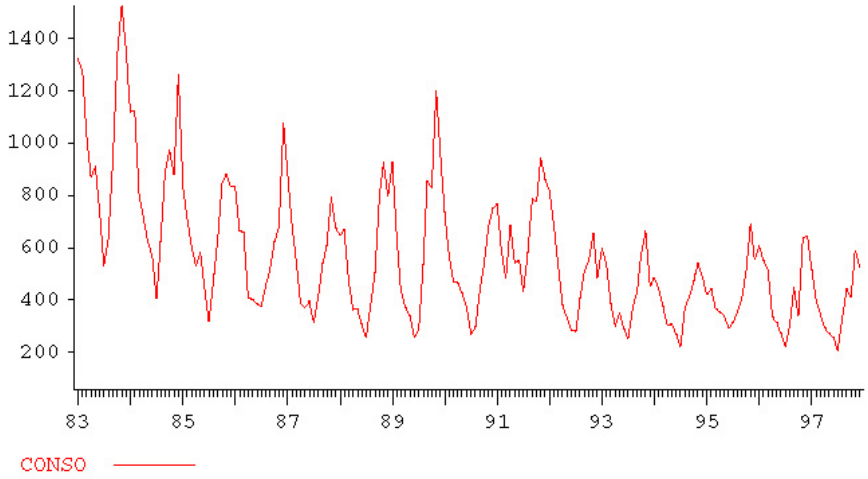
Figure 3. Les données des taux d'intérêt des certificats de dépôt



Exemple 7. CONSO, la consommation de fuel lourd en France

La figure 4 présente les données de janvier 1983 à juin 1999.

Figure 4. Les données de consommation de fuel lourd en France



2 Les méthodes

Notons y la variable étudiée mesurée (on pourrait l'appeler *VENTES* mais nous voulons faire plus court) selon un espacement régulier, $t = 1, 2, ..., T$, et $y_1, y_2, ..., y_T$, les données observées. Nous allons principalement traiter de la prévision mais l'analyse de données temporelles peut aussi se baser sur du lissage. Le lissage par moyennes mobiles est illustré dans les exemples 1 et 2 (paragraphes 3.1 et 3.2). Une moyenne mobile simple d'ordre 5, par exemple, est définie comme

$$\bar{y}_t = \frac{y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2}}{5}.$$

Pour les ordres k pairs, nous utiliserons aussi des moyennes mobiles centrées qui sont une moyenne mobile d'ordre 2 d'une moyenne mobile simple d'ordre k. Nous mentionnerons aussi la moyenne mobile de Spencer d'ordre 15 qui est une moyenne mobile pondérée avec des poids symétriques, dont certains sont négatifs :

$$\frac{1}{320}[-3, -6, -5, 3, 21, 46, 67, 74, 67, 46, 21, 3, -5, -6, -3].$$

2.1 Les méthodes de prévision

Le choix d'une méthode de prévision repose sur l'ensemble d'information, c'est-à-dire l'information disponible et que l'on veut exploiter. L'origine de prévision est T et l'horizon de prévision est h. On veut prévoir la valeur future inconnue y_{T+h}, notée $\hat{y}_T(h)$, mais aussi étudier la distribution de probabilité des *erreurs de prévision* $e_T(h) = y_{T+h} - \hat{y}_T(h)$. Nous examinerons surtout les *erreurs de prévision d'horizon* 1 : $e_{T+1} = y_{T+1} - \hat{y}_T(1)$, pour des raisons qui apparaîtront ultérieurement.

Les *méthodes extrapolatives* telles que les courbes de croissance, les méthodes de décomposition saisonnière, le lissage exponentiel et la méthode de Box et Jenkins pour les modèles ARMA (autorégressif-moyenne mobile) utilisent l'information présente et passée : y_T, y_{T-1}, ... pour prévoir y_{T+1}. Les *méthodes explicatives* comme la régression linéaire simple ou multiple emploient le présent et le passé de la variable étudiée mais aussi d'une ou plusieurs variables explicatives : x_T, x_{T-1}, ..., y_T, y_{T-1}, ... Les *méthodes systémiques et économétriques* étudient les relations entre les variables dans les deux sens : si y_T est fonction de x_T, alors x_T peut aussi être fonction de y_T. De l'information *qualitative* peut être employée. On peut la représenter à l'aide de *variables binaires*. Par exemple, l'effet de la première crise pétrolière sera exprimé par une variable valant 0 jusqu'en 1973 et 1 après.

Un autre aspect fondamental qu'il faudrait traiter est celui de la fonction de coût. Le coût d'une erreur de prévision n'est pas nécessairement symétrique, ce que le critère des moindres carrés ou "mean square error" (MSE) (bien pratique en régression multiple) implique. Son rôle tend à diminuer. Les praticiens jugent souvent de la performance d'une méthode par le critère de l'erreur absolue moyenne en pourcentage ("mean absolute percentage error") ou MAPE.

2.2 Les méthodes déterministes

A défaut de variable explicative, les modèles déterministes où la variable est une fonction déterminée du temps mais dépendant de paramètres à estimer, ne sont plus très appréciés. La tendance linéaire ou exponentielle relève du passé. On note bien le recours aux courbes de croissance (Gompertz, logistique), qui sont plus utiles pour la prévision à moyen ou à long terme mais sont aussi employées en marketing, et aussi dans la prévision de ventes par correspondance (prévision du plafond des ventes cumulées d'un article). L'estimation des paramètres est effectuée par régression non linéaire ou par des techniques ad hoc. Un autre domaine où les modèles déterministes sont encore en vigueur est celui de la décomposition saisonnière. On suppose que la série est la juxtaposition, par addition ou par multiplication, des composantes de *tendance* (T), l'allure générale du phénomène, de *cycle conjoncturel* (C), alternance de périodes d'expansion et de contraction, des *variations saisonnières* (S), liées aux rythme des saisons météorologiques soit directement (agriculture, énergie, etc.), soit indirectement (fêtes, vacances, soldes, etc.) et des *variations accidentelles* ou *erreurs* (E), de nature aléatoire. La plupart des séries mensuelles ou trimestrielles de l'entreprise sont de nature saisonnière. La tendance récente à l'annualisation du travail procure de nouveaux débouchés à cette approche. A part son intérêt en prévision, la décomposition saisonnière permet aussi de déterminer les *données corrigées des variations saisonnières*. Nous illustrerons quatre méthodes de décomposition saisonnière dans l'exercice 2 (paragraphe 3.2). Nous emploierons des données corrigées des variations saisonnières dans les exercices 3 et 7 (paragraphes 3.3 et 3.7).

2.3 Les méthodes non déterministes

Les méthodes qui ne reposent pas sur un modèle déterministe sont la prévision par moyenne mobile et la prévision par lissage exponentiel. La prévision par *moyenne mobile d'ordre k* consiste à

calculer la moyenne des k dernières données, comme indiqué plus haut, et à l'employer comme prévision. On applique la méthode de manière glissante. Soyons clairs, cette méthode est une très mauvaise méthode de prévision, à une exception près, celle des prix liés au marché et à condition de choisir $k = 1$, c'est-à-dire employer la méthode de *prévision naïve* où le prix de clôture d'aujourd'hui est la prévision pour le cours de demain.

Les méthodes de *lissage exponentiel* datent du début des années soixante. La simplicité de ces méthodes et leur faible coût en dépit des performances qu'elles procurent justifient amplement leur utilisation. Nous traitons ici du *lissage exponentiel simple* ("simple exponential smoothing"). L'idée du lissage exponentiel simple est simpliste mais efficace. Il s'agit de calculer la prévision pour le temps $T + 1$, comme moyenne pondérée de

- la dernière observation disponible
- la dernière prévision calculée,

au moyen de la formule

$$F_{t+1} = \alpha y_t + (1-\alpha) F_t \qquad (1)$$

et ceci par récurrence sur t, en employant une prévision initiale F_0. Il faut choisir la *constante de lissage* α, comprise entre 0 et 1, le cas $\alpha = 1$ correspondant à la méthode de *prévision naïve*. On procède par évaluation d'un critère (MSE ou MAPE). On montre que $\hat{y}_T(h) = F_{t+1}$. Une interprétation intéressante de la définition est la forme de correction par l'erreur

$$F_{t+1} = F_t + \alpha e_t .$$

Le nombre réduit d'opérations explique l'emploi du lissage exponentiel dans beaucoup de systèmes de gestion de stocks. Finalement, le plus difficile est d'expliquer le nom de la méthode. Ses principaux avantages sont qu'elle est simple à comprendre et à expliquer, simple à mettre en œuvre, d'un coût total très bas et que son intérêt pratique a été démontré (M-Competition; Makridakis *et al.*, 1984).

Pour des séries avec tendance et/ou saisonnalité, il y a plusieurs autres méthodes, une dizaine en tout, parmi lesquelles la plus connue est celles de Brown et de Holt-Winters. Une autre approche pour traiter la saisonnalité consiste à appliquer le lissage exponentiel simple ou de Brown sur la série corrigée des variations saisonnières pour calculer des prévisions désaisonnalisées et à restituer ensuite la saisonnalité. Ceci sera illustré dans les exemples 2, 3 et 7 (paragraphes 3.2, 3.3 et 3.7).

2.4 Les méthodes économétriques

Avant d'aborder les modèles de séries chronologiques, considérons la *régression linéaire multiple*. Cette méthode statistique est très importante en prévision parce qu'elle permet d'introduire les facteurs extérieurs qui influencent la valeur future de la variable étudiée. Pour les prévisions de ventes, ce sont les prix et les promotions de l'entreprise qui viennent d'abord en tête comme variables explicatives. Idéalement, les prix et les promotions des principaux concurrents devraient être considérés. S'ils sont probablement disponibles pour le passé, ils ne le sont sûrement pas pour le futur. Or, pour être utile en matière de prévision, un modèle de régression doit être tel que les variables explicatives soient connues ou prévisibles. Ce n'est évidemment pas le cas pour ce qui concerne les concurrents.

Le modèle de régression linéaire multiple est décrit par l'équation

$$y = b_1 x_1 + b_2 x_2 + \ldots + b_k x_k + e.$$

L'interprétation du coefficient de régression b_j est délicate. Il mesure l'augmentation de la variable dépendante y si la variable explicative x_j augmente d'une unité, mais toutes choses étant égales par ailleurs. On emploie presque toujours une constante dans le modèle, c'est-à-dire qu'une des variables vaut 1 identiquement, par exemple $x_1 = 1$. Il faut estimer ces coefficients de régression. On suppose disposer de T données relatives aux $k + 1$ variables. L'application de la méthode des

moindres carrés ne pose plus de problème sauf parfois d'ordre numérique, compte tenu du nombre fini de chiffres utilisés dans les calculs et d'effets de quasi-colinéarité (le fait que deux ou plusieurs des variables explicatives aient entre elles une relation presque linéaire). La qualité globale du modèle est mesurée par le coefficient de détermination noté R^2 qui représente la proportion de la variance de *y* expliquée par l'ensemble des x_j.

Les aspects statistiques deviennent ici plus importants. Pour étudier la sensibilité des coefficients à des variations de l'échantillon de données, il faut imposer des suppositions relativement abstraites dont il est parfois difficile d'apprécier l'importance :

1) que les variables explicatives sont mesurées *sans erreur* et sans *relation linéaire* entre elles ;
2) que la relation entre la variable dépendante et les variables explicatives soit bien spécifiée, donc *linéaire* vis-à-vis des paramètres ;
3) que la distribution de probabilité des erreurs soit *normale centrée* et d'*écart-type constant* (supposition de *normalité* et d'*homoscédasticité*), ce qui exclut, en principe, la présence de *données aberrantes* et d'effet de taille quand les données sont relatives à des pays, à des entreprises, etc. ;
4) que les erreurs soient mutuellement *indépendantes*, donc assimilables à un échantillon aléatoire simple.

On ne peut pas savoir si ces suppositions sont remplies mais on peut souvent détecter les problèmes les plus graves en examinant des graphiques, notamment le graphique des résidus e_t ou erreurs estimées en fonction des valeurs ajustées à l'aide du modèle (ou prévisions). On peut ainsi détecter

1) une mauvaise spécification du modèle ;
2) la présence de données aberrantes ;
3) l'hétéroscédasticité ;
4) l'autocorrélation des erreurs.

L'*autocorrélation des erreurs* n'a de sens que pour données chronologiques ou assimilées. C'est la supposition qui peut le plus invalider les résultats parce qu'une *autocorrélation positive* (le fait que les erreurs successives aient tendance à être proches les unes des autres au lieu de varier de manière aléatoire) rend les coefficients de régression statistiquement significatifs, de manière artificielle et à un degré difficilement imaginable. Ainsi, en ne prenant pas garde à de l'autocorrélation positive, un chercheur universitaire a pu croire, le temps d'un exposé de séminaire, qu'il avait démontré que les cours de clôture hebdomadaire de la Bourse de Bruxelles sont prévisibles à 12 semaines d'intervalle. Le *test de Durbin-Watson* permet (quand on l'emploie!) de détecter les situations les plus graves. En revanche, la normalité des erreurs semble une condition très académique.

Du point de vue pratique, il y a essentiellement deux types d'utilisateurs de la régression multiple: ceux qui ont trop peu de variables et ceux qui en ont trop. Faut-il utiliser toutes les variables explicatives disponibles ? Dans le cas contraire, comment choisir les variables ? Faut-il employer toutes les observations disponibles ? Dans le cas contraire, de quel droit rejeter certaines données ? Les réponses à ces questions ne sont pas évidentes.

Pour la sélection des variables explicatives, il existe bien des *méthodes pas à pas* ("*stepwise*") mais elles présentent des dangers: elles ne font parfois pas entrer dans le modèle les variables explicatives qu'il faudrait et elles font parfois entrer des variables non pertinentes mais néanmoins statistiquement corrélées, par un artefact du grand nombre de *tests statistiques* effectués. Il ne faut pas oublier que quand on effectue *K* tests statistiques sur des données indépendantes au niveau de probabilité de 5%, la probabilité de rejeter au moins une des *K* hypothèses vaut $1 - (0{,}95)^K$. Pour $K = 20$, on trouve non pas 5% mais 64%. On recommande plutôt de choisir les variables explicatives en tenant compte de motivations théoriques ou simplement du bon sens, de préférence avant de regarder les données, sans trop s'occuper des *probabilités de signification* ('*P*-values').

Pour la sélection des données, le but est d'avoir un modèle qui convienne dans la plupart des cas. On peut parfois éliminer des observations aberrantes (pas parce qu'elles n'entrent pas bien dans le modèle mais sur des considérations objectives, connaissance d'un incident ou d'une particularité). On peut aussi faire intervenir des informations qualitatives via des variables binaires ou indicatrices afin de refléter l'existence de plusieurs relations différentes tout en procédant à une estimation globale. Dans le cas d'existence de r catégories il faudra utiliser $r - 1$ variables binaires indicatrices quand le modèle comporte une constante (et pas une seule ni r).

Nous voulons employer la régression multiple sur des séries chronologiques. Or une des suppositions de base, celle de l'indépendance, n'a aucune raison d'être vérifiée. Le *test de Durbin-Watson* permet d'éviter les situations les plus dangereuses. Les *variations saisonnières* peuvent être représentées par des *variables binaires indicatrices* ("dummy variables"), par exemple JAN = 1 pour tous les mois de janvier, et JAN = 0 pour les autres mois. On peut être amené à employer des variables explicatives sous forme *retardée*, et même employer la variable dépendante retardée comme variable explicative (ce qui est en dehors des suppositions ci-dessus puisque la variable dépendante est pratiquement toujours affectée d'erreur). On recourt aussi aux différences $\nabla x_t = x_t - x_{t-1}$ et aux taux d'accroissement. En outre, le coefficient de détermination est généralement très élevé à cause de la tendance commune sans que cela reflète la valeur explicative du modèle. En conséquence, il faut prendre des précautions pour éviter les erreurs de spécification du modèle. Nous illustrerons la régression multiple et la plupart des problèmes qu'elle pose dans l'exercice 4 (paragraphe 3.4).

Enfin, il faut déterminer un intervalle de prévision. Cet intervalle basé sur l'hypothèse de normalité, est d'autant plus large qu'on s'éloigne des valeurs moyennes des variables explicatives. Le grand problème de la régression multiple pour la prévision, c'est qu'il faut aussi prévoir les variables explicatives. Par exemple, Ashley (1988, p. 374) conclut que "This means (for example) that it is a waste of time to include contemporaneous national level variables in a corporate or regional forecasting model if forecasts well beyond a quarter of two ahead are required.".

2.5 Erreurs de prévision et autocorrélation

Avant de considérer les modèles de séries chronologiques, il n'est pas inutile de justifier leur emploi par l'examen des *erreurs de prévision d'horizon* 1, ou *résidus*, fournies par d'autres méthodes. On peut imaginer deux manières de juger des qualités d'une prévision : soit appliquer la méthode sur données récentes et comparer les résultats avec ceux d'autres méthodes, soit appliquer la méthode sur toute la série et juger si les résidus ne contiennent plus d'information pertinente. Plus précisément, on peut avoir deux exigences pour les erreurs de prévision d'horizon 1 ("1-*step ahead forecast error*"):
(1) elles varient autour de 0 (sans quoi les prévisions seraient systématiquement biaisées) ;
(2) l'erreur au temps t ne contient pas d'information susceptible de prévoir l'erreur au temps $t + 1$ et aux temps suivants.

La première question peut être étudiée en effectuant un test statistique sur la moyenne, en supposant l'indépendance des erreurs. La seconde question consiste à éprouver précisément l'hypothèse d'indépendance, ce qui n'est concevable que dans un contexte plus large que celui du champ de la statistique classique basé sur des échantillons aléatoires. C'est la statistique des *processus aléatoires* (ou stochastiques) ("stochastic process", "random process"). Un processus constitué de variables aléatoires indépendantes de même distribution est un processus de type "*bruit blanc*" ("white noise"). On éprouve donc le caractère aléatoire des résidus au moyen d'un test de bruit blanc ("*test for randomness*"). La statistique de test la plus simple est l'autocorrélation des résidus, la corrélation entre la série $\{e_t\}$ et la série $\{e_{t-k}\}$, pour k fixé (1, 2, ...). Il faut définir un concept de population pour formuler le test d'hypothèse. Mais la classe est trop vaste et il faut

supposer la *stationnarité* du processus aléatoire pour pouvoir définir une *autocorrélation de retard k* du processus. Un processus $\{Y_t\}$ est stationnaire du second ordre si et seulement si les Y_t ont les mêmes espérances mathématiques ou moyennes $E(Y_t)$, mettons m, et des variances identiques, notées γ_0 dans la suite, et en outre que l'*autocovariance* $\gamma_k = E[(Y_t - m)(Y_{t-k} - m)]$ ne dépend pas de t. On peut alors effectuer un test individuel (pour un retard k) ou un test global. L'autocorrélation échantillon de retard k, r_k, est *significative* (au niveau de 5%) si $T^{1/2} r_k < -1{,}96$ ou si $T^{1/2} r_k > 1{,}96$. Nous illustrerons l'emploi des autocorrélations des séries résiduelles dans l'exercice 3 (paragraphe 3.3).

Comme la plupart des séries économiques et financières ne peuvent pas être considérées comme générées par un processus stationnaire, on doit d'abord les rendre stationnaires avant d'appliquer le test de bruit blanc. A cette fin, on emploie souvent la différence $\nabla y_t = y_t - y_{t-1}$, parfois une différence saisonnière $\nabla_{12} y_t = y_t - y_{t-12}$, dans le cas de données mensuelles ou $\nabla_4 y_t = y_t - y_{t-4}$, pour des données trimestrielles, ou une combinaison de plusieurs différences. Par exemple, pour une série de cours boursiers (généralement en logarithmes), la méthode de prévision naïve est justifiée si la série provient d'un processus de *marche au hasard* ("random walk process"), c'est-à-dire si la série $\{Y_t = \nabla y_t\}$ a un comportement de processus bruit blanc. Quand on calcule les autocorrélations pour des retards allant de 1 à 24, par exemple, on doit s'attendre en moyenne à 5% de rejets (même si l'hypothèse est vraie) soit 1,2 rejets, en moyenne. Il est recommandé de ne considérer que les retards suspects a priori (1, 2, et 4 ou 12). On emploie aussi les tests globaux ("portmanteau tests") qui portent sur plusieurs retards, comme le test de Ljung-Box. Pour conclure ceci, signalons qu'il existe des tests plus résistants aux valeurs aberrantes et aussi des tests de stationnarité (tests de racine unité). Nous illustrerons l'emploi des différences ordinaires et saisonnières dans les exercices 4 à 6 (paragraphes 3.4 à 3.6) et des transformations dans l'exercice 7 (paragraphe 3.7).

2.6 Les modèles ARIMA

Les modèles de séries chronologiques se basent sur les processus ARIMA qui sont définis un peu plus loin. Avant cela, on peut introduire le concept de forme ou de représentation ARIMA d'une méthode de prévision. Il s'agit d'une relation entre les observations éventuellement retardées et les erreurs de prévision d'horizon 1, éventuellement retardées. Par exemple, on peut montrer très simplement qu'il existe une forme ARIMA du lissage exponentiel simple (1)

$$y_t - y_{t-1} = e_t - (1-\alpha)e_{t-1}. \qquad (2)$$

La forme ARIMA de la prévision par moyenne mobile d'ordre 4 est la suivante:

$$y_t - \frac{1}{4}y_{t-1} - \frac{1}{4}y_{t-2} - \frac{1}{4}y_{t-3} - \frac{1}{4}y_{t-4} = e_t,$$

ou, si l'on note B l'opérateur de retard, tel que $By_t = y_{t-1}$:

$$(1 + \frac{3}{4}B + \frac{2}{4}B^2 + \frac{1}{4}B^3)\nabla\ y_t = e_t,$$

où on représente $\nabla = 1 - B$, l'opérateur de différence ordinaire, tel que $\nabla y_t = y_t - y_{t-1}$.

Un processus ARIMA est un processus aléatoire qui vérifie une telle équation où les e_t sont des variables aléatoires constituant un processus bruit blanc. Plus précisément, un processus ARIMA(p, d, q) vérifie une équation aux différences stochastique de la forme

$$(1 - \phi_1 B - ... - \phi^p B^p)\nabla^d y_t = (1 - \theta_1 B - ... - \theta^q B^q)e_t,$$

où $\phi_p(B) = (1 - \phi_1 B - ... - \phi^q B^q)$ est le polynôme autorégressif, de degré p en B, alors que $\theta_q(B) = (1 - \theta_1 B - ... - \theta^q B^q)$ est le polynôme moyenne mobile, de degré q en B et ∇^d est l'opérateur de différence appliqué d fois. Par exemple, si $p = 2$ et $q = 1$, les deux polynômes

s'écrivent $\phi_2(B) = 1 - \phi_2 B - \phi_2 B^2$ et $\theta_1(B) = 1 - \theta_1 B$. L'avantage d'un processus ARIMA est qu'il est très facile à prévoir. Par exemple, pour le modèle (2) :

$$y_{T+1} = y_T + e_{T+1} - (1-\alpha)e_T$$
$$y_{T+2} = y_{T+1} + e_{T+2} - (1-\alpha)e_{T+1}$$
$$y_{T+3} = y_{T+2} + e_{T+3} - (1-\alpha)e_{T+2}$$

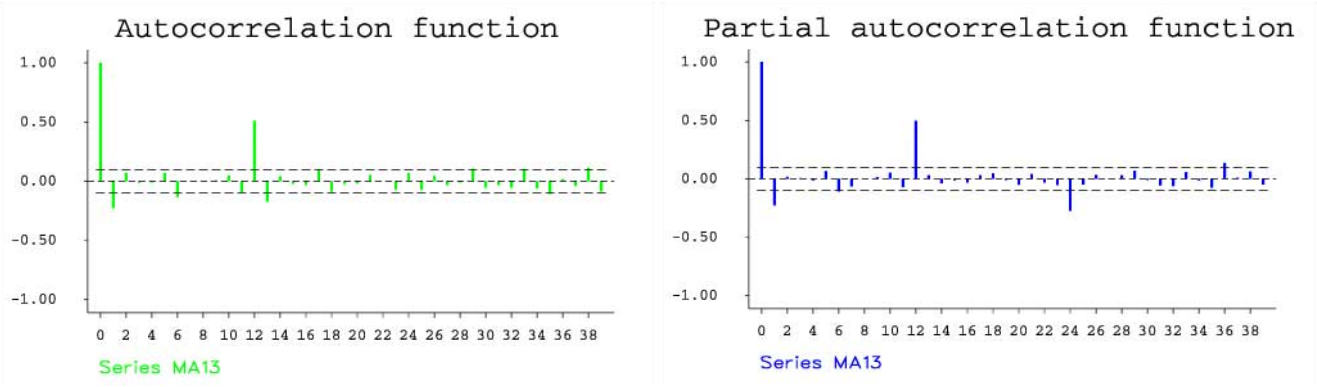
d'où on déduit successivement les prévisions d'horizon 1, 2, 3, par récurrence en remplaçant les valeurs inconnues des erreurs e_{T+1}, e_{T+2}, e_{T+3} par zéro. Pour pouvoir modéliser une série donnée par un modèle ARIMA, il faut surtout spécifier un bon modèle, ensuite estimer les paramètres, comme ici (1 – α), et valider le modèle (en examinant le comportement alétoire des résidus). C'est en quoi consiste la méthode de Box et Jenkins. En plus des processus ARIMA, il s'avère utile de définir les processus SARIMA ou ARIMA saisonniers. Outre la différence ordinaire, ils emploient la différence saisonnière définie au paragraphe 2.5, et les polynômes autorégressifs et/ou moyenne mobile sont factorisés en un produit d'un polynôme ordinaire, en B, et d'un polynôme saisonnier, en B^s. L'exemple 5 (paragraphe 3.5) est presque le plus simple qu'on puisse imaginer. L'exemple 7 (paragraphe 3.7) est plus complexe.

Auparavant, il faut étudier les processus AR, MA, ARMA et enfin ARIMA. Par exemple, on montre qu'un processus MA(q) a une fonction d'autocorrélation *tronquée* au-delà du retard q, c'est-à-dire que ces autocorrélations du processus sont nulles. Par exemple, les autocorrélations du processus MA(1) d'équation

$$y_t = e_t - \theta e_{t-1} \qquad (3)$$

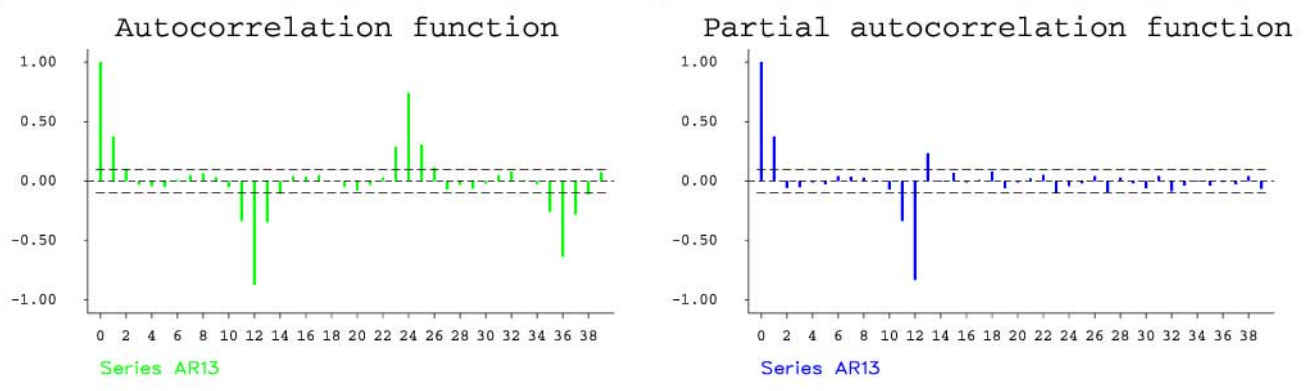
sont nulles pour un retard supérieur strictement à 1 parce que la variance vaut $(1 + \theta^2)\sigma^2$ et l'autocovariance de retard 1 vaut $-\theta\sigma^2$, donc l'autocorrélation de retard 1 est donnée par $\rho_1 = -\theta/(1 + \theta^2)$, celles de retard 2, 3, … étant nulles. Bien sûr les autocorrélations d'une série de longueur finie ne seront pas nulles, tout au plus non significatives (et encore !). On observe le corrélogramme d'une telle série dans la figure 5.

Figures 5 et 6. Fonctions d'autocorrélation et d'autocorrélation partielle d'une série de longueur 400 générée par le processus $y_t = (1 - 0{,}2\,B)(1 + 0{,}9\,B^{12})\,e_t$.



La fonction d'autocorrélation d'un processus AR est trop complexe pour permettre une spécification correcte ainsi que le montre la figure 7. On recourt alors à la fonction d'autocorrélation partielle qui est tronquée au delà du retard p pour un processus AR(p). Le corrélogramme partiel de la figure 8 fait bien apparaître une troncation. Celui de la figure 6 ne nous apprend rien.

Figures 7 et 8. Fonctions d'autocorrélation et d'autocorrélation partielle d'une série de longueur 400 générée par le processus $(1 - 0{,}2\,B)(1 + 0{,}9\,B^{12})\,y_t = e_t$.



Bien souvent, les processus ARMA s'avèrent nécessaires pour obtenir une représentation avec un nombre de paramètres réduit (pour une série de 100 données, on considère parfois qu'estimer plus de trois paramètres est excessif). Les séries non stationnaires ne sont pas représentables par un processus AR ou ARMA stationnaire (ce qui entraîne que les racines du polynôme autorégressif sont supérieures à 1 en module) mais bien éventuellement par un processus ARIMA. Pour être utilisable pour la prévision, un processus MA ou ARMA doit être inversible (ce qui entraîne que les racines du polynôme moyenne mobile sont supérieures à 1 en module). Les séries avec une saisonnalité nécessitent le recours à un ou deux polynômes saisonniers en plus d'une différence saisonnière.

2.7 La méthode de Box et Jenkins

La *méthode de Box et Jenkins* (1976) est donc l'application de la méthode scientifique à la modélisation de séries chronologiques. Pour l'appliquer, on peut effectuer les étapes suivantes, quitte à revenir en arrière si le résultat n'est pas satisfaisant :
(1) la familiarisation avec les données;
(2) l'analyse préliminaire;
(3) la spécification du modèle (ou identification);
(4) l'estimation des paramètres;
(5) l'étude de l'adéquation du modèle (ou validation);
(6) la prévision;
(7) l'interprétation des résultats.
La classe de modèles est celle des processus ARIMA mais elle peut être étendue à un modèle de régression avec erreurs ARIMA (y compris l'emploi de variables binaires, ce qui s'appelle alors *analyse d'intervention*), et à certaines classes de modèles non linéaires.

(1) La familiarisation avec les données consiste à s'informer sur le domaine d'application, les théories existantes, les objectifs poursuivis, la qualité des données (précision, exactitude), la périodicité inhérente au phénomène, l'homogénéité dans le temps, les événements qui ont pu influencer la série. Elle comporte un examen graphiques des données visant à repérer les changements de structure dans la série, les erreurs grossières, les conséquences d'interventions (changements législatifs ou économiques, accidents majeurs, grèves, etc.).

(2) L'analyse préliminaire commence par l'exercice d'options : abandonner une partie des données au début de la série, corriger les données aberrantes, suppléer les données manquantes, transformer les données (logarithmes, inverse, racine carrée, ...), changer de variable (division par une autre série). On essaie ensuite de rendre la série stationnaire en s'aidant de graphiques, notamment.

(3) Spécification du modèle (ou identification). Comme indiqué ci-dessus, elle se base sur la forme des autocorrélations et autocorrélations partielles ce qui conduit à un ou plusieurs modèles ARMA, mais généralement plusieurs itérations des étapes (3) à (5) sont nécessaires. On tient compte des éléments les plus marqués. On estime les paramètres du modèle et on recommence avec les résidus (voir plus loin).

(4) Estimation des paramètres. Les paramètres sont les coefficients des polynômes AR et MA. On minimise le critère MSE (méthode des moindres carrés non linéaires) ou, mieux, on maximise la pseudo-fonction de vraisemblance exacte (celle d'un processus normal). On recourt à cette fin à des *procédures numériques itératives*.

(5) L'étude de l'adéquation du modèle (ou validation) consiste à vérifier si l'optimisation non linéaire a abouti et si le modèle est correct. Si le modèle n'est pas valable, il faut reprendre l'analyse à partir d'une des étapes précédentes, de préférence en exploitant l'information acquise.

(6) La prévision découle immédiatement du modèle retenu. On obtient aussi les variances des erreurs de prévision d'horizon 1, 2, ... Sous la supposition de normalité, on détermine la distributions des valeurs futures et les intervalles de prévision.

(7) L'interprétation des résultats. Elle n'est pas toujours simple mais n'est pas essentielle.

2.8 L'analyse spectrale

Les méthodes décrites ci-dessus concernent l'approche temporelle des séries chronologiques. Il est aussi possible de traiter l'approche fréquentielle ou spectrale. En termes mathématiques, il s'agit de la transformée de Fourier de l'approche temporelle. A priori, l'information doit être analogue mais il y a certaines applications où l'approche spectrale est instructive. Il en est ainsi dans le traitement de la saisonnalité, ou ceux où les phénomènes oscillatoires prédominent. Pour fixer les idées, supposons un processus stationnaire où le temps est exprimé en mois. Le spectre ou densité spectrale est la répartition de la dispersion, plus précisément de la variance, en fonction des fréquences :
- les basses fréquences (correspondant aux longues périodes) sont relatives au long terme, plusieurs années ;
- les hautes fréquences (correspondant aux courtes périodes) sont relatives au court terme, quelques mois au plus ;
- les moyennes fréquences (correspondant aux périodes moyennes) sont relatives au moyen terme.

Les fréquences sont exprimées comme des nombres entre 0 et 2. La période est l'inverse de la fréquence et varie donc entre l'infini (= 1/0) et 0,5 (= 1/2). Nous noterons f la fréquence. Certains auteurs expriment la fréquence en fréquence angulaire, mesurée en radians. Si on note ω (omega grec) la pulsation ou fréquence angulaire, on a $\omega = 2\pi f$, où $\pi = 3{,}1415...$. La fréquence angulaire varie entre 0 et π. La période est alors exprimée par $2\pi/\omega$.
Notons $S(f)$ la densité spectrale en la fréquence f , f dans [0 ;0,5]. La variance est la « somme » des $S(f)$, ce qu'on écrit

$$\sigma^2 = 2\int_0^{0,5} S(f)df \ .$$

(4)

Plus généralement, la densité spectrale est la transformée de Fourier de la fonction d'autocovariance :

$$\gamma_k = 2\int_0^{0,5} S(f)e^{2\pi ikf}df \ ,$$

ce qui donne en particulier (4) dans le cas où $k = 0$. Inversement, on peut écrire la transformée de Fourier inverse :

$$S(f) = \sum_{k=-\infty}^{\infty} e^{-2\pi i k f} \gamma_k ,$$

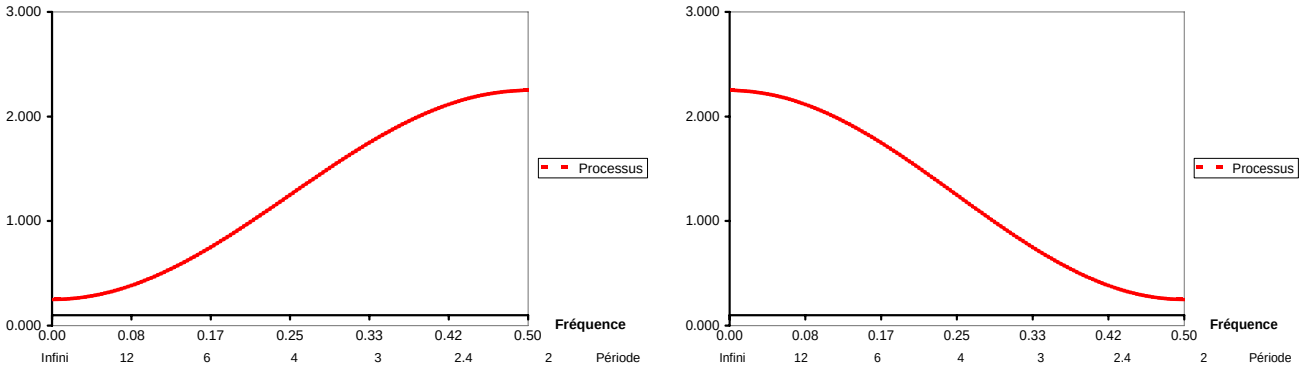
tenant compte que $\gamma_k = \gamma_{-k}$. Par exemple, considérons un processus MA(1) défini par (3). Or nous avons vu (paragraphe 2.6) que l'autocovariance de retard 0 vaut $(1 + \theta^2)\sigma^2$ et les autocovariances de retard +1 et –1 valent $-\theta\,\sigma^2$ de sorte qu'on a

$$\begin{aligned} S(f) &= (1+\theta^2)\sigma^2 - \sigma^2\theta(e^{2\pi i f} + e^{-2\pi i f}) \\ &= (1+\theta^2)\sigma^2 - 2\sigma^2\theta\cos(2\pi f) \\ &= \sigma^2[1+\theta^2 - 2\theta\cos(2\pi f)] \end{aligned}$$

Si $\theta < 0$, il y a de l'autocorrélation positive d'ordre 1, et le spectre est plus élevé en 0,5 qu'en 0 (figure 9). Si $\theta > 0$, il y a de l'autocorrélation négative d'ordre 1, ce qui se traduit par un spectre plus élevé en 0 qu'en 0,5 (figure 10). Si $\theta = 0$, le processus est de type bruit-blanc et son spectre est constant, c'est-à-dire toutes les fréquences ont même intensité. Une approche alternative équivalente consiste à définir $S(f)$ comme le carré du module du polynôme moyenne mobile évalué en $e^{2\pi i f}$:

$$S(f) = \sigma^2 \left|1 - \theta e^{2\pi i f}\right|^2 = \sigma^2 \left[1 + \theta^2 - 2\theta\cos(2\pi f)\right].$$

Figures 9 et 10. Spectres de processus MA(1) avec $\theta = 0{,}5$, à gauche, et $\theta = -0{,}5$, à droite.



Dans le cas d'un processus AR(p), on fait de même avec l'inverse du polynôme autorégressif, par exemple pour un processus AR(4), défini par l'équation

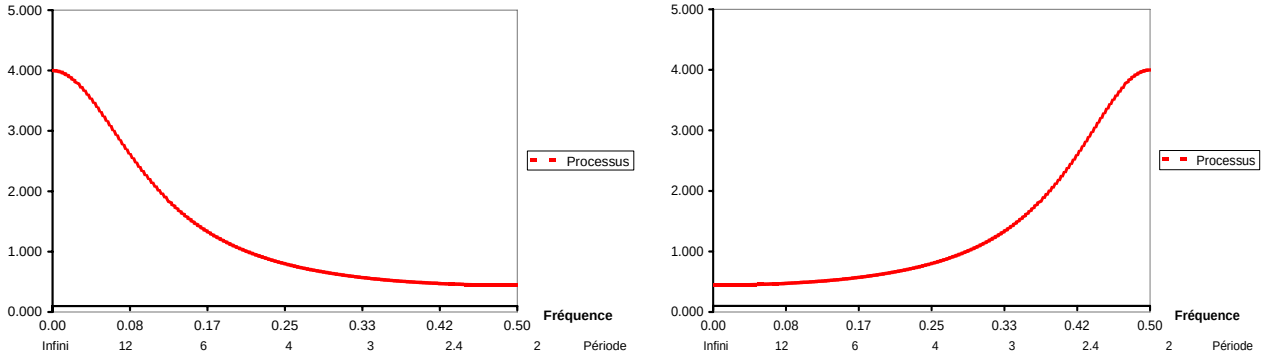
$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \phi_3 y_{t-3} - \phi_4 y_{t-4} = e_t, \qquad (5)$$

où la variance innovation vaut 1 :

$$S(f) = \frac{1}{\left|1 - \phi_1 e^{2\pi i f} - \phi_2 e^{2\pi i (2f)} - \phi_3 e^{2\pi i (3f)} - \phi_4 e^{2\pi i (4f)}\right|^2} . \qquad (6)$$

Les figures 11 et 12 montrent les cas de processus AR(1), comme (5) avec $\phi_1 = \phi$, $\phi_2 = \phi_3 = \phi_4 = 0$.

Figures 11 et 12. Spectres de processus AR(1) avec $\phi = 0{,}5$, à gauche, et $\phi = -0{,}5$, à droite.



Pour une série mensuelle, une saisonnalité dans la série va se manifester par un contenu fréquentiel important aux fréquences qui correspondent à la période 12 (car 1 an = 12 mois) mais aussi ses sous-multiples 12/2 = 6, 12/3 = 4, 12/4 = 3, 12/5 = 2,4 et 12/6 = 2. Considérons par exemple le processus suivant :

$$y_t - 0{,}8\, y_{t-12} = e_t - 0{,}6\, e_{t-1}, \qquad (7)$$

dont la variance des innovations vaut σ^2. Il s'agit bien d'un processus stationnaire dont les racines du polynôme autorégressif sont de module $(1/0{,}8)^{1/12} > 1$. Il est aussi inversible parce que la racine du polynôme moyenne mobile $(1/0{,}6) > 1$. Le spectre est défini par le rapport des carrés de modules suivants :

$$S(f) = \sigma^2 \frac{\left|1 - 0{,}6e^{2\pi i f}\right|^2}{\left|1 - 0{,}8e^{2\pi i (12 f)}\right|^2}.$$

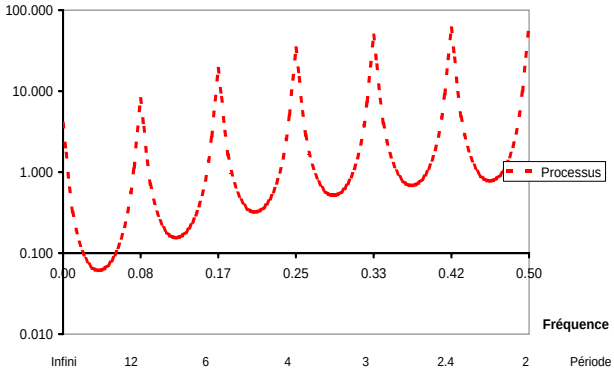
Le spectre comportera des pointes aux fréquences multiples de 1/12. On considère parfois les graphiques de spectres en échelle logarithmique. L'échelle logarithmique a pour effet de tasser les grandes valeurs. Pour le processus défini par (7), le spectre est illustré dans la figure 13.

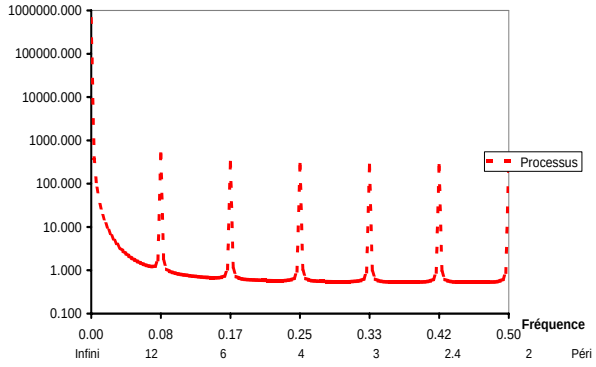
Jusqu'à présent nous avons défini la densité spectrale d'un processus stationnaire. Le cas de processus non stationnaires est plus délicat. On entend par là le cas de processus ARIMA ou SARIMA ayant des racines unités, à cause d'une différence ou d'une différence saisonnière. Dans ce cas, l'approche par les autocorrélations ne tient plus. Néanmoins, on peut définir le spectre généralisé par la seconde approche (voir Gourieroux et Montfort, 1990), étant entendu qu'il pourra être infini pour certaines fréquences. Par exemple le spectre généralisé du processus de promenade aléatoire, défini par l'équation $y_t - y_{t-1} = e_t$ et de variance innovation égale à σ^2 :

$$S(f) = \sigma^2 \frac{1}{\left|1 - e^{2\pi i f}\right|^2} = \frac{\sigma^2}{2 - 2\cos(2\pi f)}.$$

Comme cos(0) = 1, à la fréquence 0, le spectre est proportionnel à l'inverse de 2 – 2 = 0, donc est infini. Un exemple différent est traité dans la figure 14.

Figures 13 et 14. A gauche, spectre en logarithmes du processus stationnaire d'équation $(1-0.8B^{12})\,y_t = (1-0.6B)\,e_t$. A droite, approximation (obtenue en remplaçant les racines 1 par 0,99) du spectre en logarithmes du processus non stationnaire $\nabla\nabla_{12}y_t = (1-0.6B)(1-0.8B^{12})e_t$.





Pour estimer la densité spectrale à l'aide d'une série, plusieurs techniques sont envisageables. Il existe des méthodes employant la transformée de Fourier des autocorrélations estimées et en lissant les valeurs obtenues dans une fenêtre étroite autour d'une fréquence donnée. Nous illustrerons plutôt (voir l'exercice 6, paragraphe 3.6) l'approche autorégressive qui consiste à ajuster la série par un modèle AR d'ordre suffisamment grand et calculer le spectre du processus correspondant.

2.9 Conclusion

Les modèles de séries chronologiques tels qu'ils ont été décrits ici ne suffisent souvent pas pour la représentation et la prévision de toutes les données. Il faut envisager notamment d'inclure des variables explicatives (pour tenir compte de l'effet des prix) et des interventions (pour tenir compte des promotions et des incidents éventuels). De nombreuses généralisations de ces modèles sont néanmoins développées et permettent de réduire légèrement l'incertitude. Après discussion des exemples du paragraphe 3, il sera clair que les modèles de séries chronologiques peuvent rendre de grands services.

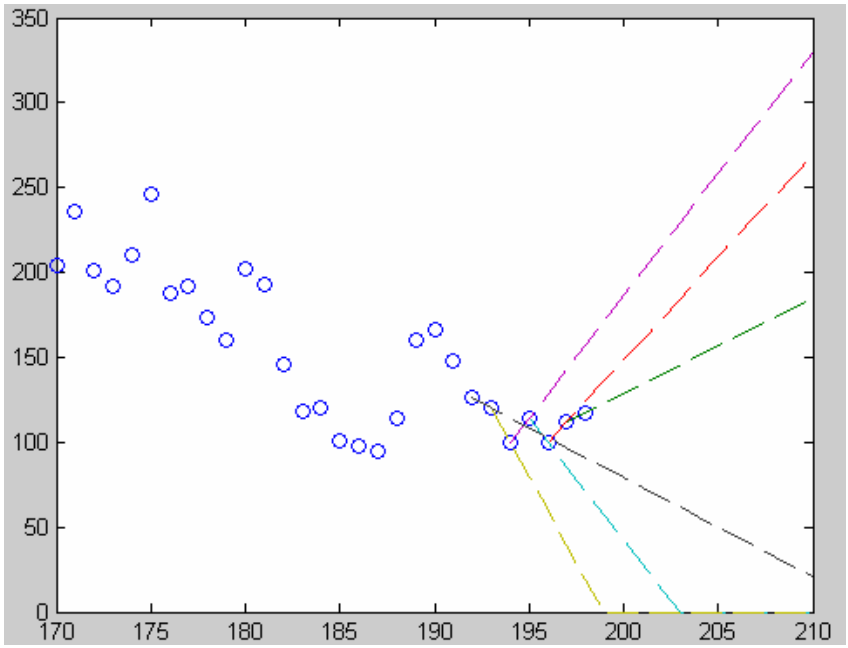
3 Exemples

3.1 Exemple 1

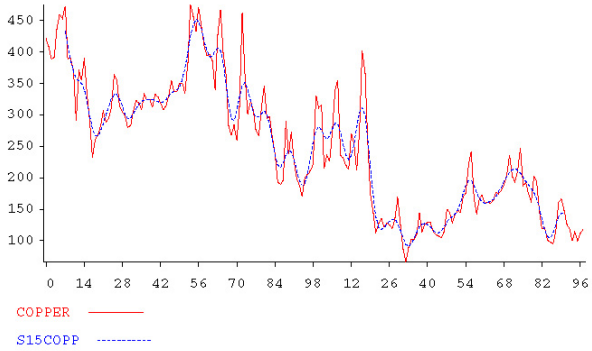
Il est parfois surprenant d'entendre comment des prévisions sont déterminées, en appliquant un taux de croissance uniforme, voire même en traçant une droite à main levée. A titre d'exemple, sur la série des prix du cuivre (1970-1997), la figure 15 montre les droites passant par les 6 dernières paires de points consécutifs. On a donc la droite passant par les points de 1991 et 1992, celle passant par les points de 1992 et 1993, etc. On voit que les six droites sont de pentes extrêmement différentes. Une prévision qui serait établie sur la base de la droite passant par 1991 et 1992 donnerait des prévisions sur la droite jaune inclinée vers le bas à 45 degrés environ. Un an après, les prévisions seraient sur la droite beaucoup plus inclinée, pour arriver ensuite sur la droite verte de pente positive, et ainsi de suite. Il est clair qu'une telle série n'est pas un bon exemple pour un ajustement linéaire mais une droite tracée à main levée à travers les 29 points illustrés donnerait de meilleurs résultats.

La série des prix du cuivre (1800-1997) se prête bien à du lissage comme le montrent les figures 16 et 17. Nous avons comparé le lissage par moyenne mobile, à gauche une moyenne mobile simple d'ordre 15, et à droite une moyenne mobile de Spencer d'ordre 15. La première masque le mieux les variations conjoncturelles au profit de la tendance. La seconde conserve mieux les variations conjoncturelles. Pensons à la grande crise des années 1930-33. La moyenne mobile de Spencer d'ordre 15 la conserve presque intégralement alors que la moyenne mobile simple la lisse.

Figure 15. Les prix du cuivre et des droites de prévision passant par des paires de points.



Figures 16 et 17. Lissage de la série des prix du cuivre par une moyenne mobile simple d'ordre 15 et par une moyenne mobile de Spencer d'ordre 15.



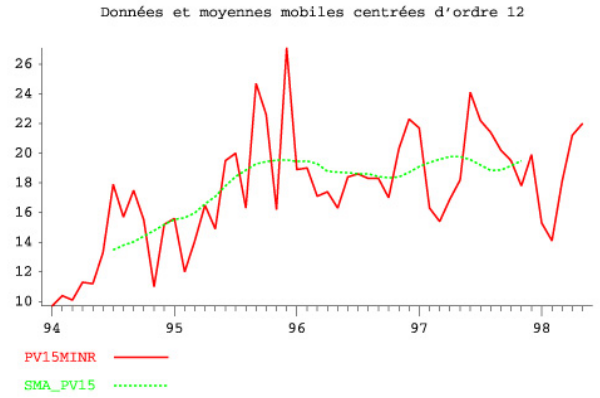
3.2 Exemple 2

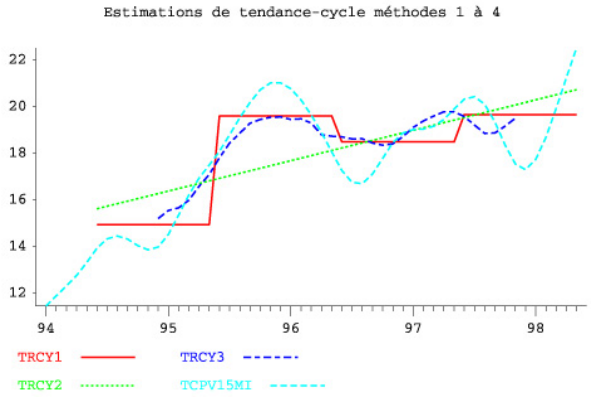
Les données de pourcentage de retards au décollage ont été introduites dans Time Series Expert 2.4 (TSE). Comme la description l'a laissé entrevoir, on soupçonne un comportement saisonnier. Les méthodes suivantes de décomposition saisonnière, sous forme de modèle additif, ont été utilisées pour l'analyse :

méthode 1 : comparaison aux moyennes annuelles
méthode 2 : comparaison à la tendance linéaire
méthode 3 : comparaison aux moyennes mobiles centrées sur 12 mois
méthode 4 : méthode Census X11.

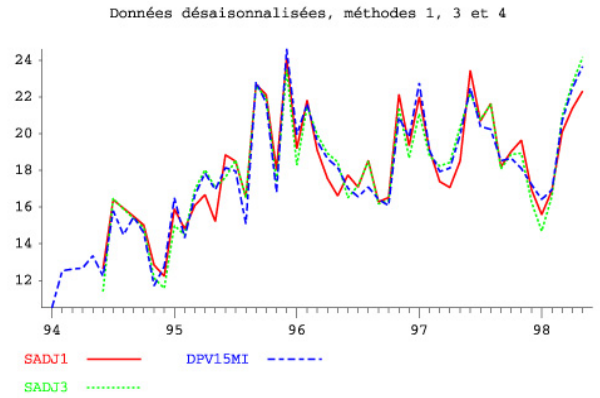
Pour les méthodes 1 à 3, puisqu'il n'y a que 5 données pour la dernière année, les 5 premières données de l'année 1994 n'ont pas été utilisées. Toutes les données ont été employées pour la méthode 4. Les résultats détaillés sont donnés dans l'annexe A.

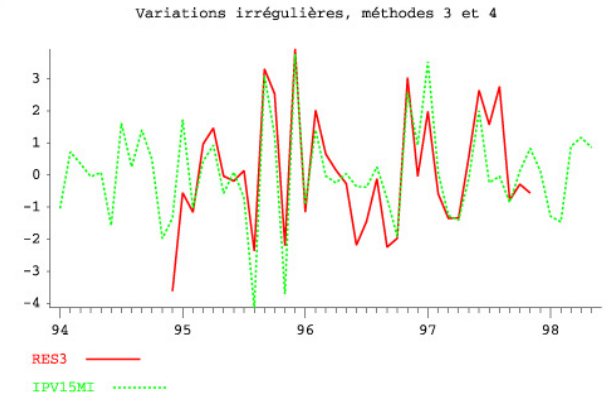
Figures 18 et 19. Pourcentages de retards au décollage. Les données et les moyennes mobiles centrées d'ordre 12, à gauche. Quatre estimations de la tendance-cycle, à droite.





Figures 20 et 21. Les données corrigées des variations saisonnières fournies par les méthodes 1, 3 et 4, à gauche. Les variations irrégulières des méthodes 3 et 4, à droite.





La figure 18 montre que les moyennes mobiles centrées d'ordre 12 révèlent une tendance croissante du pourcentage de retards jusqu'en 1996, suivie d'un plateau. On peut voir la valeur des moyennes mobiles centrées d'ordre 12 dans le sous-tableau B2 de la méthode 4 (annexe A, tableau A.4). Par exemple, ceci permet de calculer le coefficient saisonnier *provisoire* de décembre. A cet effet, il faut calculer les écarts entre données et moyennes mobiles centrées d'ordre 12. Pour décembre, on a les écarts suivants : 15,20 – 15,19 = 0,01 ; 27,10 – 19,55 = 7,55 ; 22,30 – 18,71 = 3,59. On calcule la moyenne de ces trois nombres, soit 11,15/3 = 3,72. On procéderait de même pour chaque mois. Pour obtenir le coefficient définitif donné dans le tableau de la méthode 3 (annexe A, tableau A.3), S(12)=3,628, il faudrait encore normaliser les coefficients saisonniers à une somme de 0.

La méthode 4, la méthode Census X11, est trop complexe pour essayer de retrouver tous les résultats. Néanmoins, on peut retrouver *à partir des autres résultats*, voir tableau A.4 de l'annexe A, la donnée désaisonnalisée de décembre 1995 (sous-tableau D11) et la valeur de la composante d'erreur à la même date (sous-tableau D13). En effet, le coefficient saisonnier de décembre 1995 vaut 2,49. La donnée corrigée des variations saisonnières de décembre 1995 vaut donc 27,10 – 2,49 = 24,61. L'erreur à la même date vaut 24,61 – 21,01 = 3,60.

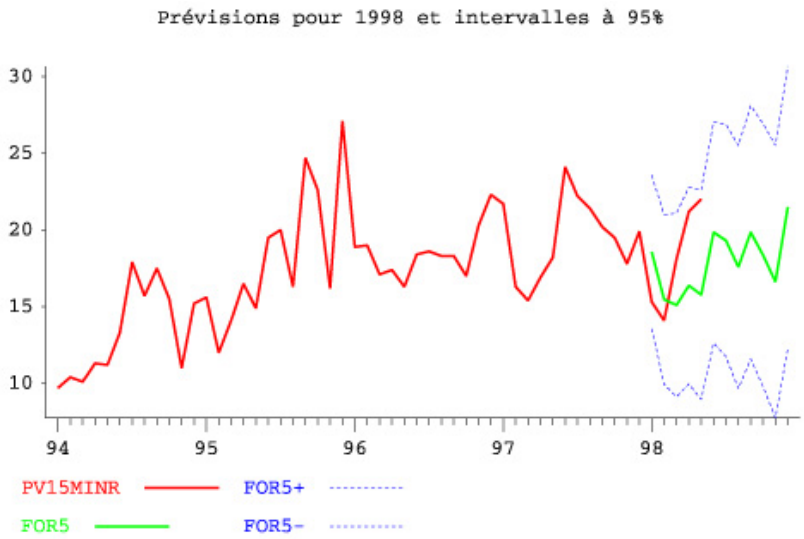
La figure 19 remontre les moyennes mobiles centrées d'ordre 12 (sous le nom TRCY3) ainsi que les estimations de tendance-cycle des 3 autres méthodes (TRCY1, TRCY2 et TCPV15MI). On peut rejeter la méthode 1 parce que la tendance-cycle est sur une courbe en escalier, peu réaliste, et la méthode 2 parce que la tendance-cycle est sur une droite. Ni l'une, ni l'autre ne s'approchent de la tendance-cycle donnée par les moyennes mobiles centrées d'ordre 12 ou par la méthode X-11.

La figure 20 fournit les données corrigées des variations saisonnières des méthodes 1 (SADJ1), 3 (SADJ3) et 4 (DPV15MI). Elles sont relativement proches.

La figure 21 donne les estimations des erreurs des méthodes 3 et 4. On peut interpréter les valeurs de novembre 1995 (négative) et décembre 1995 (positive), sachant qu'il y a eu une grève à Air France et à Air Inter du 9 au 11 novembre et une grève presque totale dans le chemins de fer

français du 29 novembre au 16 décembre 1995. La première a réduit un peu le trafic à Bruxelles a donc diminué la proportion de retards en novembre 1995. La seconde, de longue durée, a entraîné le recours à l'avion et donc des retards plus nombreux en décembre 1995.
On peut aussi se servir de cet exemple pour illustrer l'emploi du lissage exponentiel. On a calculé des prévisions en employant du lissage exponentiel simple sur les données corrigées des variations saisonnières et en restituant ensuite la composante saisonnière (méthode 5). Les résultats détaillés sont fournis dans le tableau A.5 de l'annexe A. L'objectif étant de calculer des prévisions pour l'année 1998 dans son entièreté, les 5 premières données de 1998 n'ont pas été employées dans les calculs.
Notons que la valeur estimée de la constante de lissage du lissage exponentiel simple vaut 0,464. On peut interpréter ceci : la constante de lissage optimale est située entre les cas extrêmes classiques correspondant à $\alpha = 0$ et $\alpha = 1$. On peut aussi interpréter la qualité des prévisions du tableau en annexe et de la figure jointe qui donnent les prévisions (FOR5) obtenues par la méthode 5 avec les intervalles de prévision à 95% (FOR5+ et FOR5-). Les prévisions obtenues en restituant la saisonnalité sont modérément bonnes avec MAPE = 19,7 %. D'après le graphique de la figure 22 qui montre les intervalles de prévision à 95 %, ces intervalles sont très larges, traduisant la très grande incertitude.

Figure 22. Pourcentages de retards. Prévision par lissage exponentiel avec correction saisonnière



3.3 Exemple 3

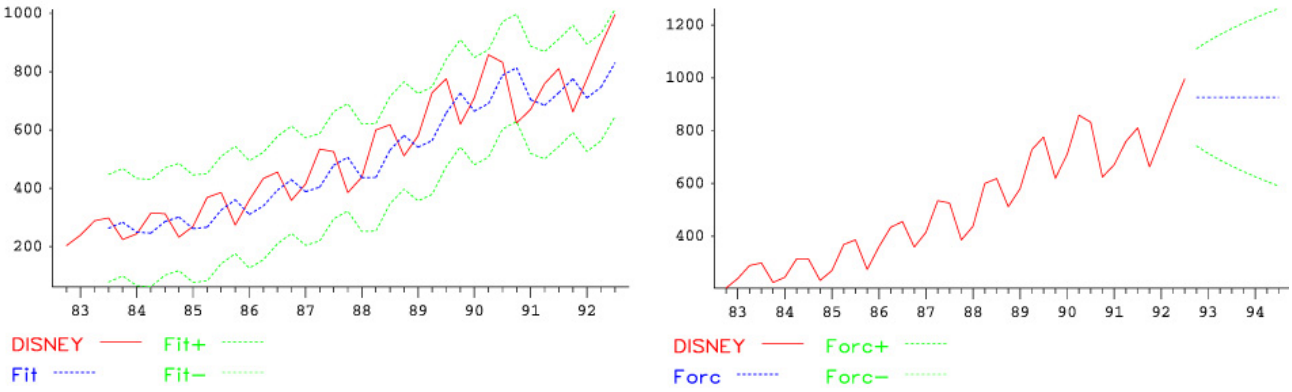
Considérons la série des revenus trimestriels de la société de Walt Disney Company, tels que publiés dans les rapports périodiques. Nous avons analysé les résultats obtenus par les deux méthodes suivantes :

- (SES) lissage exponentiel simple
- (EWS) lissage exponentiel simple avec correction saisonnière.

Voici les résultats obtenus par Time Series Expert qui emploie la representation ARIMA des methods de lissage exponential, comme expliqué par Broze et Mélard (1990). Les résultats détaillés (quelque peu édités) sont présentés en annexe (tableaux B.1 et B.2 de l'annexe B). Pour chacune des deux méthodes, on montre aussi, après réestimation sur l'ensemble des données, les graphiques obtenus pour les prévisions d'horizon 1 et d'origine variable, à gauche, et pour les prévisions d'horizon variable et d'origine au 3e trimestre de 1992, à droite.
Les résultats de la méthode SES (figures 23 et 24) sont relatifs au lissage exponentiel simple caractérisé par une tendance localement constante et l'absence de saisonnalité. Après une phase d'estimation non linéaire, la constante de lissage a comme valeur finale $\alpha = 0{,}52$, voir annexe B, tableau B.1. Les mesures de synthèse de l'ajustement sont notamment la variance des erreurs qui est

égale à 7894. Les erreurs de prévision sur les 4 derniers trimestres sont synthétisées par MSE = 19340 et MAPE = 13,3 %.

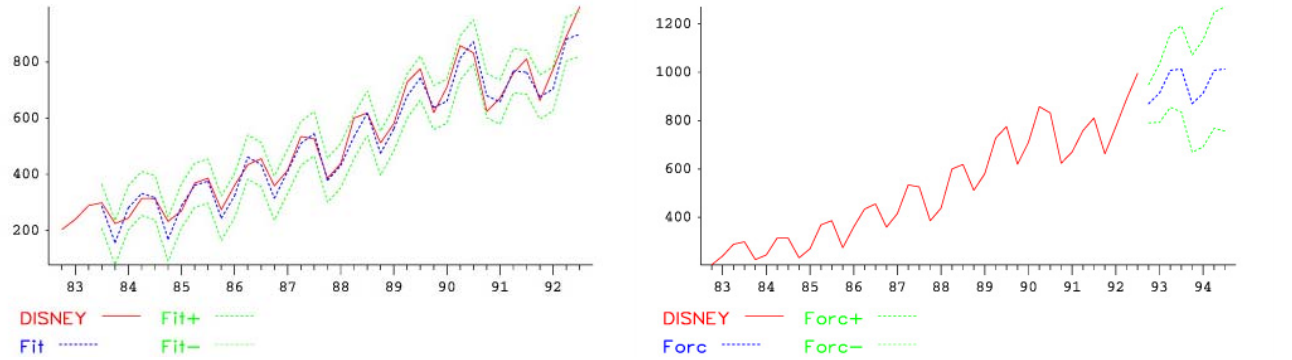
Figures 23 et 24. Les revenus de Walt Disney Company et les prévisions par la méthode SES (d'horizon 1, d'origine variable, à gauche, d'horizon variable, d'origine au 3e trimestre de 1991, à droite)



Les résultats de la méthode EWS (figures 25 et 26) sont relatifs au lissage exponentiel simple avec correction saisonnière. On suppose donc que la série est caractérisée par une tendance localement constante mais sur la série corrigée des variations saisonnières. En finale, la saisonnalité est restituée aux prévisions. Notons que les "coefficients saisonniers" donnés par le programme (pour les trimestres 1 à 4 : 41,2, 97,5, –0,8, –137,9, respectivement) sont obtenus comme des moyennes à travers les années de la série en différence.

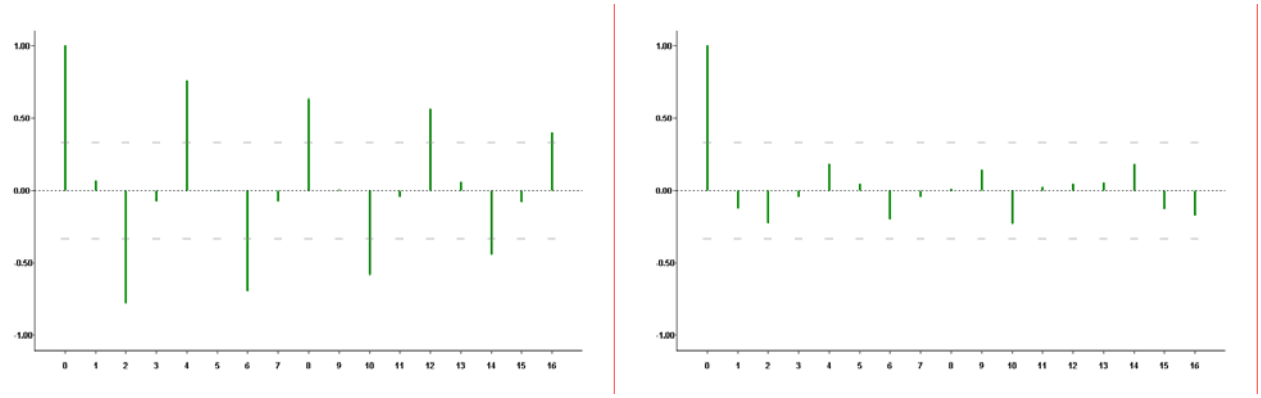
Cette fois, la constante de lissage a comme valeur finale $\alpha = 1{,}28$, voir annexe B, tableau B.2. Les mesures de synthèse de l'ajustement sont notamment la variance des erreurs qui est égale à 1295, beaucoup plus petite que pour le lissage exponentiel simple sans correction saisonnière. Les erreurs de prévision sur les 4 derniers trimestres sont synthétisées par MSE = 8824 et MAPE = 8,4 %. Les prévisions s'avèrent donc meilleures.

Figures 25 et 26. Les revenus de Walt Disney Company et les prévisions par la méthode EWS (d'horizon 1, d'origine variable, à gauche, d'horizon variable, d'origine au 3e trimestre de 1991, à droite).



De plus, afin de motiver l'approche par les modèles ARIMA, on montre (figures 27 et 28) les corrélogrammes des résidus obtenus par les deux méthodes. Visiblement, les erreurs de prévision d'horizon 1 provenant de la méthode SES contiennent beaucoup d'information permettant de prévoir leurs valeurs suivantes, ce qui n'est pas les erreurs de prévision d'horizon 1 provenant de la méthode EWS.

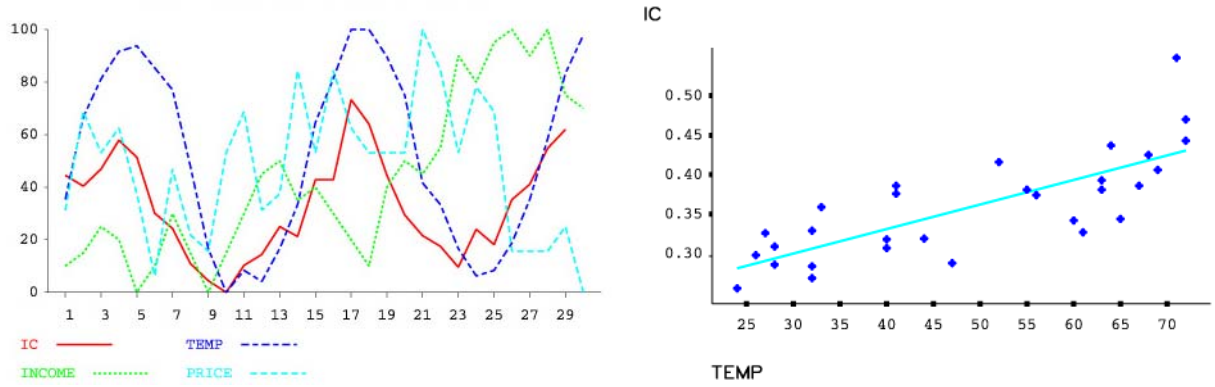
Figures 27 et 28. Les corrélogrammes des résidus ou erreurs de prévision d'horizon 1 obtenus par les méthodes SES (à gauche) et EWS (à droite).



3.4 Exemple 4

Nous avons employé le module expert de régression multiple (ESREG) de TSE. Ce module effectue une régression d'une variable (marquée Dep) en fonction de variables explicatives (marques Exp) mais en considérant également des variables candidates (marquées Can). Ces variables ne sont pas employées dans le modèle dont les résultats sont affichés mais elles peuvent être ajoutées manuellement une à une à ce modèle et le caractère significatif du coefficient de régression correspondant est examiné. Une variable candidate suggérée est indiquée comme étant dans le modèle, "In model". La variable dépendante est celle des ventes, IC. La première donnée n'est pas employée à cause de la présence de la variable retardée LAGTEMP.

Figures 29 et 30. Les données ramenées de 0 à 100 (à gauche) et les ventes de crème glacée en fonction de la temperature (à droite).



La figure 29 montre les données, ramenées à un intervalle allant de 0 à 100, en fonction du temps. Cela permet de constater qu'il y a de la saisonnalité dans plusieurs des variables : IC, TEMP et LAGTEMP.

La figure 30 représente les ventes (IC) en fonction de la température (TEMP) et justifie, si c'était nécessaire, l'emploi de la variable TEMP, la température, pour prédire les ventes de crèmes glacées. Le tableau 31 contient les résultats de la régression avec seulement une constante de régression, toutes les autres variables étant candidates. Le coefficient de détermination R^2 est évidemment nul et l'écart-type des ventes vaut 0,0668, mais le plus intéressant est la proposition d'ajouter LAGTEMP dans le modèle.

Tableaux 31 et 32. Les résultats des deux premiers modèles.

Modèle avec constante seulement

Diagnostic for the following model. Dependent variable is: IC
Explanatory variables are: CONSTANT
Candidate variables are: DATE, INCOME, LAGTEMP, PRICE, TEMP.

- -

Coefficient of determination (R-square)	0.0000
Adjusted R-square	0.0000
Residual variance	0.00446
Residual standard deviation	0.06676

- - - - - - ORDINARY LEAST SQUARES ESTIMATES - - - -

Variables	Coefficient	StdError	Student stat.	2-tail signif.
CONSTANT	0.359	0.01240	28.92	0.0000

- -

Residual autocorrelations analysis:
Durbin-Watson (DW) statistic = 0.387
Approximate acceptance interval : [1.257 - 2.743]
Significant positive 1st order correlation detected.

- -

FINAL DIAGNOSIS PHASE. MY ADVICE:
You should introduce the following candidate variable(s) in the model: LAGTEMP (0.00655108).

Modèle avec LAGTEMP comme variable explicative

Diagnostic for the following model. Dependent variable is: IC
Explanatory variables are: CONSTANT, LAGTEMP.
Candidate variables are: DATE, INCOME, PRICE, TEMP.

- -

Coefficient of determination (R-square)	0.2435
Adjusted R-square	0.2154
Residual variance	0.00350
Residual standard deviation	0.05913

- - - - - ORDINARY LEAST SQUARES ESTIMATES - - - -

Variables	Coefficient	StdError	Student stat.	2-tail signif.
CONSTANT	0.260	0.03516	7.40	0.0000
LAGTEMP	0.00204	0.00069	2.95	0.0065

- -

Residual autocorrelations analysis:
Durbin-Watson (DW) statistic = 0.426
Approximate acceptance interval : [1.257 - 2.743]
Significant positive 1st order correlation detected.

- -

FINAL DIAGNOSIS PHASE. MY ADVICE:
You should introduce the following candidate variable(s) in the model: DATE (0.02522170).
I recommend to keep these variables into the linear model:
CONSTANT, LAGTEMP

Dans le tableau 32, le coefficient de détermination corrigé vaut 0,215 et l'écart-type résiduel vaut 0,0591. Le coefficient de LAGTEMP est bien significatif, valant 0,00204 avec une erreur-type de 0,00069 et une statistique de Student égale à 0,00204/0,00069 = 2,95. Le programme propose d'ajouter DATE. La variable DATE représente l'indice de temps *t*. Le tableau 32 conclut à un effet de tendance que confirme la figure 35 lequel illustre les prédictions obtenues à l'aide du modèle du tableau 32 comparativement aux données des ventes. La statistique de Durbin-Watson est très éloignée de 2, avec 0,426, ce qui révèle de l'autocorrélation dans les résidus.
Le tableau 33 contient la sortie de l'étape qui suit l'étape suivante (d'où la présence de deux variables supplémentaires au lieu d'une). La figure 36 montre les résidus de la régression du tableau 33 en fonction du temps. Il indique que les périodes de 4 semaines qui comportent l'Independence Day ont des résidus positifs (c'est-à-dire des valeurs observées très supérieures aux valeurs prédites par le modèle actuel) parmi les plus élevés. Dès lors, 4 variables binaires ont été construites sur la base de la position des jours fériés correspondants. Par exemple, la variable INDEPEND prend la valeur 1 durant les périodes où tombe cette fête (4, 17 et 30) et 0 les autres périodes. Elles ont été appelées INDEPEND, LABOUR, MEMORIAL et THANKSGI, déclarées comme variables candidates en relançant le traitement du modèle du tableau 33, à l'issue duquel elle observe la fenêtre reprise au tableau 37. En conséquence, on ajoute la variable INDEPEND au modèle, ce qui donne lieu au tableau 34 et aux prévisions de la figure 38. L'économiste sera déçu que les variables économiques n'aient pas été sélectionnées. Cependant, le modèle est excellent à tous points de vue, y compris le fait que la statsitique de Durbin-Watson est proche de 2.

Un modèle supplémentaire considère un modèle pour IC (voir tableau 40) avec toutes les variables disponibles comme variables explicatives. Notons qu'on peut donner, a priori, le signe attendu des coefficients de régression dans le modèle avec toutes les variables explicatives (tableau 40): positifs : DATE (tendance croissante), INCOME (effet de pouvoir d'achat), INDEPEND, TEMP (effet température) ; négatifs : PRICE (effet prix) ; incertains : les autres fêtes, LAGTEMP. Par ailleurs, on a regardé un modèle de fondement économique avec seulement le revenu (INCOME) et les prix (PRICE) comme variables explicatives (voir tableau 41). Enfin la figure 39 contient le graphique des résidus de ce dernier modèle en fonction des prédictions de IC.

Tableaux 33 et 34. Les résultats des modèles des deux premiers modèles.

Modèle avec LAGTEMP, DATE et TEMP

Diagnostic for the following model. Dependent variable is: IC
Explanatory variables are: CONSTANT, DATE, LAGTEMP, TEMP.
Candidate variables are: INCOME, PRICE.

Coefficient of determination (R-square) 0.8562
Adjusted R-square 0.8389
Residual variance 0.00072
Residual standard deviation 0.02679

ORDINARY LEAST SQUARES ESTIMATES

Variables	Coefficient	StdError	Student stat.	2-tail signif.
CONSTANT	0.170	0.02066	8.22	0.0000
DATE	0.00258	0.00061	4.24	0.0003
LAGTEMP	-0.00253	0.00062	-4.07	0.0004
TEMP	0.00546	0.00060	9.14	0.0000

Colinearity analysis:
Unexpected estimated coefficient for variable TEMP
Correlation coefficient between LAGTEMP and TEMP (0.860)

Residual autocorrelations analysis:
Durbin-Watson (DW) statistic = 1.657
Approximate acceptance interval : [1.257 - 2.743]
No significant autocorrelation coefficients found.

FINAL DIAGNOSIS PHASE. MY ADVICE:
I recommend to keep these variables into the linear model:
CONSTANT, DATE, LAGTEMP, TEMP,
Forecasts saved on file: PRED.DB
Residuals saved on file: RES.DB

Modèle avec plusieurs variables

Diagnostic for the following model. Dependent variable is: IC
Explanatory variables are: CONSTANT, DATE, INDEPEND, LAGTEMP, TEMP.
Candidate variables are: INCOME, LABOUR, MEMORIAL, PRICE, THANKSGI.

Coefficient of determination (R-square) 0.9199
Adjusted R-square 0.9065
Residual variance 0.00042
Residual standard deviation 0.02041

ORDINARY LEAST SQUARES ESTIMATES

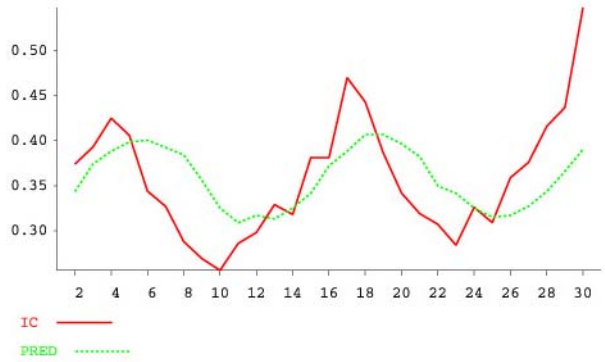
Variables	Coefficient	StdError	Student stat.	2-tail signif.
CONSTANT	0.190	0.01643	11.59	0.0000
DATE	0.00238	0.00047	5.11	0.0000
INDEPEND	0.06110	0.01399	4.37	0.0002
LAGTEMP	-0.00231	0.00048	-4.87	0.0001
TEMP	0.00477	0.00048	9.91	0.0000

Colinearity analysis:
Unexpected estimated coefficient for variable TEMP
Correlation coefficient between LAGTEMP and TEMP (0.860)

Residual autocorrelations analysis:
Durbin-Watson (DW) statistic = 2.207
Approximate acceptance interval : [1.257 - 2.743]
No significant autocorrelation coefficients found.

FINAL DIAGNOSIS PHASE. MY ADVICE:
I recommend to keep these variables into the linear model:
CONSTANT, DATE, INDEPEND, LAGTEMP, TEMP,

Figures 35 et 36. Les prévisions du modèle du tableau 31 (à gauche) et les résidus du modèle du tableau 32 (à droite).



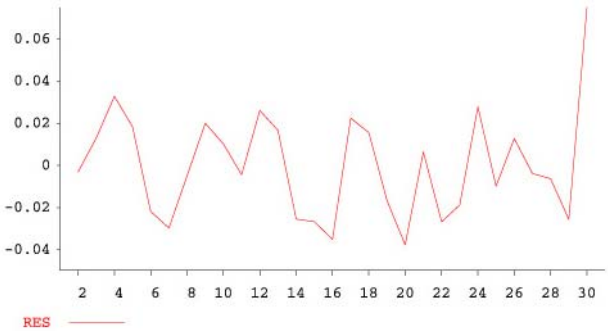
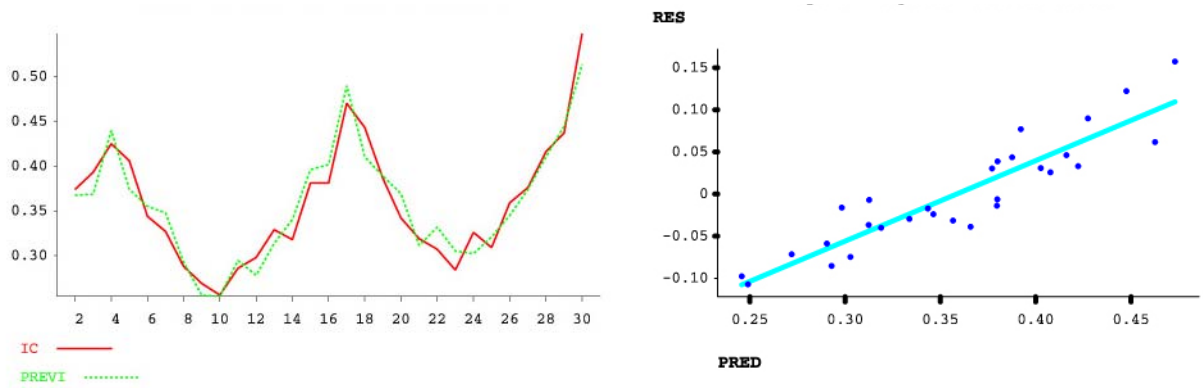


Figure 37. Contenu de la fenêtre après avoir déclaré les variables candidates qui correspondent aux jours fériés

```
--- Names - Coefficient information - Advices ---------- Role --- ExpSign -------
  IC                                                      Dep    ---
  CONSTANT       0.170  Ok   0.0000  In Model             ---    Exp    +
  DATE           0.003  Ok   0.0003  In Model             ---    Exp    +
  LAGTEMP       -0.003  Ok   0.0004  In Model             ---    Exp    ?
  TEMP           0.005  !    0.0000  In Model             ---    Exp    -
  INCOME                                                 ---    Can    +
  INDEPEND                          In Model             ---    Can    ?
  LABOUR                                                 ---    Can    ?
  MEMORIAL                                               ---    Can    ?
  PRICE                                                  ---    Can    -
  THANKSGI                                               ---    Can    ?
  PRED                                                   ---    ---    ?
  RES                                                    ---    ---    ?
  YEAR                                                   ---    ---    ?
----------- Keys: {Space, +, -, D, E, C, I} {Esc, RETURN, arrows} ---------------
```

Figures 38 et 39. Les prévisions du modèle du tableau 33 (à gauche) et le diagramme des résidus du modèle du tableau 34 en fonction des valeurs prédites (à droite).



Tableaux 40 et 41. Les résultats des deux derniers modèles.

Modèle avec toutes les variables

INITIAL MODEL RESULTS :
RESIDUAL VARIANCE .4196E-3 RES. STD ERROR .205E-1
R-SQUARE .9361 R-SQUARE CORRECTED .9059
DURBIN-WATSON 2.3652

DEPENDENT VARIABLE : IC

VARIABLE	VALUE OLS	S.E. OLS	T OLS	2-TAIL PROB	
CONSTANT	.540102	.2078236	2.599	.018	*
DATE	.293719E-2	.9969322E-3	2.946	.008	***
INCOME	-.119322E-2	.1571002E-2	-.760	.457	
INDEPEND	.580488E-1	.1536558E-1	3.778	.001	***
LABOUR	-.116470E-1	.1747965E-1	-.666	.513	
LAGTEMP	-.229712E-2	.5783902E-3	-3.972	.001	***
MEMORIAL	.427203E-3	.1475156E-1	.029	.977	
PRICE	-.900883	.4826614	-1.866	.077	
TEMP	.460774E-2	.5763788E-3	7.994	.000	***
THANKSGI	-.130970E-1	.1705623E-1	-.768	.452	

DESCRIPTIVES STATISTICS

VARIABLE	STD-ERR.	MAXIMUM	AVERAGE	MINIMUM
DATE	8.5147	30.000	16.000	2.0000
INCOME	6.2282	96.000	84.828	76.000
INDEPEND	.30993	1.0000	.10345	.00000
LABOUR	.25788	1.0000	.68966E-1	.00000
LAGTEMP	16.174	72.000	48.345	24.000
MEMORIAL	.30993	1.0000	.10345	.00000
PRICE	.84288E-2	.29200	.27548	.26000
TEMP	16.640	72.000	49.379	24.000
THANKSGI	.25788	1.0000	.68966E-1	.00000

Modèle avec le revenu et les prix

Diagnostic for the following model. Dependent variable is: IC
Explanatory variables are: CONSTANTE, INCOME, PRICE.
Candidate variables are: DATE, INDEPEND, LABOUR, LAGTEMP, MEMORIAL, TEMP, THANKSGI.

Coefficient of determination (R-square) 0.0650
Adjusted R-square -0.0070
Residual variance 0.00449
Residual standard deviation 0.06699

ORDINARY LEAST SQUARES ESTIMATES

Variables	Coefficient	StdError	Student stat.	2-tail signif.
CONSTANT	0.873	0.474	1.84	0.0768
INCOME	0.00033	0.00205	0.16	0.8725
PRICE	-1.971	1.516	-1.30	0.2050

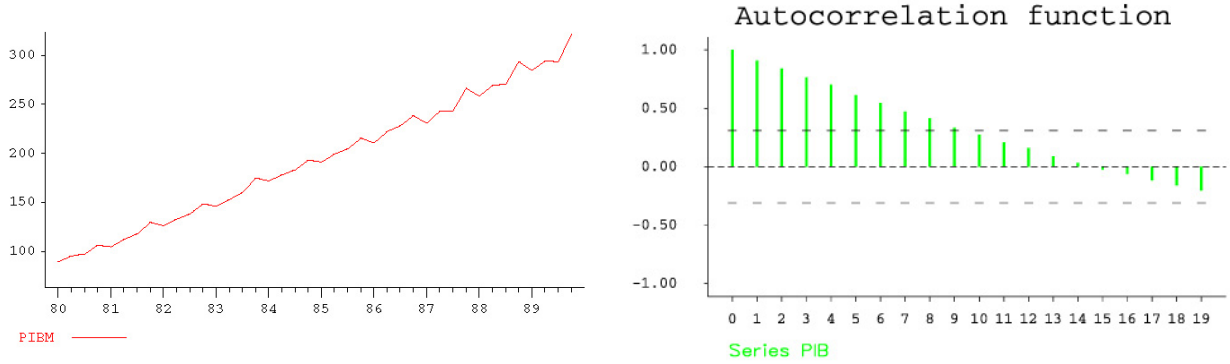
Residual autocorrelations analysis:
Durbin-Watson (DW) statistic = 0.423
Approximate acceptance interval : [1.257 - 2.743]

FINAL DIAGNOSIS PHASE. MY ADVICE:
The previous model was better than the new one.
You should introduce the following candidate variable(s) in the model: INDEPEND(0.00053343).
Instability of estimates over time is due to structural changes.
It means that the relation for the dependent variable has changed over time. So you should try to understand what are these changes.
I recommend to withdraw ONE of these variables from the model: INCOME, PRICE,
Forecasts saved on file: PREVI.DB
Residuals saved on file: RES.DB

3.5 Exemple 5

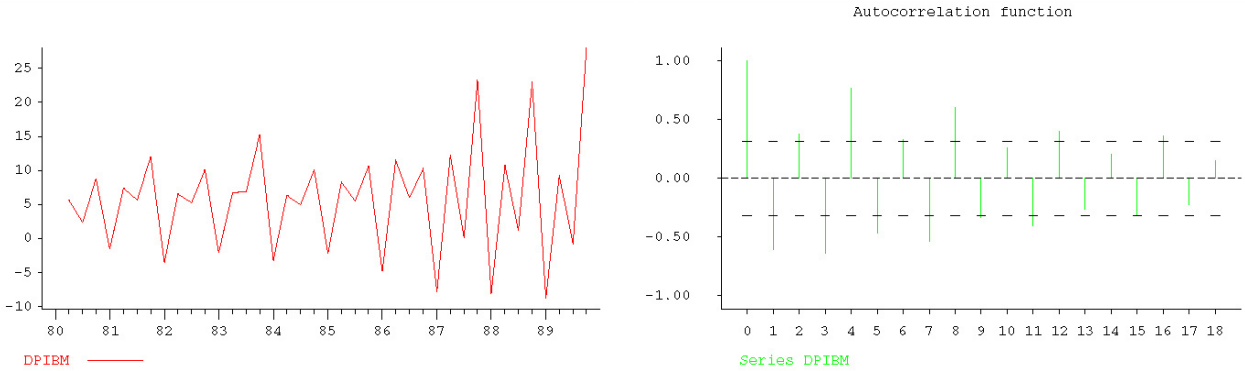
La série du produit intérieur brut de l'Italie (1er trimestre 1980 – 4e trimestre 1989) est représentée dans la figure 42. De manière évidente, cette série ne peut pas avoir été produite par un processus stationnaire, ne fut-ce qu'à cause de la forte tendance presque linéaire alors qu'un processus stationnaire est de moyenne constante. Pour cette raison, regarder les autocorrélations de la série n'a en principe pas de sens puisqu'elles n'existent pas pour le processus générateur. Le corrélogramme de la série brute est néanmoins présenté (figure 43). À noter la décroissance presque linéaire qui est très fréquente pour des séries produites par des processus non stationnaires.

Figures 42 et 43. Les données de la série PIB et le corrélogramme de la série



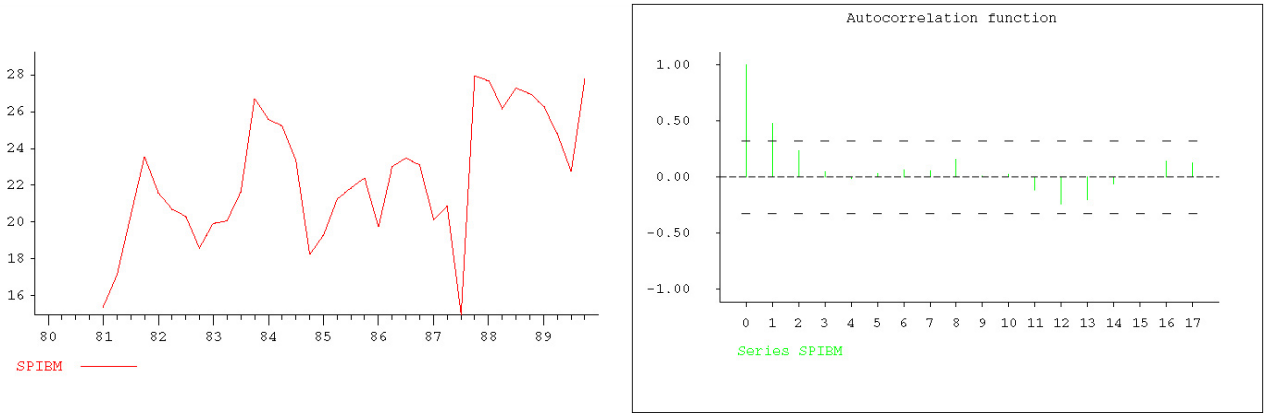
Dans le cas présent, on pourrait envisager d'enlever une tendance linéaire par soustraction mais ce n'est généralement pas recommandé. Nous allons plutôt employer une différence ordinaire (figure 44). La série en différence ne peut raisonnablement pas avoir été produite par un processus stationnaire, cette fois à cause de la saisonnalité qui se manifeste par des variations de nature périodique, avec un creux au 1er trimestre de chaque année et un pic au 4e trimestre de chaque année. Le corrélogramme est néanmoins proposé dans la figure 45. À noter les fortes autocorrélations de retard 4, 8, 12, etc., qui confirment la saisonnalité.

Figures 44 et 45. Les données de la série ∇PIB et le corrélogramme de la série



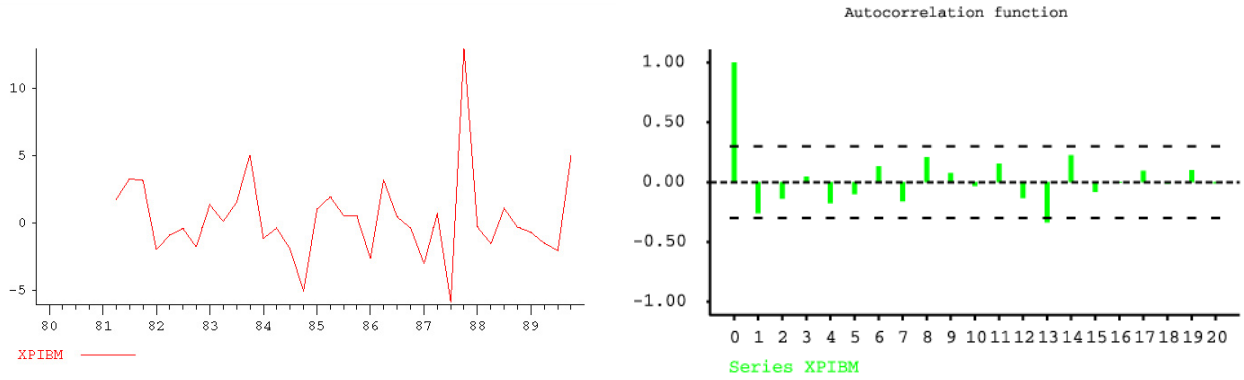
En employant une différence saisonnière sur la série brute, il apparaît encore des variations de niveau et même une tendance légèrement croissante alors qu'un processus stationnaire est de moyenne constante. La série de la figure 46 peut donc difficilement avor été générée par un processus stationnaire. Pour cette raison, regarder les autocorrélations n'a pas de sens. Le corrélogramme (figure 47) montre que seul le retard 1 est statistiquement significatif. Or, un processus non stationnaire a souvent des autocorrélations élevées qui ne décroissent pas rapidement vers 0. Ce n'est pas le cas ici.

Figures 46 et 47. Les données de la série ∇_4PIB et le corrélogramme de la série



Enfin, la série en différence ordinaire et saisonnière peut être raisonnablement être produite par un processus stationnaire, voir figure 48. Il y a bien un pic isolé au 4e trimestre de 1987. Non seulement le corrélogramme de la figure 49 est compatible avec un processus stationnaire, mais il est même compatible avec un processus bruit blanc. En effet, ici les autocorrélations sont presque toutes dans la bande à 95 %, avec une exception pour le retard 13, soit 3 ans et un trimestre. Comme cette autocorrélation est tout juste significative à 5%, il ne faut pas s'en préoccuper. On peut donc accepter l'hypothèse de bruit blanc.

Figures 48 et 49. Les données de la série $\nabla\nabla_4$PIB et le corrélogramme de la série

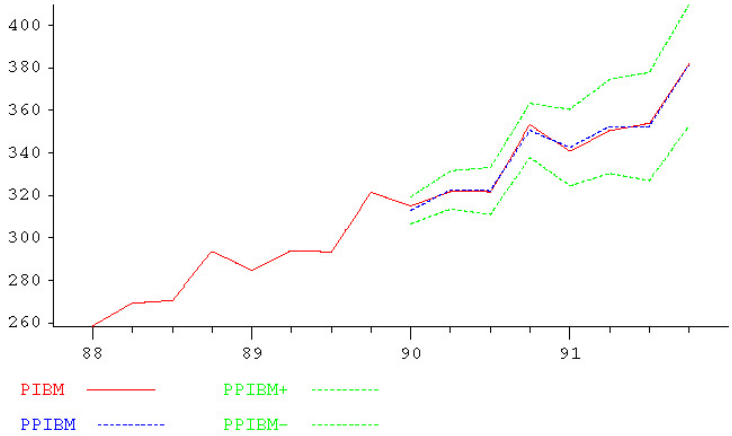


On écrit donc l'équation du modèle

$$\nabla\nabla_4 \text{PIB}_t = e_t.$$

Par conséquent $\text{PIB}_t = \text{PIB}_{t-1} + \text{PIB}_{t-4} - \text{PIB}_{t-5} + e_t$. La prévision pour le 1er trimestre de 1990 vaut la donnée du 4e trimestre de 1989 + la donnée du 1er trimestre de 1989 – la donnée du 4e trimestre de 1988 = 321,26 + 284,81 – 293,50 = 312,57. C'est ce qu'on trouve (voir annexe C, tableau C.1). La donnée est 315,12 donc l'erreur est égale à 2,54. Les prévisions de la série pour 1990-1991 sont données en bleu, comparées avec les données en rouge. Elles sont tellement proches des données qu'on a de la peine à les distinguer dans la figure 50. Le modèle basé sur la série en différences ordinaire et saisonnière fournit un critère MAPE égal à 0,7 % pour les années 1990 et 1991. C'est excellent.

Figure 50. Les données de la série PIB et les prévisions obtenues par le modèle $\nabla\nabla_4 \text{PIB}_t = e_t$

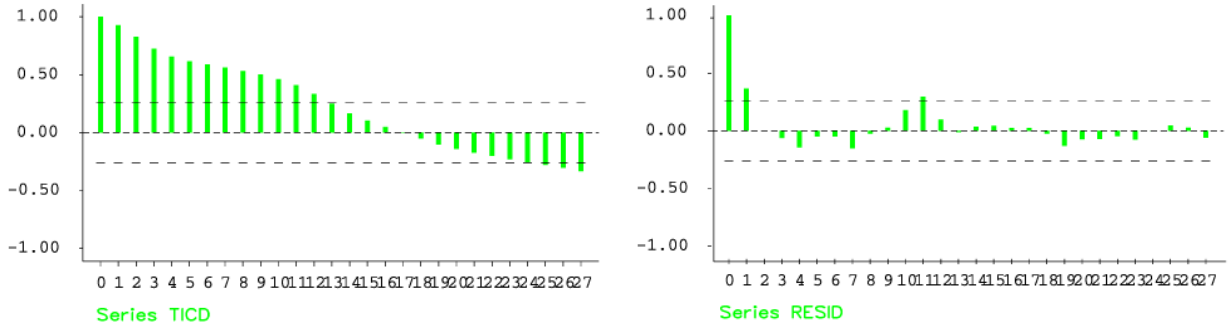


Notons qu'on aurait pu éventuellement considérer un modèle qui serait basé sur la série en différence saisonnière. L'équation du modèle est $\nabla_4 \text{PIB}_t = m + e_t$, où la constante m vaut 22,38. Par conséquent la prévision pour le 1er trimestre de 1990 vaut la donnée du 1er trimestre de 1989 + 22,38 = 284,81 + 22,38 = 307,18. C'est ce qu'on trouve. Ce modèle fournit un critère MAPE égal à 3,0 % pour les années 1990 et 1991, donc moins bon que le modèle avec les différences ordinaire et saisonnière.

3.6 Exemple 6

Le graphe de la série TICD a été montré dans la figure 3. Les corrélogrammes contenus dans les figures 51 et 52 justifient pourquoi la série ne peut pas avoir été générée par un processus stationnaire et qu'elle doit être analysée en différence première. Les autocorrélations de ∇TICD sortent de la bande à 95% pour les retards 1 et 11. L'autocorrélation de retard 1 est explicable par l'inertie des systèmes économiques mais pas celle de retard 11. Si c'était 12, on pourrait envisager une indication de saisonnalité. Il est toutefois bien connu que la plupart des séries financières n'ont pas de saisonnalité. Les rares exceptions sont relatives à des pays à vocation touristique et pour des séries comme la masse monétaire. Comme ce n'est pas 12 mais 11, il est donc préférable d'ignorer cette dernière et de la mettre sur le compte de la probabilité de 5 % avec laquelle une autocorrélation de retard spécifié peut sortir des limites. Il reste néanmoins l'autocorrélation de retard 1. On conclut que la série n'a pas été engendrée par un processus de type bruit blanc et à la présence d'autocorrélation d'ordre 1. Ceci est compatible avec un modèle moyenne mobile d'ordre 1, d'autant plus que le test global ne conduit pas non plus au rejet de l'hypothèse. On est donc tenté de conclure à un modèle ARIMA(0,1,1).

Figures 51 et 52. Les corrélogrammes des séries TICD, à gauche, et ∇TICD, à droite.



Voici, dans le tableau 53, un extrait des résultats fournis par TSE (les résultats complets sont repris dans l'annexe D, tableau D.1).

Tableau 53. Extrait des résultats du tableau D.1 de l'annexe D.

```
SERIES READ FROM DISK,          NAME IS TICD.DB                  ,LENGTH   61
        TIME INTERVAL : FROM DEC1974 TO DEC1979.
  /\/\/\/\/\/\/\/\/\/\
=== IDENTIFICATION OF A TIME SERIES MODEL
=== MODEL DESCRIPTION                  FORM      DEGREE/ORD PARAMETERS NUMBER
- DIFFERENCE                         REGULAR              1
=== SUMMARY MEASURES  <V>
TOTAL  NUMBER  OF  PARAMETERS =  1  STANDARD DEVIATION =  .540535
=== DATA ANALYSIS WITH  60 DATA POINTS BEGINNING AT TIME  JAN1975===
MEAN =  .766667E-01 ,T-STATISTIC =   1.10        (FOR TESTING ZERO MEAN)
=SIGNIFICANT AUTOCORRELATIONS (USING BARTLETT LIMITS) <A(OR)S>
 .2 - 1 %       1:  .3989
  1 - 5 %      11:  .3423
```

Nous avons vu plus haut (paragraphe 2.6) que la formule donnant l'autocorrélation d'ordre 1 d'un processus MA(1), d'équation (3) est donnée par $\rho_1 = -\theta/(1 + \theta^2)$. En égalant le second membre à l'autocorrélation estimée sur la série, $r_1 = 0{,}399$, et en résolvant l'équation $-\theta/(1 + \theta^2) = 0{,}399$, on est conduit à une estimation de θ : $\hat{\theta} = -0{,}498$. Une telle équation a deux solutions mais la seconde est supérieure à 1 en valeur absolue et ne satisfait pas à la condition d'inversibilité. Que peut-on reprocher à cette méthode, application de la méthode des moments pour l'estimation des paramètres d'un modèle statistique ? Elle n'est pas assez générale et elle est difficile à appliquer pour des modèles plus complexes, comportant plusieurs paramètres. Ici elle conduit à une équation du second degré en θ. Dans le cas de plusieurs paramètres, ce ne serait pas aussi simple et on aurait des calculs numériques intenses. L'équation pourrait aussi ne conduire à aucune solution. En outre, la méthode n'a pas de très bonnes propriétés statistiques. On lui préfère l'estimation par la méthode du maximum de vraisemblance.

Nous avons ajouté une constante au modèle pour prendre en compte une éventuelle tendance dans les taux d'intérêt. Notons que la moyenne 0,077 n'est pas statistiquement significative, puisque la statistique de Student correspondante vaut 1,10 < 2.
Nous avons estimé les paramètres en employant TSE. Voici, dans le tableau 54, un extrait des résultats qu'il a fourni (annexe D, tableau D.2).

Tableau 54. Résultats de l'estimation par TSE d'un modèle MA(1), avec constante, sur ∇ TICD

```
 SERIES READ FROM DISK,          NAME IS TICD.DB                ,LENGTH   61
         TIME INTERVAL : FROM DEC1974 TO DEC1979.
   /\/\/\/\/\/\/\/\/\/\
  1 PARAMETERS WITH STARTING VALUES :
     1     MA   1  .00000
 === ESTIMATION BY MAXIMIZATION OF THE EXACT (LOG)LIKELIHOOD
      (FAST ALGORITHM WITH TOLERANCE  1.0E-05)
 === MODEL DESCRIPTION                  FORM      DEGREE/ORD PARAMETERS NUMBER
 - DIFFERENCE                         REGULAR              1
 - ADDITIVE CONSTANT                  AUTOMATIC
 - ARMA MODEL
       MOVING AVERAGE POLYNOMIAL      REGULAR              1     MA  nn      1
 NON LINEAR ESTIMATION:
 ITER SUM OF SQ MA   1
  0    17.24       .000
  1    13.86      -.404
  2    13.72      -.509
  3    13.71      -.494
  4    13.71      -.496
  5    13.71      -.495
  6    13.71      -.495
 =ITERATION STOPS - RELATIVE CHANGE IN EACH COEFFICIENT LESS THAN  1.00000E-03
 FINAL VALUES OF THE PARAMETERS                    WITH 95% CONFIDENCE LIMITS
              NAME       VALUE         STD ERROR   T-VALUE  LOWER          UPPER
      1      MA   1    -.49532         .11590        -4.3    -.73         -.26
 THE FOLLOWING PARAMETERS WERE ESTIMATED SEPARATELY
          MEAN     7.66667E-02
 === SUMMARY MEASURES  <V>
 TOTAL  NUMBER  OF  PARAMETERS =  2  STANDARD DEVIATION =  .485097
```

Après 6 itérations au cours desquelles on voit décroître la fonction objectif, exprimée comme une somme de carrés, l'estimation de θ est obtenue et vaut –0,495. Par ailleurs, la constante estimée comme la moyenne des valeurs de ∇ TICD vaut 0,077. Le coefficient θ est significatif à 5 %. En effet, l'erreur-type vaut 0,116, donc la statistique de Student vaut –4,3 et est supérieure à –1,96 en valeur absolue. L'écart-type résiduel, estimation de l'écart-type des innovations, vaut 0,485. Le tableau 55 montre les autocorrélations de la série résiduelle. L'équation du modèle est donc

$$\nabla \mathrm{TICD}_t - 0{,}0767 = (1 + 0{,}495B)\, e_t. \qquad (8)$$

Tableau 55. Analyse résiduelle du modèle du tableau 54.

```
=== RESIDUAL ANALYSIS WITH  60 RESIDUALS, BEGINNING AT TIME JAN1975===
 MEAN = -.872985E-03 ,T-STATISTIC =   -.01        (FOR TESTING ZERO MEAN)
 =OUTLIERS <R(OR)S>
  .2 - 1 % JAN1975: -1.323     OCT1979:  1.290
   1 - 5 % NOV1978:  .9776
 =SIGNIFICANT AUTOCORRELATIONS (USING BARTLETT LIMITS) <A(OR)S>
   1 - 5 %      11:  .2870
 =LJUNG-BOX PORTMANTEAU TEST STATISTICS ON RESIDUAL AUTOCORRELATIONS <L>
  ORDER D.F. STATISTIC  SIGNIFICANCE
  24     23   11.24        .981
  ---> WRITTEN TO FILE : RESID.DB , LENGTH =   60
```

L'hypothèse de validité du modèle n'est pas rejetée. Les tests individuels de bruit blanc basés sur les autocorrélations comme sur les autocorrélations partielles ne mettent pas en cause l'hypothèse de bruit blanc des innovations. Seule subsiste l'autocorrélation de retard 11 qui n'est pas explicable. Le test global portant sur les 24 premières autocorrélations fournit une statistique de Ljung-Box *Q'* égale à 11,24. Un regard sur le tableau des quantiles de la loi χ^2 révèle que la valeur critique à utiliser est de 35,2. En effet, le nombre de degrés de liberté à prendre en considération est égal à 24 – 1 soit 23. Puisque 11,24 < 35,2, on ne rejette pas l'hypothèse. Le tableau 56 contient les prévisions pour l'année 1980.

Tableau 56. Prévision au moyen du modèle du tableau 54.

```
=== FORECASTING FROM  DEC1979  WITH FRESH DATA <F>
   DATE   OBSERVATION   FORECAST      ERROR  % ERROR  95% FORECAST INTERVAL
JAN1980                   13.328                          12.377     14.278
FEB1980                   13.404                          11.694     15.115
MAR1980                   13.481                          11.257     15.705
APR1980                   13.558                          10.918     16.197
MAY1980                   13.634                          10.636     16.632
JUN1980                   13.711                          10.393     17.029
JUL1980                   13.788                          10.178     17.397
AUG1980                   13.864                           9.984     17.744
SEP1980                   13.941                           9.809     18.073
OCT1980                   14.018                           9.648     18.387
NOV1980                   14.094                           9.499     18.689
DEC1980                   14.171                           9.361     18.981
```

On peut aisément vérifier ces prévisions. D'abord,

$$\hat{Y}_{1979.12}(1) = m - \theta\hat{e}_{1979.12} = 0,0767 + 0,495 \times (-0,342) = -0,092$$

en allant récupérer le résidu de décembre 1979 (non contenu dans les sorties). Par conséquent

$$\hat{y}_{1979.12}(1) = y_{1979.12} + \hat{Y}_{1979.12}(1) = 13,420 - 0,092 = 13,328.$$

Pour $h > 1$, les prévisions $\hat{Y}_{1979.12}(h)$ valent $m = 0,077$, d'où $\hat{y}_{1979.12}(h) = \hat{y}_{1979.12}(h-1) + 0,077$.

De même, on peut vérifier les intervalles de prévision à 95%. Le quantile d'ordre 0,975 de la loi normale centrée réduite valant 1,96, comme $\theta = -0,495$, la quantité $(1 - \theta)^2$ vaut 2,236. Sachant que l'écart-type résiduel vaut 0,485, la demi largeur de l'intervalle de prévision d'horizon h vaut

$$1,96 \times 0,485 \times [1 + (h - 1).2,236]^{1/2} . \qquad (9)$$

Nous avons aussi estimé un modèle sans constante. A priori, ce modèle peut être justifié par le fait que la moyenne de la série n'est pas statistiquement significative dans le premier modèle. Le modèle estimé (voir annexe D, tableau D.3) a pour équation

$$\nabla \mathrm{TICD}_t = (1 + 0,500B)\, e_t. \qquad (10)$$

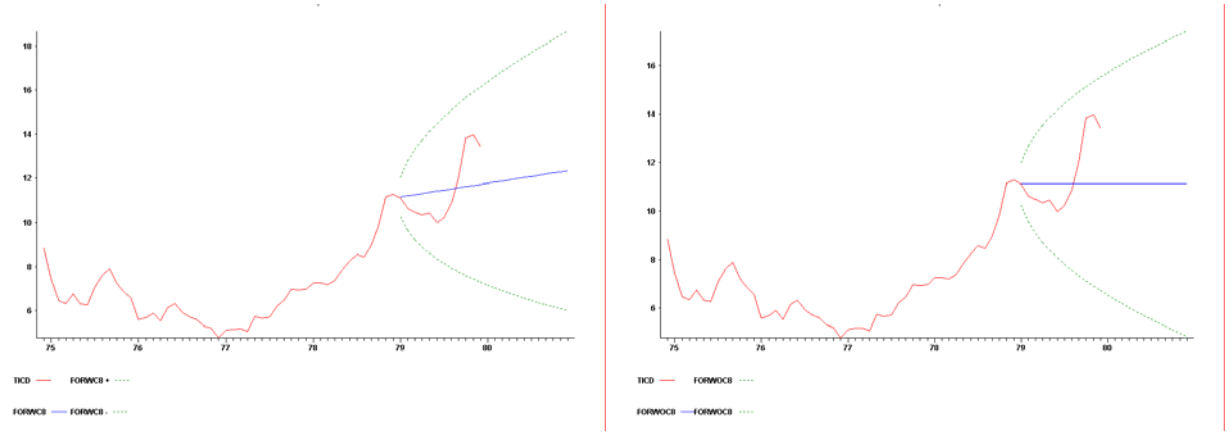
Le paramètre unique estimé est significatif à 5 %. L'écart-type résiduel est légèrement supérieur : 0,482.

Nous avons réestimé les deux modèles (8) et (10) sur la série partielle qui s'arrête en décembre 1978. Les modèles estimés (voir annexe D, tableaux D.4 et D.5) ont pour équation

$$\nabla \mathrm{TICD}_t - 0,051 = (1 + 0,459B)\, e_t \quad \text{et} \quad \nabla \mathrm{TICD}_t = (1 + 0,465B)\, e_t.$$

Les figures 57 et 58 montrent les prévisions obtenues par les deux méthodes. Les résultats ne sont pas exceptionnels au cours de cette période troublée mais le modèle avec constante donne des résultats plus satisfaisants.

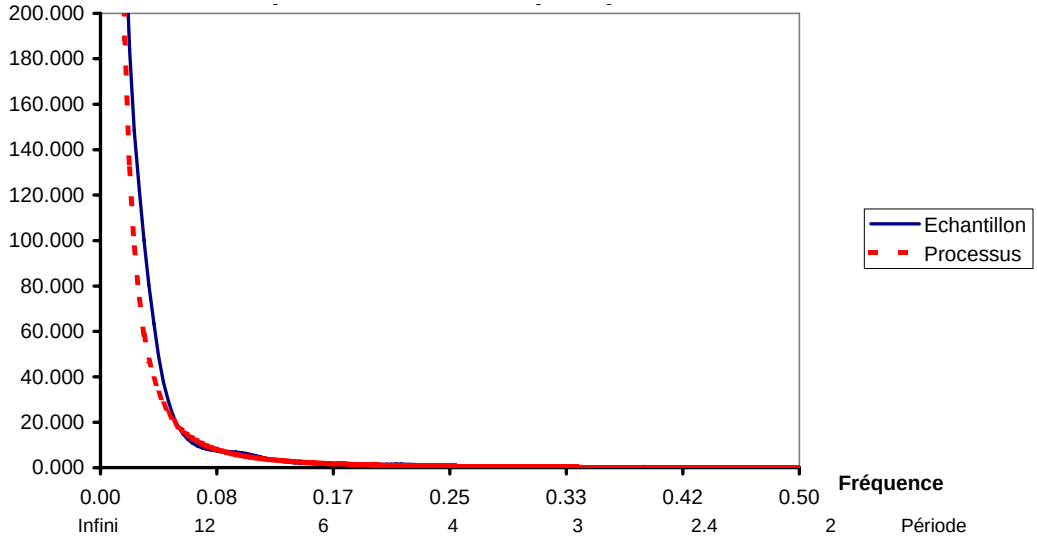
Figures 57 et 58. Les prévisions pour 1979 des modèles respectivement avec et sans constante.



Notons que les prévisions du modèle sans constante sont égales quel que soit l'horizon, comme pour la méthode de prévision naïve ou le lissage exponentiel simple. En effet, le lissage exponentiel simple, de constante de lissage α, possède la forme ARIMA(0,1,1) sans constante avec $\theta = 1 - \alpha$. L'estimation de la constante de lissage par maximum de vraisemblance conduit en effet à l'estimation de α égale à 1,465 dans le cas de la série écourtée et 1,500 dans le cas de la série

complète. Le fait que la constante de lissage soit en dehors de l'intervalle [0 ; 1] peut surprendre mais ceci illustre la supériorité de la modélisation ARIMA par rapport aux méthodes élémentaires de prévision.
Nous avons employé l'analyse spectrale sur une série artificielle de longueur 400 qui simule la série TICD. Cette série, appelée DIFMA1_5, est générée par le processus d'équation (10). Son spectre a été estimé dans le tableur Excel par l'approche autorégressive, en employant le plus grand ordre possible $p = 16$. Cela signifie qu'on a régressé les observations sur elles-mêmes, avec un décalage de p. On a alors employé les estimations obtenues dans le dénominateur d'une expression similaire à (6) mais où le degré 4 est remplacé par 16. Dans la figure 59, on a également représenté une approximation du spectre généralisé du processus.

Figure 59. Spectre estimé de la série DIFMA1_5 et spectre généralisé du processus ARIMA(0,1,1).



3.7 Exemple 7

On essaye de modéliser la consommation de fuel lourd en France en tonnes. Les données portent sur la période qui va de janvier 1983 à juin 1999. Cette période a été caractérisée par une diminution simultanée du prix du pétrole brut et du cours du dollar après les crises pétrolières de 1973-1974 et de 1980. Il y a bien eu les variations à la hausse du prix du pétrole après l'invasion du Koweit (août 1990) et lors de la guerre du Golfe (janvier 1991). Cette consommation de fuel lourd reprend la consommation du secteur résidentiel (chauffage des locaux et des maisons), des différentes industries et des centrales électriques. La consommation d'Electricité de France (EDF) est très fluctuante parce que le fuel lourd sert d'appoint à l'énergie nucléaire et à l'énergie hydro-électrique. Le graphique de la série est présenté dans la figure 1.
On veut modéliser la série de consommation de fuel lourd afin de prédire son évolution. Afin de choisir un modèle, on réserve les données de janvier 1998 à juin 1999. La modélisation s'arrête donc en décembre 1997. Plusieurs méthodes sont employées :

- le lissage exponentiel double avec correction pour la saisonnalité, en version multiplicative ;
- le lissage exponentiel de Winters en version multiplicative ;
- des modèles ARIMA, d'abord sur les données brutes, puis sur les données transformées par logarithmes.

Le logiciel Time Series Expert est employé.
A. Le lissage exponentiel est d'abord utilisé. La méthode de lissage double de Brown est choisie parce qu'il y a une tendance mais elle est appliquée sur la série corrigée des variations saisonnières, après transformation logarithmique, et la saisonnalité est restituée aux prévisions avant transformation exponentielle. La raison est que la saisonnalité présente plutôt les caractéristiques d'un modèle multiplicatif. Le tableau 60 est un extrait de la sortie. La figure 61 montre les

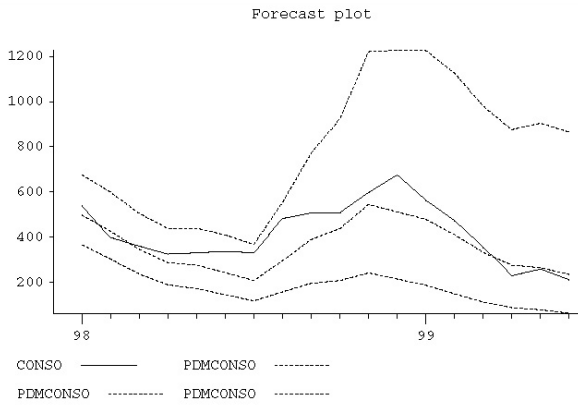
prévisions obtenues. La figure 62 présente le graphique des autocorrélations de la série d'erreurs de prévision sur la période 1983-1997. Elle montre que de l'information se cache dans les résidus.

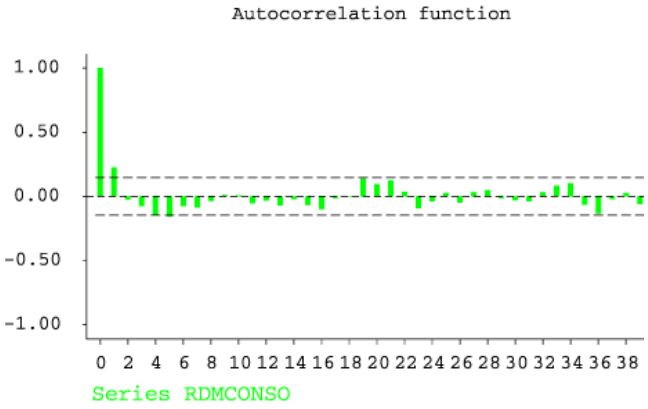
Tableau 60. Extrait des résultats d'un modèle de Brown, avec correction saisonnière multiplicative.

```
 === FITTING OF THE PARAMETERS OF CLASSIC SHORT-TERM FORECASTING METHODS
     (CO-AUTHORS: O. ROULAND AND L. BROZE)
 EXPONENTIAL SMOOTHING METHOD
   -TREND    COMPONENT : LINEAR TREND, BROWN MODEL
   -SEASONAL COMPONENT : NONE
 === MODEL DESCRIPTION                    FORM      DEGREE/ORD PARAMETERS NUMBER
 - SEASONAL PERIOD                          12
 - NORMALIZED BOX COX TRANSFORMATION  LOGARITHMS              BOXC 1      0
 - DIFFERENCE                         REGULAR                2
 - ADDITIVE DETERMINISTIC SEASONALITY AUTOMATIC
 - CLASSIC SHORT TERM FORECASTING     SEE ABOVE
 NON LINEAR ESTIMATION:
 ITER SUM OF SQ ALPHA
  0   1262.1E+03  .500
  1   1164.7E+03  .391
 …
 16   1111.4E+03  .244
 =ITERATION STOPS - RELATIVE CHANGE IN EACH COEFFICIENT LESS THAN  1.00000E-03
 FINAL VALUES OF THE PARAMETERS                        WITH 95% CONFIDENCE LIMITS
              NAME      VALUE         STD ERROR   T-VALUE  LOWER          UPPER
      1      ALPHA     .24400       2.60955E-02     9.4     .19          .30

 THE FOLLOWING PARAMETERS WERE ESTIMATED SEPARATELY
           SEAS 1   -4.4429
           SEAS 2   -44.646
           SEAS 3   -29.911
           SEAS 4    14.216
           SEAS 5    73.680
           SEAS 6   -41.310
           SEAS 7   -20.003
           SEAS 8    266.49
           SEAS 9   -40.612
           SEAS10   -79.268
           SEAS11    48.683
           SEAS12   -142.88
 …
 === SUMMARY MEASURES  <V>
 TOTAL  NUMBER  OF  PARAMETERS = 12  STANDARD DEVIATION =  81.0484
 === FORECASTING FROM  DEC1997  WITH FRESH DATA <F>
    DATE    OBSERVATION    FORECAST      ERROR  % ERROR  95% FORECAST INTERVAL
 JAN1998         540.0       497.3       42.7      7.9       366.6       674.8
 …
 JUN1999         211.0       234.8      -23.8     11.3        63.8       864.5
  CUMULATED ERROR          :    1015.2 (= 13.6%); MEAN ERROR:       56.4
  MEAN ABSOLUTE ERROR (MAE):      68.1 (= 16.4%);
  ROOT MEAN SQUARE ERROR   :      84.1 (= 20.3%); MEAN SQUARE ERROR: 7.077E+03
  MEAN ABSOLUTE PERCENTAGE ERROR (MAPE):   16.0% .
```

Figures 61 et 62. Prévisions du modèle du tableau 60 (à gauche) et corrélogramme des résidus (à droite).



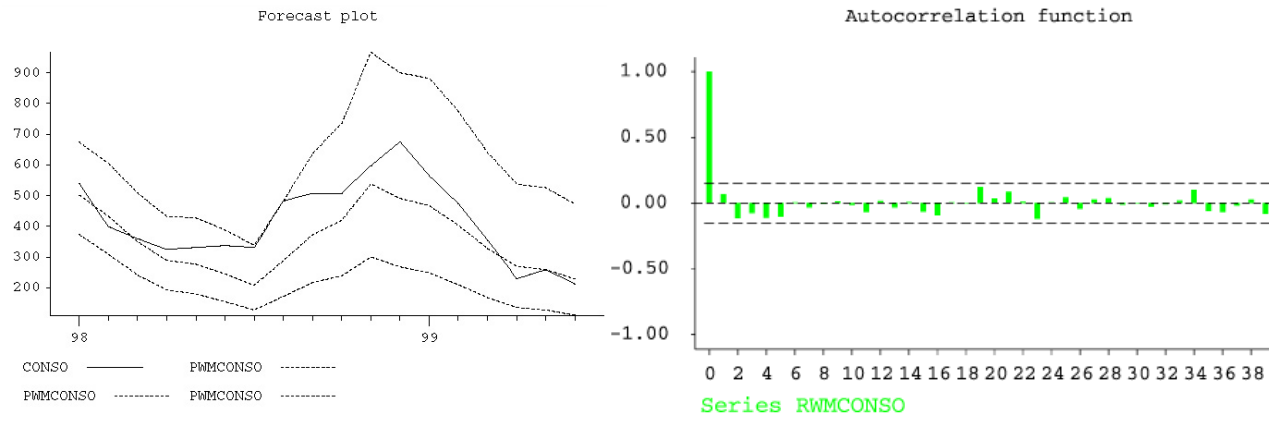


B. Le lissage exponentiel de Winters est utilisé ensuite, en version multiplicative. Plus précisément, c'est la version additive de la méthode qui est appliquée sur les logarithmes des données. Le tableau 63 est un extrait de la sortie. La figure 64 montre les prévisions obtenues. La figure 65 présente le graphique des autocorrélations de la série d'erreurs de prévision sur la période 1983-1997.

Tableau 63. Extrait des résultats d'un modèle de Winters en version multiplicative.

```
 === FITTING OF THE PARAMETERS OF CLASSIC SHORT-TERM FORECASTING METHODS
     (CO-AUTHORS: O. ROULAND AND L. BROZE)
 EXPONENTIAL SMOOTHING METHOD
   -TREND    COMPONENT : LINEAR TREND, HOLT MODEL
   -SEASONAL COMPONENT : ADDITIVE
 === MODEL DESCRIPTION                    FORM      DEGREE/ORD PARAMETERS NUMBER
 - SEASONAL PERIOD                                      12
 - NORMALIZED BOX COX TRANSFORMATION  LOGARITHMS                BOXC 1      0
 - DIFFERENCE                         REGULAR             1
 - DIFFERENCE                         SEASONAL            1
 - CLASSIC SHORT TERM FORECASTING     SEE ABOVE
 NON LINEAR ESTIMATION:
 ITER SUM OF SQ ALPHA     GAMMA     DELTA
  0   1239.6E+03  .500      .100      .100
 …
 11   1156.1E+03  .539     2.734E-06  .173
 =ITERATION STOPS - RELATIVE CHANGE IN SUM OF SQUARES LESS THAN  1.00000E-06
 FINAL VALUES OF THE PARAMETERS                        WITH 95% CONFIDENCE LIMITS
           NAME      VALUE         STD ERROR   T-VALUE  LOWER          UPPER
      1    ALPHA     .53884        .56847          .9    -.58          1.7
      2    GAMMA    2.73366E-06    1.3229          .0    -2.6          2.6
      3    DELTA     .17291        .14534         1.2    -.11          .46
 = UNDERLYING SARIMA MODEL
   DEGREE OF REGULAR DIFFERENCING   =   1
   DEGREE OF SEASONAL DIFFERENCING  =   1 (PERIOD  =   12)
   MA   1:   .4612   MA   2:   .0000   MA   3:   .0000   MA   4:   .0000
   MA   5:   .0000   MA   6:   .0000   MA   7:   .0000   MA   8:   .0000
   MA   9:   .0000   MA  10:   .0000   MA  11:   .0000   MA  12:   .9203
   MA  13:  -.3814
 === SUMMARY MEASURES  <V>
 SUM OF SQUARES :     COMPUTED =  .115615E+07  ADJUSTED =  .101236E+07
 VARIANCE ESTIMATES :   BIASED =  6062.01      UNBIASED =  6172.90
 TOTAL  NUMBER  OF  PARAMETERS =  3  STANDARD DEVIATION =  78.5678
 INFORMATION CRITERIA :     AIC =  2111.11          SBIC =  2124.88
 === RESIDUAL ANALYSIS WITH 167 RESIDUALS, BEGINNING AT TIME FEB1984===
 MEAN =  3.87597     ,T-STATISTIC =    .64        (FOR TESTING ZERO MEAN)
 =SIGNIFICANT AUTOCORRELATIONS (USING BARTLETT LIMITS) <A(OR)S>
   1 - 5 %      46:  .2159
 =SIGNIFICANT PARTIAL AUTOCORRELATIONS <P(OR)S>
   1 - 5 %      46:  .1612
 =LJUNG-BOX PORTMANTEAU TEST STATISTICS ON RESIDUAL AUTOCORRELATIONS <L>
  ORDER D.F. STATISTIC  SIGNIFICANCE
  24      21   17.32        .691
 === FORECASTING FROM  DEC1997  WITH FRESH DATA <F>
    DATE   OBSERVATION   FORECAST      ERROR  % ERROR  95% FORECAST INTERVAL
 JAN1998         540.0      500.9       39.1      7.2       372.7      673.3
 …
 JUN1999         211.0      228.2      -17.2      8.2       110.4      471.7
  CUMULATED ERROR          :    1116.6 (= 14.9%); MEAN ERROR:       62.0
  MEAN ABSOLUTE ERROR (MAE):      71.9 (= 17.3%);
  ROOT MEAN SQUARE ERROR   :      90.7 (= 21.8%); MEAN SQUARE ERROR: 8.222E+03
  MEAN ABSOLUTE PERCENTAGE ERROR (MAPE):   16.5% .
```

Figures 64 et 65. Prévisions du modèle du tableau 63 (à gauche) et corrélogramme des résidus (à droite).

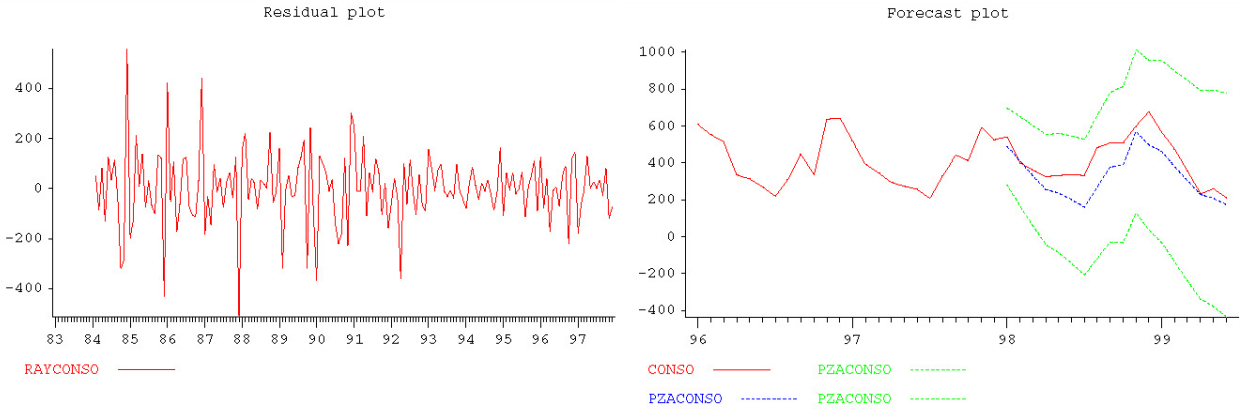


La méthode de Winters n'est ici pas supérieure à la précédente, même si elle intègre une composante saisonnière. Le paramètre GAMMA (correspondant dans la théorie à celui de même nom noté γ) est pratiquement nul et il n'est pas statistiquement significatif. C'est un peu surprenant pour cette série parce qu'un paramètre γ nul signifie une pente constante, alors qu'elle a plutôt l'air négative. La qualité des prévisions à la fois à l'aide des graphiques (figures 61 et 64) est comparable mais à l'avantage de la méthode de Brown. Il est correct de dire que les prévisions de la méthode de Winters sont moins bonnes parce que la courbe des données sort presque de la bande formée par les limites des intervalles de prévision alors que celles de la méthode de Brown sont bien situées dans la bande. Mais c'est dû essentiellement au fait que l'écart-type résiduel de la

méthode de Winters est légèrement plus petit (78,6 contre 81,0) donnant donc des intervalles de prévision proportionnellement plus étroits. La figure 65 présente les autocorrélations de la série résiduelle, c'est-à-dire la série des erreurs de prévision d'horizon 1. Une information à ce sujet est reprise dans la rubrique "Significant autocorrelations" du tableau 63. Mais elle concerne le retard 46 sans aucun intérêt. Notons que la sortie donne la forme ARIMA de la méthode de lissage exponentiel de Winters. Elle a servi en particulier pour déterminer les prévisions et les intervalles de prévision.

C. Un modèle ARIMA (non précisé) est ensuite appliqué aux données brutes. La figure 66 montre la série en différences et différences saisonnières à partir duquel un modèle ARMA est spécifié et estimé. La figure 67 montre le graphique des prévisions accompagnées des intervalles de prévision.

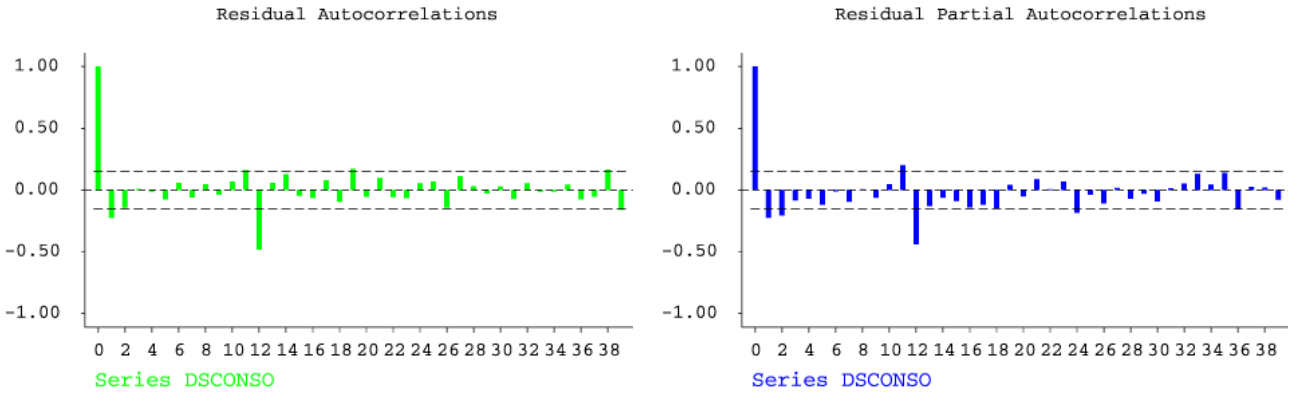
Figures 66 et 67. Résidus d'un modèle non spécifié (à gauche) et prévisions déduites de ce modèle (à droite).



On constate que les intervalles de prévision (obtenues à partir d'un modèle qui n'est pas décrit ici) vont en s'élargissant alors que les données sont de moins en moins dispersées au cours du temps. Comment peut-on l'expliquer ? Un examen des formules qui donnent les prévisions, par exemple (9), montrent que les intervalles de prévision sont de plus en plus larges quand l'horizon augmente. Le fait que la dispersion des données diminue avec le temps n'a pas d'influence. Cela donne simplement une surestimation de l'écart-type résiduel. Ceci est la raison pourquoi la modélisation ARIMA sur les données brutes a été abandonnée au profit d'une modélisation sur les logarithmes des données? C'est confirmé par l'aspect de la figure 4.

D. Un modèle ARIMA est ensuite appliqué aux données en logarithmes. On choisit de prendre des différences et des différences saisonnières. Les figures 68 et 69 contiennent respectivement les autocorrélations et autocorrélations partielles de la série ainsi obtenue.

Figures 68 et 69. Corrélogramme (à gauche) et corrélogramme partiel (à droite) de la série log(CONSO) en différence et différence saisonnière.



Un modèle est alors estimé. Les principaux résultats sont repris dans le tableau 70. L'équation du modèle peut s'écrire

$$\nabla\nabla_{12}\log(\text{CONSO}_t) = e_t - 0{,}81e_{t-12}\,. \qquad (11)$$

Les figures 71 et 72 contiennent respectivement les autocorrélations et autocorrélations partielles des résidus. Le corrélogramme de la série résiduelle est tronqué au-delà du retard 1, tandis que plusieurs autocorrélations partielles sont statistiquement significatives. La série des résidus du modèle peut donc être représentée par un modèle MA(1) qui, combiné avec le modèle SMA(1) de l'équation (11) donne un modèle MA multiplicatif, d'équation

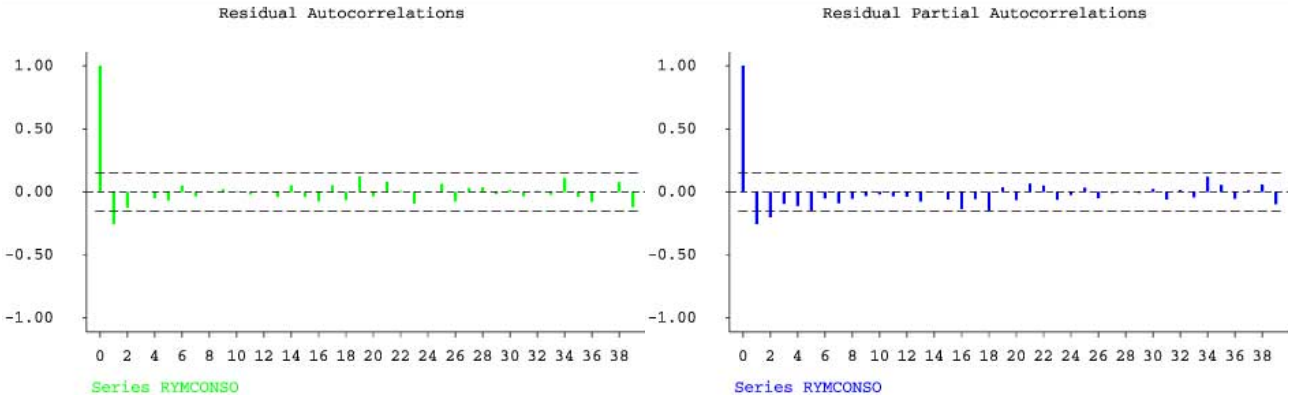
$$\nabla\nabla_{12}\log(\text{CONSO}_t) = (1-\theta_1 B)(1-\Theta_1 B^{12})e_t\,.$$

Les paramètres de ce modèle sont estimés dans le tableau 73.

Tableau 70. Extrait des résultats d'un premier modèle ARIMA.

```
=== MODEL DESCRIPTION                    FORM      DEGREE/ORD PARAMETERS NUMBER
- SEASONAL PERIOD                                  12
- NORMALIZED BOX COX TRANSFORMATION  LOGARITHMS                 BOXC 1      0
- DIFFERENCE                         REGULAR              1
- DIFFERENCE                         SEASONAL             1
- ARMA MODEL
      MOVING AVERAGE POLYNOMIAL      SEASONAL             1     SMA nn      1
NON LINEAR ESTIMATION:
ITER SUM OF SQ SMA  1
 0   2013.6E+03  .000
 …
 6   1288.1E+03  .814
=ITERATION STOPS - RELATIVE CHANGE IN EACH COEFFICIENT LESS THAN  1.00000E-03
FINAL VALUES OF THE PARAMETERS                      WITH 95% CONFIDENCE LIMITS
           NAME      VALUE          STD ERROR   T-VALUE  LOWER          UPPER
     1      SMA  1    .81448       7.17621E-02    11.3     .67          .96
=== SUMMARY MEASURES  <V>
SUM OF SQUARES :     COMPUTED =  .128814E+07  ADJUSTED =  .119139E+07
VARIANCE ESTIMATES :   BIASED =  7134.05      UNBIASED =  7177.03
TOTAL  NUMBER  OF  PARAMETERS =  1  STANDARD DEVIATION =  84.7174
INFORMATION CRITERIA :    AIC =  2126.26          SBIC =  2133.14
=== RESIDUAL ANALYSIS WITH 167 RESIDUALS, BEGINNING AT TIME FEB1984===
MEAN =  1.62024     ,T-STATISTIC =    .25        (FOR TESTING ZERO MEAN)
=SIGNIFICANT AUTOCORRELATIONS (USING BARTLETT LIMITS) <A(OR)S>
 .01-.2 %       1: -.2554
 .2 - 1 %      46:  .2490
=SIGNIFICANT PARTIAL AUTOCORRELATIONS <P(OR)S>
 .01-.2 %       1: -.2554
 .2 - 1 %       2: -.2013
  1 - 5 %      45: -.1623
=LJUNG-BOX PORTMANTEAU TEST STATISTICS ON RESIDUAL AUTOCORRELATIONS <L>
 ORDER D.F. STATISTIC  SIGNIFICANCE
 24     23   24.13        .397
=== FORECASTING FROM  DEC1997  WITH FRESH DATA <F>
   DATE   OBSERVATION   FORECAST      ERROR  % ERROR  95% FORECAST INTERVAL
JAN1998         540.0      497.7       42.3      7.8       361.8      684.7
…
JUN1999         211.0      237.5      -26.5     12.6        56.2     1003.9
 CUMULATED ERROR          :     957.6 (= 12.8%); MEAN ERROR:       53.2
 MEAN ABSOLUTE ERROR (MAE):      66.5 (= 16.0%);
 ROOT MEAN SQUARE ERROR   :      83.6 (= 20.1%); MEAN SQUARE ERROR: 6.986E+03
 MEAN ABSOLUTE PERCENTAGE ERROR (MAPE):   15.6% .
```

Figures 71 et 72. Corrélogramme (à gauche) et corrélogramme partiel (à droite) de la série résiduelle du modèle du tableau 12.



Ce modèle estimé dans le tableau 73 est assez satisfaisant. Les figures 74 et 75 contiennent respectivement les autocorrélations et autocorrélations partielles de la série résiduelle. Toutes les autocorrélations et toutes les autocorrélations partielles sont dans la bande à 95%. Il n'y a donc plus d'information pertinente dans les résidus. On note toutefois quelques points atypiques, peu compatibles avec l'hypothèse de normalité, notamment un résidu anormalement élevé en avril 1991, un autre en août 1989 et un résidu anormalement bas en octobre 1996. Le premier a une

explication simple : la première guerre du Golfe qui a produit une augmentation des prix pétroliers et a donc profité au carburant le moins cher. La figure 76 montre le graphique des prévisions accompagnées des intervalles de prévision.

Tableau 73. Extrait des résultats d'un second modèle ARIMA.

```
=== MODEL DESCRIPTION                    FORM      DEGREE/ORD PARAMETERS NUMBER
- SEASONAL PERIOD                              12
- NORMALIZED BOX COX TRANSFORMATION  LOGARITHMS                BOXC 1      0
- DIFFERENCE                        REGULAR              1
- DIFFERENCE                        SEASONAL             1
- ARMA MODEL
      MOVING AVERAGE POLYNOMIAL     REGULAR              1     MA  nn      1
      MOVING AVERAGE POLYNOMIAL     SEASONAL             1     SMA nn      1
NON LINEAR ESTIMATION:
ITER SUM OF SQ MA   1    SMA  1
 0   2013.6E+03  .000      .000
 …
 9   1155.6E+03  .423      .865
=ITERATION STOPS - RELATIVE CHANGE IN SUM OF SQUARES LESS THAN  1.00000E-06
FINAL VALUES OF THE PARAMETERS                        WITH 95% CONFIDENCE LIMITS
           NAME      VALUE         STD ERROR   T-VALUE  LOWER          UPPER
     1      MA   1    .42277       7.09819E-02     6.0     .28          .56
     2      SMA  1    .86453       8.17244E-02    10.6     .70         1.0
=== SUMMARY MEASURES  <V>
SUM OF SQUARES :     COMPUTED =  .115562E+07  ADJUSTED =  .104658E+07
VARIANCE ESTIMATES :   BIASED =  6266.97      UNBIASED =  6342.93
TOTAL  NUMBER  OF  PARAMETERS =  2  STANDARD DEVIATION =  79.6425
INFORMATION CRITERIA :    AIC =  2108.87          SBIC =  2119.20
=== RESIDUAL ANALYSIS WITH 167 RESIDUALS, BEGINNING AT TIME FEB1984===
MEAN =  3.34433     ,T-STATISTIC =    .54        (FOR TESTING ZERO MEAN)
=OUTLIERS <R(OR)S>
 .01-.2 % APR1991:  292.8
 .2 - 1 % AUG1989:  210.5     OCT1996: -214.6
  1 - 5 % JUL1986:  170.7     DEC1986:  190.5     SEP1989:  182.3
          AUG1990: -169.7     DEC1992: -165.0     DEC1993: -175.6
=SIGNIFICANT AUTOCORRELATIONS (USING BARTLETT LIMITS) <A(OR)S>
  1 - 5 %      46:  .2172
=SIGNIFICANT PARTIAL AUTOCORRELATIONS <P(OR)S>
  1 - 5 %      46:  .1719
=LJUNG-BOX PORTMANTEAU TEST STATISTICS ON RESIDUAL AUTOCORRELATIONS <L>
 ORDER D.F. STATISTIC  SIGNIFICANCE
 24     22   16.92         .768
=== FORECASTING FROM  DEC1997  WITH FRESH DATA <F>
   DATE   OBSERVATION   FORECAST      ERROR  % ERROR  95% FORECAST INTERVAL
JAN1998         540.0      494.6       45.4      8.4       366.5      667.4
…
JUN1999         211.0      232.0      -21.0     10.0       103.1      522.2
 CUMULATED ERROR          :    1055.3 (= 14.1%); MEAN ERROR:       58.6
 MEAN ABSOLUTE ERROR (MAE):      69.7 (= 16.8%);
 ROOT MEAN SQUARE ERROR   :      87.1 (= 21.0%); MEAN SQUARE ERROR: 7.585E+03
 MEAN ABSOLUTE PERCENTAGE ERROR (MAPE):   16.2% .
```

Figures 74 et 75. Corrélogramme (à gauche) et corrélogramme partiel (à droite) de la série résiduelle du modèle du tableau 73.

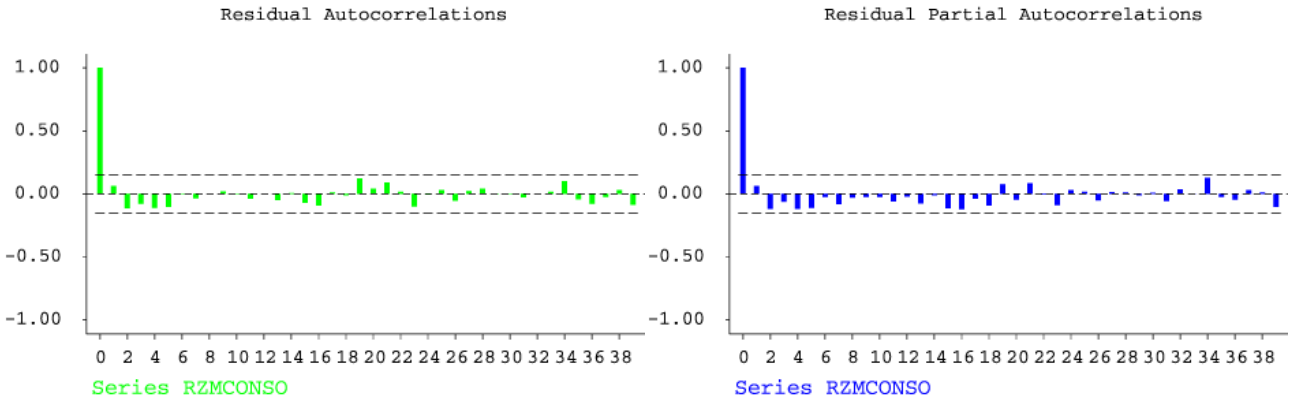
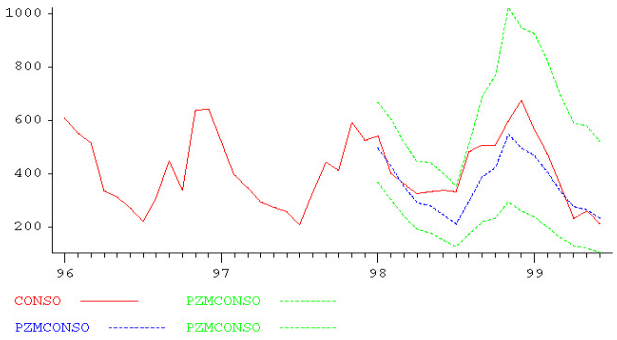


Figure 76. Prévisions déduites du modèle du tableau 73.



E. Le tableau 77 correspond à un modèle amélioré avec une analyse d'intervention à la date d'avril 1991. L'analyse d'intervention utilise des variables de régression et des paramètres de manière à modéliser l'effet d'une intervention. La modélisation la plus simple emploie une variable binaire, appelée K9104, égale à 1 en une date, ici avril 1991 et 0 ailleurs et un paramètre d'intensité noté b_0. Cette intervention est appliquée à la variable CONSO directement. Combiné avec le modèle ARIMA déjà obtenu, l'équation peut être écrite comme suit :

$$\nabla\nabla_{12}\log(\mathrm{CONSO}_t - b_0\mathrm{K9104}_t) = (1-\theta_1 B)(1-\Theta_1 B^{12})e_t\,.$$

Les paramètres de ce modèle sont estimés dans le tableau 77. On remarque que les trois paramètres sont statistiquement différents de 0. L'écart-type résiduel (situé sur la ligne "Standard deviation") est inférieur (77,79 au lieu de 79,65) mais c'est pratiquement toujours le cas et ne peut pas justifier l'intervention. Le critère BIC, plus adéquat dans ce contexte parce qu'il pénalise sévèrement les paramètres excédentaires, montre aussi une diminution (2114 au lieu de 2219).

Tableau 77. Extrait des résultats d'un dernier modèle ARIMA.

```
===KNOWLEDGE ABOUT INTERVENTIONS (BOX-TIAO)
 DIRECTIVE     TYPE          DATE        STEP NATURE     PARAM/VALUE COMMENTS
 I9104         BOX-TIAO      APR1991          VALUE      KI9104:
      1 DIRECTIVE(S),      1 PARAMETER(S),      0 CONSTANT(S).
=== MODEL DESCRIPTION                  FORM      DEGREE/ORD PARAMETERS NUMBER
- SEASONAL PERIOD                                12
- BOX-TIAO INTERVENTION              SEE ABOVE            1     KIdddd      1
- NORMALIZED BOX COX TRANSFORMATION  LOGARITHMS                BOXC 1      0
- DIFFERENCE                        REGULAR              1
- DIFFERENCE                        SEASONAL             1
- ARMA MODEL
      MOVING AVERAGE POLYNOMIAL     REGULAR              1     MA  nn      1
      MOVING AVERAGE POLYNOMIAL     SEASONAL             1     SMA nn      1
NON LINEAR ESTIMATION:
ITER SUM OF SQ MA   1    SMA  1    KI9104
 0   2013.6E+03  .000      .000      .000
…
10   1090.2E+03  .388      .853      218.
=ITERATION STOPS - RELATIVE CHANGE IN SUM OF SQUARES LESS THAN  1.00000E-06
FINAL VALUES OF THE PARAMETERS                    WITH 95% CONFIDENCE LIMITS
            NAME      VALUE         STD ERROR   T-VALUE  LOWER          UPPER
     1      MA   1    .38784       7.27981E-02     5.3     .25          .53
     2      SMA  1    .85297       7.92327E-02    10.8     .70         1.0
     3      KI9104    218.19        57.066         3.8    1.06E+02    3.30E+02
=== SUMMARY MEASURES  <V>
SUM OF SQUARES :     COMPUTED =  .109023E+07  ADJUSTED =  992631.
VARIANCE ESTIMATES :   BIASED =  5943.90      UNBIASED =  6052.63
TOTAL  NUMBER  OF  PARAMETERS =  3  STANDARD DEVIATION =  77.7986
INFORMATION CRITERIA :    AIC =  2100.54          SBIC =  2114.31
=== RESIDUAL ANALYSIS WITH 167 RESIDUALS, BEGINNING AT TIME FEB1984===
MEAN =  3.06368     ,T-STATISTIC =    .51        (FOR TESTING ZERO MEAN)
=OUTLIERS <R(OR)S>
 .2 - 1 % AUG1989:  207.0     OCT1996: -215.0
  1 - 5 % JUL1986:  170.9     DEC1986:  192.2     SEP1989:  174.0
          AUG1990: -169.2     DEC1992: -162.5     DEC1993: -173.8
=SIGNIFICANT AUTOCORRELATIONS (USING BARTLETT LIMITS) <A(OR)S>
  1 - 5 %      46:  .2081
=SIGNIFICANT PARTIAL AUTOCORRELATIONS <P(OR)S>
=LJUNG-BOX PORTMANTEAU TEST STATISTICS ON RESIDUAL AUTOCORRELATIONS <L>
 ORDER D.F. STATISTIC  SIGNIFICANCE
 24     22   20.20        .571
=== FORECASTING FROM  DEC1997  WITH FRESH DATA <F>
   DATE   OBSERVATION   FORECAST      ERROR  % ERROR  95% FORECAST INTERVAL
JAN1998         540.0      495.0       45.0      8.3       369.1      663.8
FEB1998         399.0      427.7      -28.7      7.2       303.2      603.4
MAR1998         360.0      352.2        7.8      2.2       238.9      519.3
APR1998         325.0      284.7       40.3     12.4       185.6      436.6
MAY1998         331.0      279.3       51.7     15.6       175.6      444.2
JUN1998         337.0      246.0       91.0     27.0       149.6      404.6
JUL1998         331.0      210.7      120.3     36.4       124.1      357.5
AUG1998         482.0      298.4      183.6     38.1       170.7      521.8
SEP1998         506.0      388.5      117.5     23.2       216.1      698.7
OCT1998         507.0      421.3       85.7     16.9       228.1      778.3
NOV1998         596.0      550.3       45.7      7.7       290.3     1043.1
DEC1998         674.0      496.1      177.9     26.4       255.3      963.8
JAN1999         563.0      469.1       93.9     16.7       232.8      945.2
FEB1999         472.0      405.4       66.6     14.1       195.3      841.4
MAR1999         354.0      333.8       20.2      5.7       156.3      712.9
APR1999         230.0      269.8      -39.8     17.3       122.9      592.2
MAY1999         258.0      264.7       -6.7      2.6       117.4      596.7
JUN1999         211.0      233.1      -22.1     10.5       100.8      539.2
 CUMULATED ERROR          :    1049.8 (= 14.0%); MEAN ERROR:       58.3
 MEAN ABSOLUTE ERROR (MAE):      69.1 (= 16.6%);
 ROOT MEAN SQUARE ERROR   :      86.4 (= 20.8%); MEAN SQUARE ERROR: 7.462E+03
 MEAN ABSOLUTE PERCENTAGE ERROR (MAPE):   16.0% .
           0 POINTS BELOW THE LOWER LIMIT (TOTAL:   18 POINTS)
           0 POINTS ABOVE THE UPPER LIMIT (TOTAL:   18 POINTS)
```

Si l'on compare les performances des différentes méthodes de prévision employées ici, par exemple au moyen du critère MAPE, on note que le modèle avec intervention (avec MAPE = 16,0 %) est à peine meilleur que le modèle final sans intervention (MAPE = 16,2 %). En ce qui concerne la

comparaison des performances en prévision des différentes méthodes des parties A à E, les résultats se valent et ne sont pas très bons. Visiblement la série de consommation de fuel lourd n'est pas facile à prévoir, en particulier durant l'année 1998 et le premier milieu de 1999. Il serait intéressant d'analyser la série débarrassée des consommations d'EDF.

4 Logiciels, cours et aspects de calcul

4.1 Logiciels

Les grands éditeurs de logiciels statistiques et économétriques proposent des modules d'analyse des séries temporelles : SAS (module SAS/ETS), SPSS (module Trends), Statistica, Statgraphics, Gauss, EViews, TSP, SPlus, etc. Certains de ces éditeurs permettent de tester leur programme pendant une période limitée. Il existe des logiciels spécialisés comme ForecastPro, Autobox de Automatic Forecasting System (AFS ; qui distribue aussi un logiciel gratuit FreeFore) et Scientific Computing Associates (SCA ; particulièrement adapté aux modèles multivariés). Outre FreeFore et le logiciel statistique libre R, quelques logiciels gratuits sont disponibles, notamment Demetra (élaboré par la Commission européenne) qui sert d'interface à X-12-ARIMA (Bureau of the Census) et Tramo/Seats (Banque d'Espagne), Tramo/Seats for Windows (Banque d'Espagne) et Time Series Expert (TSE) employé ici. Le site de Prat propose plusieurs logiciels d'analyse de séries temporelles. Le logiciel E4 est une boîte à outils de Matlab qui peut être téléchargée mais nécessite une clé d'activation à demander aux auteurs. Il est basé sur Terceiro (1990) et couvre des modèles (univariés ou multivariés) qui peuvent être mis sous une forme d'espace état, dont les modèles ARIMA.
D'excellentes sources d'informations pour ce qui concerne les logiciels d'analyse des séries temporelles sont fournies par le site de l'Econometrics Journal On Line mis à jour par Marius Ooms. Le site lié au livre de Armstrong (2001) contient également des informations sur des logiciels ainsi que sur des colloques et des compétitions entre méthodes de prévision.

4.2 Cours

La deuxième édition du livre de l'auteur (voir le site web de Mélard) présente une proposition de cours multimédia qui comporte les chapitres suivants :

1. Concepts et définitions
2. Régression linéaire simple
3. Courbes de croissance
4. Lissage par moyenne mobile
5. Méthodes de décomposition saisonnière
6. Méthodes de lissage exponentiel
7. Régression linéaire multiple
8. Autocorrélation et erreurs de prévision
9. Modèles ARIMA
10. Méthode de Box et Jenkins
11. Régression à erreurs autocorrélées
12. Méthode X-12-ARIMA
13. Méthode TRAMO/SEATS.

Les logiciels utilisés pour les exemples et exercices sont Microsoft Excel (jusqu'au chapitre 8), Time Series Expert et Demetra (pour les chapitres 11 à 13). Il existe d'autres cours en ligne, principalement en anglais. Citons les cours de Richard Weber à Cambridge, de Susan Thomas à l'Indira Gandhi Institute of Development Research à Bombay, de Aad van der Vaart à Amsterdam, de Michael Falk à Würzburg, du CNAM à Paris sous la responsabilité d'Alain Montfort et celui de la société StatSoft qui édite Statistica.

4.3 Aspects de calcul

Les aspects de calcul sont très importants dans l'analyse des séries temporelles, particulièrement dans la modélisation ARIMA. Pour preuve, Newbold *et al.* (1994) ont appliqué plusieurs logiciels sans citer leurs noms sur 5 séries avec des modèles très simples. Ils montrent que les estimations obtenues peuvent être très différentes et les prévisions également. Un facteur important de différence est la méthode d'estimation employée. Certains logiciels proposent encore une des méthodes approchées de la méthode d'estimation par maximum de vraisemblance qui ont été proposées par Box et Jenkins (1976), dont la méthode conditionnelle, alors que les résultats de simulation de Ansley et Newbold (1980) ont montré la supériorié de l'estimation par maximum de vraisemblance exacte. Voir Mélard (1984) pour l'algorithme FLIKAM, utilisé dans Time Series Expert, SPSS et Statistica. Le même algorithme est employé ici pour le lissage exponentiel (voir Broze et Mélard, 1990) et la prévision par moyenne mobile.

Il n'y a pas d'ouvrage récent qui expose les aspects de calcul pour l'analyse des séries temporelles, sauf Pollock (1999). Terceiro (1990) aborde plusieurs sujets pertinents sans rentrer dans les détails de calcul. Les sujets à traiter incluent non seulement l'évaluation de la vraisemblance exacte, mais aussi l'évaluation des autocovariances d'un processus ARMA, l'optimisation, l'estimation de la matrice d'information de Fisher, les procédures de spécification, les particularités dues aux transformations ou aux variances inégales (modèles hétéroscédastiques), etc. Plusieurs articles de vulgarisation mentionnent ces problèmes, notamment Mélard (1989, 1990b). Le cours de statistique d'informatique de Mélard dans son site web ne traite pas explicitement des modèles de séries temporelles mais fournit des éléments introductifs pour les aspects numériques.

Remerciements

Plusieurs des exemples donnés ont déjà eu deux vies, comme projet de stagiaire ou travail d'étudiant (je remercie notamment Mme Baldassarini d'ISTAT pour l'exemple 5 et Matthieu Sadoulet et Peter Basten pour l'exemple 7) et comme question d'examen. Je remercie également Jean-Jacques Droesbeke, ainsi qu'Atika Cohen, Abdelhamid Ouakasse et Hassane Njimi.

Bibliographie sommaire

Livres et articles

ARMSTRONG, S. (Ed.) (2001). *Principle of forecasting: a handbook for researchers and practitioners*, Springer, New York.

ANSLEY, C. F. et NEWBOLD, P. (1980). Finite sample properties of estimates for autoregressive moving average models, *Journal of Econometrics* **13**, 159-183.

ASHLEY, R. (1988). On the relative worth of recent macroeconomic forecasts, *International Journal of Forecasting* **4**, 363-376.

BOURBONNAIS, R. et TERRAZA, M. (2004). *Analyse des séries temporelles*, Dunod, Paris.

BOX, G. E. P. et JENKINS, G. M. (1976). *Time Series Analysis Forecasting and Control*, Holden-Day, San Francisco (édition révisée).

BROZE, L. et MÉLARD, G. (1990). Exponential smoothing: estimation by maximum likelihood, *Journal of Forecasting* **9**, 445-455.

COUTROT, B. et DROESBEKE, J.-J. (1990). *Les méthodes de prévision*, Que Sais-je? n°2157, Presses Universitaires de France, Paris (2e éd.).

DROESBEKE, J.-J., FICHET, B. et TASSI, Ph. (éditeurs) (1994). *Modélisation ARCH - Théorie statistique et applications dans le domaine de la finance*, Collection Association pour la Statistique et ses Utilisations, Editions de l'Université de Bruxelles, Bruxelles et Editions Ellipses, Paris.

GOURIEROUX, C. et MONFORT, A. (1990). *Séries temporelles et modèles dynamiques*, Economica, Paris.

GRANGER, C. W. J. (1980). *Forecasting in Business and Economics*, Academic Press, New York.

HAMILTON, J. (1994). *Time Series Analysis*, Princeton University Press, Princeton.

KADIYALA, K. R. (1970). Testing for the Independence of Regression Disturbances, *Econometrica* **38**, 97-117
LEVIN, R. I. et RUBIN, D. S. (1998). *Statistics for Management*, Prentice Hall, Upper Saddle River (NJ) (7e éd.).
LÜTKEPOHL, H. (1993). *Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin, (2e éd.).
MARTINO, Joseph P. (1983). *Technological Forecasting for Decision Making*, Elsevier, New York.
MAKRIDAKIS, S., ANDERSEN, A., CARBONE, R., FILDES, R., HIBON, M., LEWANDOWSKI, R., NEWTON, J., PARZEN, E. et WINKLER, R. (1984). *The Forecasting Accuracy of Major Time Series Methods*, Wiley, Chichester.
MAKRIDAKIS, S., WHEELWRIGHT, S. S. et HYNDMAN, R. J. (1997). *Forecasting: Methods and Applications*, Wiley, New York (3e éd.).
MÉLARD, G. (1984). Algorithm AS197: A fast algorithm for the exact likelihood of autoregressive-moving average models, *Journal of the Royal Statistical Society Series C Applied Statistics* **33**, 104-114.
MÉLARD, G. (1989). Estimation des paramètres de modèles ARMA, In J.J. DROESBEKE, B. FICHET et Ph. TASSI (éditeurs), *Séries chronologiques - Théorie et pratique des modèles ARMA* (chapitre 4), pp. 75-91.
MÉLARD, G. (1990a). *Méthodes de prévision à court terme*, Editions de l'Université de Bruxelles, Bruxelles, et Editions Ellipses, Paris.
MÉLARD, G. (1990b). Méthodes numériques dans la modélisation de séries chronologiques, *Cahiers du Centre d'Etudes de Recherche Opérationnelle* **32**, 153-180.
NEWBOLD, P., AGIAKLOGLOU, C. et MILLER, J. (1994). Adventures with ARIMA software, *International Journal of Forecasting* **10**, 573-581.
PINDYCK, R. S. et RUBINFELD, D. L. (1976). *Econometric Models and Economic Forecasts*, McGraw-Hill, New York.
POLLOCK, D.S.G. (1999). *A handbook of time-series analysis, signal processing and dynamics*, Academic Press, London.
TERCEIRO, J. (1990). *Estimation of Dynamic Econometric Models with Errors in Variables*, Springer-Verlag, Berlin.

Sites Web (consultés le 11 juillet 2006)
FALK Michael, cours à Würzburg,
http://statistik.mathematik.uni-wuerzburg.de/timeseries/index.php?id=preamble
MÉLARD, Guy, présentation du futur cours multimédia,
http://homepages.ulb.ac.be/~gmelard/Hawaii02.pdf
MÉLARD, Guy, cours de statistique informatique,
http://homepages.ulb.ac.be/~gmelard/statinfo.html
MONTFORT, Alain, cours du CNAM à Paris,
http://dnf2.cnam.fr/offre2005/ue.php?code_formation=STA107&spe_dem=S34&pole_dem=P3&de=list_ues.php.
OOMS, Marius (2005) The Econometrics Journal On Line,
http://www.econ.vu.nl/econometriclinks/software.html.
STATSOFT, cours sur Statistica, http://www.statsoft.com/textbook/stathome.html.
THOMAS, Susan, cours à l'Indira Gandhi Institute of Development Research, Bombay,
http://www.igidr.ac.in/~susant/TEACHING/TSA/.
VAN DER VAART, Aad, cours à la V. U. Amsterdam
http://www.math.vu.nl/stochastics/onderwijs/timeseries/).
WEBER, Richard, cours à Cambridge, http://www.statslab.cam.ac.uk/~rrw1/timeseries/.

Annexes

Annexe A. Détails de l'exemple 2

Tableau A.1. Résultats de la méthode 1.

```
 Decomposition method results :
 Comparison to  the annual means     .  additive
 Series is : PV15MINR
 The first   5 observations are dropped because decomposition is made on whole years samples.
 New sample : 1994.06 - 1998.05
 Seasonal coefficients :
  S( 6) =          0.662
  S( 7) =          1.513
  S( 8) =         -0.238
  S( 9) =          2.013
  S(10) =          0.488
  S(11) =         -1.838
  S(12) =          2.962
  S( 1) =         -0.287
  S( 2) =         -2.813
  S( 3) =         -1.987
  S( 4) =         -0.162
  S( 5) =         -0.313
 Data without seasonal variation are saved on SADJ1.DB   Range : 1994.06 - 1998.05
 Estimated (Trend & Cycle) are saved on TRCY1.DB         Range : 1994.06 - 1998.05
 Residuals are saved on RES1.DB                          Range : 1994.06 - 1998.05
```

Tableau A.2. Résultats de la méthode 2 .

```
 Decomposition method results :
 Comparison to  the linear trend       .  additive
 Series is : PV15MINR
 The first   5 observations are dropped because decomposition is made on whole years samples.
 New sample : 1994.06 - 1998.05
 Regression on annual means. Results :
   Intercept =        14.904  beta =         1.303
 Seasonal coefficients :
  S( 6) =          1.260
  S( 7) =          2.001
  S( 8) =          0.143
  S( 9) =          2.284
  S(10) =          0.650
  S(11) =         -1.783
  S(12) =          2.908
  S( 1) =         -0.450
  S( 2) =         -3.084
  S( 3) =         -2.368
  S( 4) =         -0.651
  S( 5) =         -0.910
 Data without seasonal variation are saved on SADJ2.DB   Range : 1994.06 - 1998.05
 Estimated (Trend & Cycle) are saved on TRCY2.DB         Range : 1994.06 - 1998.05
 Residuals are saved on RES2.DB                          Range : 1994.06 - 1998.05
```

Tableau A.3. Résultats de la méthode 3.

```
 Decomposition method results :
 Comparison to  the moving averages    .  additive
 Series is : PV15MINR
 The first   5 observations are dropped because decomposition is made on whole years samples.
 New sample : 1994.06 - 1998.05
 Center Moving Average of order  12 are computed.
 Moving averages are available on sample 1994.12 - 1997.11
 Seasonal coefficients :
  S( 6) =          1.893
  S( 7) =          1.432
  S( 8) =         -0.193
  S( 9) =          2.122
  S(10) =          0.634
  S(11) =         -1.129
  S(12) =          3.628
  S( 1) =          0.618
  S( 2) =         -2.488
  S( 3) =         -2.838
  S( 4) =         -1.531
  S( 5) =         -2.147
 Data without seasonal variation are saved on SADJ3.DB   Range : 1994.06 - 1998.05
 Estimated (Trend & Cycle) are saved on TRCY3.DB         Range : 1994.12 - 1997.11
 Residuals are saved on RES3.DB                          Range : 1994.12 - 1997.11
```

Tableau A.4. Résultats de la méthode 4.

```
                                    X-11 SEASONAL ADJUSTMENT PROGRAM
              A. PRIOR ADJUSTMENTS, IF ANY
          SERIES TITLE- PV15MINR
                                       PERIOD COVERED-  1/94 TO  5/98

                                  TYPE OF RUN -       ADDITIVE SEASONAL ADJUSTMENT
                           SIGMA LIMITS FOR GRADUATING EXTREME VALUES ARE 1.5 AND 2.5
                 1/94 -  5/98        ADDITIVE SEASONAL ADJUSTMENT      LONG  PRINTOUT
<B 1. ORIGINAL SERIES
<YEAR    JAN      FEB      MAR      APR      MAY      JUN      JUL      AUG      SEP      OCT      NOV      DEC    TOT
```

```
<1994  9.70   10.40   10.10   11.30   11.20   13.30   17.90   15.70   17.50   15.50   11.00   15.20 158.80
<1995 15.60   12.00   14.10   16.50   14.90   19.50   20.00   16.30   24.70   22.60   16.20   27.10 219.50
<1996 18.90   19.00   17.10   17.40   16.30   18.40   18.60   18.30   18.30   17.00   20.30   22.30 221.90
<1997 21.70   16.30   15.40   16.90   18.20   24.10   22.20   21.40   20.20   19.50   17.80   19.90 233.60
<1998 15.30   14.10   18.10   21.20   22.00 ******* ******* ******* ******* ******* ******* *******  90.70
<
<B 2. TREND CYCLE CENTERED 12-TERM  MOVING AVERAGE
<YEAR   JAN     FEB     MAR     APR     MAY     JUN     JUL     AUG     SEP     OCT     NOV     DEC     TOT
<1994 ***** ******* ******* ******* ******* *******   13.48   13.79   14.02   14.41   14.78   15.19  85.67
<1995 15.54   15.65   15.98   16.57   17.08   17.80   18.43   18.86   19.27   19.44   19.53   19.55 213.69
<1996 19.44   19.47   19.28   18.78   18.72   18.69   18.61   18.61   18.43   18.34   18.40   18.71 225.48
<1997 19.10   19.38   19.59   19.77   19.77   19.57   19.20   18.84   18.86   19.15   19.49 ******* 212.73
<
...
<D 7.       TREND CYCLE  - HENDERSON CURVE
                  13-TERM MOVING AVERAGE SELECTED.   I/C RATIO IS  2.04
<   YEAR      JAN     FEB     MAR     APR     MAY     JUN     JUL     AUG     SEP     OCT     NOV     DEC
<   1994     11.43   11.86   12.30   12.75   13.27   13.80   14.14   14.25   14.18   13.98   13.87   14.04
<   1995     14.58   15.39   16.22   16.94   17.54   18.12   18.82   19.53   20.14   20.64   20.90   20.92
<   1996     20.70   20.25   19.62   18.90   18.10   17.35   16.89   16.87   17.24   17.80   18.32   18.62
<   1997     18.73   18.84   19.07   19.47   19.99   20.45   20.58   20.29   19.52   18.45   17.49   17.10
<   1998     17.46   18.48   19.84   21.32   22.80 ******* ******* ******* ******* ******* ******* *******
<D 9. FINAL  REPLACEMENT VALUES FOR EXTREME SI  DIFFERENCES
<   YEAR      JAN     FEB     MAR     APR     MAY     JUN     JUL     AUG     SEP     OCT     NOV     DEC
<   1994   ******* ******* ******* ******* ******* ******* ******* ******* ******* ******* ******* *******
<   1995   ******* ******* ******* ******* ******* ******* *******     .91    2.50 *******    -.99    2.41
<   1996   ******* ******* ******* ******* ******* ******* ******* ******* ******* *******    1.22 *******
<   1997      -.45 ******* ******* ******* ******* ******* ******* ******* ******* ******* ******* *******
<   1998   ******* ******* ******* ******* ******* ******* ******* ******* ******* ******* ******* *******
<D10. FINAL SEASONAL  FACTORS
<   YEAR      JAN     FEB     MAR     APR     MAY     JUN     JUL     AUG     SEP     OCT     NOV     DEC
<   1994      -.81   -2.13   -2.52   -1.37   -2.14    1.07    2.10    1.22    2.04     .91    -.71    2.44
<   1995      -.90   -2.32   -2.53   -1.35   -2.04    1.36    2.04    1.22    1.89     .92    -.60    2.49
<   1996     -1.05   -2.63   -2.48   -1.24   -1.85    1.37    2.04    1.21    1.88     .92    -.59    2.49
<   1997     -1.06   -2.78   -2.53   -1.22   -1.72    1.62    1.81    1.17    1.68     .84    -.30    2.67
<   1998     -1.12   -2.84   -2.59   -1.27   -1.66 ******* ******* ******* ******* ******* ******* *******
<D11. FINAL SEASONALLY ADJUSTED SERIES
<   YEAR      JAN     FEB     MAR     APR     MAY     JUN     JUL     AUG     SEP     OCT     NOV     DEC
<   1994     10.51   12.53   12.62   12.67   13.34   12.23   15.80   14.48   15.46   14.59   11.71   12.76
<   1995     16.50   14.32   16.63   17.85   16.94   18.14   17.96   15.08   22.81   21.68   16.80   24.61
<   1996     19.95   21.63   19.58   18.64   18.15   17.03   16.56   17.09   16.42   16.08   20.89   19.81
<   1997     22.76   19.08   17.93   18.12   19.92   22.48   20.39   20.23   18.52   18.66   18.10   17.23
<   1998     16.42   16.94   20.69   22.47   23.66 ******* ******* ******* ******* ******* ******* *******
   FILE DPV15MI.db    seasonally adjusted)
   FILE TCPV15MI.db   Trend and Cycle
   FILE IPV15MI.db    Irregular Series
   FILE SPV15MI.db    Seasonal factors
<D12. FINAL TREND CYCLE  - HENDERSON CURVE
                  13-TERM MOVING AVERAGE SELECTED.   I/C RATIO IS  2.33
<   YEAR      JAN     FEB     MAR     APR     MAY     JUN     JUL     AUG     SEP     OCT     NOV     DEC
<   1994     11.44   11.87   12.30   12.75   13.31   13.92   14.32   14.44   14.31   14.03   13.85   13.97
<   1995     14.51   15.34   16.20   16.92   17.49   18.07   18.78   19.52   20.17   20.71   21.01   21.01
<   1996     20.76   20.27   19.63   18.89   18.05   17.25   16.74   16.69   17.10   17.77   18.45   18.86
<   1997     19.01   19.03   19.15   19.44   19.89   20.31   20.42   20.09   19.33   18.34   17.53   17.29
<   1998     17.72   18.68   19.91   21.22   22.51 ******* ******* ******* ******* ******* ******* *******
<
<D13. FINAL IRREGULAR SERIES
<   YEAR      JAN     FEB     MAR     APR     MAY     JUN     JUL     AUG     SEP     OCT     NOV     DEC
<   1994      -.94     .67     .32    -.08     .03   -1.69    1.48     .04    1.15     .56   -2.13   -1.21
<   1995     1.99   -1.02     .44     .94    -.55     .07    -.82   -4.43    2.64     .97   -4.21    3.60
<   1996     -.81    1.36    -.05    -.24     .09    -.22    -.18     .40    -.67   -1.69    2.44     .95
<   1997     3.75     .04   -1.22   -1.32     .03    2.17    -.03     .14    -.81     .32     .57    -.06
<   1998    -1.30   -1.74     .78    1.25    1.15 ******* ******* ******* ******* ******* ******* *******
```

Tableau A.5. Résultats pour le lissage exponentiel avec correction saisonnière.

```
 SERIES READ FROM DISK,          NAME IS PV15MINR.DB             ,LENGTH   53
       TIME INTERVAL : FROM JAN1994 TO MAY1998.
   /\/\/\/\/\/\/\/\/\/\
 WARNING *** MODEL FITTING IS PERFORMED WITH ONLY  48 DATA, ENDING
 AT TIME DEC1997.  5 FRESH DATA ARE RESERVED FOR EX-POST VALIDATION
 === FITTING OF THE PARAMETERS OF CLASSIC SHORT-TERM FORECASTING METHODS
     (CO-AUTHORS: O. ROULAND AND L. BROZE)
 EXPONENTIAL SMOOTHING METHOD
   -TREND    COMPONENT : CONSTANT
   -SEASONAL COMPONENT : NONE
  1 PARAMETERS WITH STARTING VALUES :
     1     ALPHA   .50000
 === ESTIMATION BY MAXIMIZATION OF THE EXACT (LOG)LIKELIHOOD
 === MODEL DESCRIPTION                    FORM      DEGREE/ORD PARAMETERS NUMBER
 - SEASONAL PERIOD                              12
 - DIFFERENCE                        REGULAR              1
 - ADDITIVE DETERMINISTIC SEASONALITY AUTOMATIC
 - CLASSIC SHORT TERM FORECASTING     SEE ABOVE
 NON LINEAR ESTIMATION:
 ITER SUM OF SQ ALPHA
  0    230.0       .500
  1    229.7       .461
  2    229.7       .464
  3    229.7       .464
  4    229.7       .464
  5    229.7       .464
  6    229.7       .464
 =ITERATION STOPS - RELATIVE CHANGE IN EACH COEFFICIENT LESS THAN  1.00000E-03
 FINAL VALUES OF THE PARAMETERS                    WITH 95% CONFIDENCE LIMITS
             NAME      VALUE         STD ERROR   T-VALUE  LOWER          UPPER
      1      ALPHA     .46411        .15035        3.1     .16          .77
 THE FOLLOWING PARAMETERS WERE ESTIMATED SEPARATELY
          SEAS 1   -2.9306
          SEAS 2   -3.0972
          SEAS 3   -.36389
```

```
         SEAS 4    1.2694
         SEAS 5   -.59722
         SEAS 6    4.0694
         SEAS 7   -.53056
         SEAS 8   -1.7306
         SEAS 9    2.2694
         SEAS10   -1.4972
         SEAS11   -1.7306
         SEAS12    4.8694
=== SUMMARY MEASURES  <V>
SUM OF SQUARES :     COMPUTED =  229.661      ADJUSTED =  228.013
VARIANCE ESTIMATES :   BIASED =  4.85133      UNBIASED =  6.51465
TOTAL  NUMBER  OF  PARAMETERS = 12  STANDARD DEVIATION =  2.55238
INFORMATION CRITERIA :    AIC =  238.921          SBIC =  263.764
=== RESIDUAL ANALYSIS WITH  47 RESIDUALS, BEGINNING AT TIME FEB1994===
MEAN =  .371896     ,T-STATISTIC =   1.00        (FOR TESTING ZERO MEAN)
=== FORECASTING FROM  DEC1997  WITH FRESH DATA <F>
   DATE   OBSERVATION   FORECAST      ERROR  % ERROR  95% FORECAST INTERVAL
JAN1998         15.30      18.56      -3.26     21.3       13.56      23.57
FEB1998         14.10      15.47      -1.37      9.7        9.95      20.98
MAR1998         18.10      15.10       3.00     16.6        9.12      21.09
APR1998         21.20      16.37       4.83     22.8        9.95      22.79
MAY1998         22.00      15.78       6.22     28.3        8.95      22.60
JUN1998                    19.85                           12.64      27.05
JUL1998                    19.31                           11.74      26.89
AUG1998                    17.58                            9.66      25.51
SEP1998                    19.85                           11.60      28.11
OCT1998                    18.36                            9.78      26.93
NOV1998                    16.63                            7.74      25.51
DEC1998                    21.50                           12.31      30.68
 CUMULATED ERROR          :      9.41 (= 10.4%); MEAN ERROR:       1.88
 MEAN ABSOLUTE ERROR (MAE):      3.74 (= 20.6%);
 ROOT MEAN SQUARE ERROR   :      4.09 (= 22.5%); MEAN SQUARE ERROR:  16.7
 MEAN ABSOLUTE PERCENTAGE ERROR (MAPE):   19.7% .
           0 POINTS BELOW THE LOWER LIMIT (TOTAL:    5 POINTS)
           0 POINTS ABOVE THE UPPER LIMIT (TOTAL:    5 POINTS)
 ---> WRITTEN TO FILE : FOR5.DB , LENGTH =   12
 ---> WRITTEN TO FILE : FOR5.DBM , LENGTH =   12
 ---> WRITTEN TO FILE : FOR5.DBP , LENGTH =   12
```

Annexe B. Détails de l'exemple 3

Tableau B.1. Résultats pour le lissage exponentiel sans correction saisonnière.

```
SERIES READ FROM DISK,         NAME IS DISNEY.DB              ,LENGTH   40
       TIME INTERVAL : FROM  4/1982 TO  3/1992.
  /\/\/\/\/\/\/\/\/\/\
WARNING *** MODEL FITTING IS PERFORMED WITH ONLY  36 DATA, ENDING
AT TIME  3/1991.  4 FRESH DATA ARE RESERVED FOR EX-POST VALIDATION
=== FITTING OF THE PARAMETERS OF CLASSIC SHORT-TERM FORECASTING METHODS
    (CO-AUTHORS: O. ROULAND AND L. BROZE)
EXPONENTIAL SMOOTHING METHOD
  -TREND    COMPONENT : CONSTANT
  -SEASONAL COMPONENT : NONE
 1 PARAMETERS WITH STARTING VALUES :
    1     ALPHA   .50000
=== ESTIMATION BY MAXIMIZATION OF THE EXACT (LOG)LIKELIHOOD
      (FAST ALGORITHM WITH TOLERANCE  1.0E-05)
=== MODEL DESCRIPTION                FORM      DEGREE/ORD PARAMETERS NUMBER
- DIFFERENCE                        REGULAR             1
- CLASSIC SHORT TERM FORECASTING     SEE ABOVE
*** 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 ALPHA
 0   2705.6E+02  .500
 1   2704.4E+02  .519
 2   2704.4E+02  .518
 3   2704.4E+02  .518
 4   2704.4E+02  .518
 5   2704.4E+02  .518
 6   2704.4E+02  .518
=ITERATION STOPS - RELATIVE CHANGE IN EACH COEFFICIENT LESS THAN  1.00000E-03
CORRELATION MATRIX
       ALPHA
ALPHA    1.00
FINAL VALUES OF THE PARAMETERS                      WITH 95% CONFIDENCE LIMITS
           NAME      VALUE         STD ERROR   T-VALUE  LOWER          UPPER
     1     ALPHA     .51791        .15482        3.3      .20          .83
ESTIMATION HAS TAKEN   .0 SEC. FOR  13 EVALUATIONS OF S.S. (MEAN TIME=,  .000)
N.B. QUICK RECURSIONS USED FROM TIME   16
= UNDERLYING SARIMA MODEL
  DEGREE OF REGULAR DIFFERENCING   =   1
  DEGREE OF SEASONAL DIFFERENCING  =   0 (PERIOD  =    4)
  MA   1:   .4821
*** WARNING-A MEAN LEVEL IS NOT INCLUDED IN THE MODEL
THE FOLLOWING CONSTANTS WERE INVOLVED IN THE LEAST SQUARES ESTIMATION METHOD
       ARMA      .99623
=== SUMMARY MEASURES  <V>
SUM OF SQUARES :     COMPUTED =  270441.      ADJUSTED =  268405.
VARIANCE ESTIMATES :   BIASED =  7668.72      UNBIASED =  7894.27
TOTAL  NUMBER  OF  PARAMETERS =  1  STANDARD DEVIATION =  88.8497
INFORMATION CRITERIA :    AIC =  428.566          SBIC =  431.824
=== RESIDUAL ANALYSIS WITH  35 RESIDUALS, BEGINNING AT TIME  1/1983===
MEAN =  30.8317     ,T-STATISTIC =   2.05        (FOR TESTING ZERO MEAN)
=OUTLIERS <R(OR)S>
```

```
  1 - 5 %  4/1990: -181.2
=SIGNIFICANT AUTOCORRELATIONS (USING BARTLETT LIMITS) <A(OR)S>
 .01-.2 %       2: -.5767          4:  .7985
  1 - 5 %       8:  .6817
=SIGNIFICANT PARTIAL AUTOCORRELATIONS <P(OR)S>
 .01-.2 %       2: -.6236          4:  .5579
 .2 - 1 %       3:  .5202
  1 - 5 %       6:  .3334
=LJUNG-BOX PORTMANTEAU TEST STATISTICS ON RESIDUAL AUTOCORRELATIONS <L>
 ORDER D.F. STATISTIC  SIGNIFICANCE
  4      3   38.13        .000
  8      7   65.87        .000
 12     11   86.41        .000
 14     13   91.09        .000
 ---> WRITTEN TO FILE : Resid1.DB , LENGTH =   35
=== FITTING INTERVALS AT THE 95% LEVEL, WITH LEAD TIME   1
           1 POINTS BELOW THE LOWER LIMIT (TOTAL:   34 POINTS)
           0 POINTS ABOVE THE UPPER LIMIT (TOTAL:   34 POINTS)
 ---> WRITTEN TO FILE : Fit1.DB , LENGTH =   34
 ---> WRITTEN TO FILE : Fit1.DBM , LENGTH =   34
 ---> WRITTEN TO FILE : Fit1.DBP , LENGTH =   34
=== FORECASTING FROM   3/1991  WITH FRESH DATA <F>
   DATE   OBSERVATION   FORECAST      ERROR  % ERROR  95% FORECAST INTERVAL
 4/1991         662.4      769.9     -107.5     16.2       595.7      944.0
 1/1992         774.1      769.9        4.2       .5       573.8      966.0
 2/1992         890.5      769.9      120.6     13.5       554.0      985.7
 3/1992         996.2      769.9      226.3     22.7       535.9     1003.8
 CUMULATED ERROR          :     243.7 (=  7.3%); MEAN ERROR:       60.9
 MEAN ABSOLUTE ERROR (MAE):     114.7 (= 13.8%);
 ROOT MEAN SQUARE ERROR   :     139.1 (= 16.7%); MEAN SQUARE ERROR: 1.934E+04
 MEAN ABSOLUTE PERCENTAGE ERROR (MAPE):   13.3% .
           0 POINTS BELOW THE LOWER LIMIT (TOTAL:    4 POINTS)
           0 POINTS ABOVE THE UPPER LIMIT (TOTAL:    4 POINTS)
 ---> WRITTEN TO FILE : Forc1.DB , LENGTH =    4
 ---> WRITTEN TO FILE : Forc1.DBM , LENGTH =    4
 ---> WRITTEN TO FILE : Forc1.DBP , LENGTH =    4
```

Tableau B.2. Résultats pour le lissage exponentiel avec correction saisonnière.

```
SERIES READ FROM DISK,          NAME IS DISNEY.DB              ,LENGTH   40
      TIME INTERVAL : FROM  4/1982 TO  3/1992.
  /\/\/\/\/\/\/\/\/\/\
WARNING *** MODEL FITTING IS PERFORMED WITH ONLY  36 DATA, ENDING
AT TIME  3/1991.  4 FRESH DATA ARE RESERVED FOR EX-POST VALIDATION
=== FITTING OF THE PARAMETERS OF CLASSIC SHORT-TERM FORECASTING METHODS
    (CO-AUTHORS: O. ROULAND AND L. BROZE)
EXPONENTIAL SMOOTHING METHOD
  -TREND    COMPONENT : CONSTANT
  -SEASONAL COMPONENT : NONE
 1 PARAMETERS WITH STARTING VALUES :
    1     ALPHA   .50000
=== ESTIMATION BY MAXIMIZATION OF THE EXACT (LOG)LIKELIHOOD
      (FAST ALGORITHM WITH TOLERANCE  1.0E-05)
=== MODEL DESCRIPTION                  FORM      DEGREE/ORD PARAMETERS NUMBER
- SEASONAL PERIOD                          4
- DIFFERENCE                     REGULAR             1
- ADDITIVE DETERMINISTIC SEASONALITY AUTOMATIC
- CLASSIC SHORT TERM FORECASTING     SEE ABOVE
*** 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 ALPHA
 0   7107.6E+01  .500
 1   4514.0E+01  .884
 2   4061.7E+01  1.17
 3   4025.5E+01  1.26
 4   4024.4E+01  1.28
 5   4024.4E+01  1.28
 6   4024.4E+01  1.28
=ITERATION STOPS - RELATIVE CHANGE IN EACH COEFFICIENT LESS THAN  1.00000E-03
CORRELATION MATRIX
       ALPHA
ALPHA    1.00
FINAL VALUES OF THE PARAMETERS                        WITH 95% CONFIDENCE LIMITS
            NAME      VALUE         STD ERROR   T-VALUE  LOWER          UPPER
     1      ALPHA     1.2800        .18276        7.0      .91          1.7
ESTIMATION HAS TAKEN   .0 SEC. FOR  13 EVALUATIONS OF S.S. (MEAN TIME=,  .000)
N.B. QUICK RECURSIONS USED FROM TIME   10
= UNDERLYING SARIMA MODEL
  DEGREE OF REGULAR DIFFERENCING   =   1
  DEGREE OF SEASONAL DIFFERENCING  =   0 (PERIOD  =    4)
  MA    1:  -.2800
*** WARNING-A MEAN LEVEL IS NOT INCLUDED IN THE MODEL
THE FOLLOWING PARAMETERS WERE ESTIMATED SEPARATELY
         SEAS 1    41.175
         SEAS 2    97.512
         SEAS 3   -.82500
         SEAS 4   -137.86
THE FOLLOWING CONSTANTS WERE INVOLVED IN THE LEAST SQUARES ESTIMATION METHOD
         ARMA      .99883
=== SUMMARY MEASURES  <V>
SUM OF SQUARES :     COMPUTED =  40243.6      ADJUSTED =  40149.8
VARIANCE ESTIMATES :   BIASED =  1147.14      UNBIASED =  1295.15
TOTAL  NUMBER  OF  PARAMETERS =  4  STANDARD DEVIATION =  35.9883
INFORMATION CRITERIA :    AIC =  366.154          SBIC =  374.298
=== RESIDUAL ANALYSIS WITH  35 RESIDUALS, BEGINNING AT TIME  1/1983===
MEAN =  10.8359     ,T-STATISTIC =   1.78        (FOR TESTING ZERO MEAN)
=OUTLIERS <R(OR)S>
=SIGNIFICANT AUTOCORRELATIONS (USING BARTLETT LIMITS) <A(OR)S>
=SIGNIFICANT PARTIAL AUTOCORRELATIONS <P(OR)S>
=LJUNG-BOX PORTMANTEAU TEST STATISTICS ON RESIDUAL AUTOCORRELATIONS <L>
```

```
 ORDER D.F. STATISTIC  SIGNIFICANCE
  4      3    3.27        .352
  8      7    5.16        .640
 12     11    8.64        .655
 14     13   11.14        .599
 ---> WRITTEN TO FILE : Resid2.DB , LENGTH =   35
=== FITTING INTERVALS AT THE 95% LEVEL, WITH LEAD TIME   1
           0 POINTS BELOW THE LOWER LIMIT (TOTAL:   34 POINTS)
           0 POINTS ABOVE THE UPPER LIMIT (TOTAL:   34 POINTS)
 ---> WRITTEN TO FILE : Fit2.DB , LENGTH =   34
 ---> WRITTEN TO FILE : Fit2.DBM , LENGTH =   34
 ---> WRITTEN TO FILE : Fit2.DBP , LENGTH =   34
=== FORECASTING FROM   3/1991  WITH FRESH DATA <F>
   DATE   OBSERVATION   FORECAST      ERROR  % ERROR  95% FORECAST INTERVAL
 4/1991         662.4      688.9      -26.5      4.0       618.4      759.5
 1/1992         774.1      730.1       44.0      5.7       615.5      844.7
 2/1992         890.5      827.6       62.9      7.1       681.7      973.5
 3/1992         996.2      826.8      169.4     17.0       655.2      998.3
 CUMULATED ERROR          :     249.8 (=  7.5%); MEAN ERROR:       62.4
 MEAN ABSOLUTE ERROR (MAE):      75.7 (=  9.1%);
 ROOT MEAN SQUARE ERROR   :      93.9 (= 11.3%); MEAN SQUARE ERROR: 8.824E+03
 MEAN ABSOLUTE PERCENTAGE ERROR (MAPE):    8.4% .
           0 POINTS BELOW THE LOWER LIMIT (TOTAL:    4 POINTS)
           0 POINTS ABOVE THE UPPER LIMIT (TOTAL:    4 POINTS)
 ---> WRITTEN TO FILE : Forc2.DB , LENGTH =    4
 ---> WRITTEN TO FILE : Forc2.DBM , LENGTH =    4
 ---> WRITTEN TO FILE : Forc2.DBP , LENGTH =    4
```

Annexe C. Détails de l'exemple 5

Tableau C.1. Résultats du modèle sur la série PIB.

```
SERIES READ FROM DISK,          NAME IS PIB.DB                  ,LENGTH   48
       TIME INTERVAL : FROM  1/1980 TO  4/1991.
  /\/\/\/\/\/\/\/\/\/\
WARNING *** MODEL FITTING IS PERFORMED WITH ONLY  40 DATA, ENDING
AT TIME  4/1989.  8 FRESH DATA ARE RESERVED FOR EX-POST VALIDATION
=== MODEL DESCRIPTION                    FORM      DEGREE/ORD PARAMETERS NUMBER
- SEASONAL PERIOD                              4
- DIFFERENCE                          REGULAR             1
- DIFFERENCE                          SEASONAL            1
*** WARNING-A MEAN LEVEL IS NOT INCLUDED IN THE MODEL
=== SUMMARY MEASURES  <V>
SUM OF SQUARES :     COMPUTED =  362.310      ADJUSTED =  362.310
VARIANCE ESTIMATES :    BIASED =  10.3517      UNBIASED =  10.3517
TOTAL  NUMBER  OF  PARAMETERS =  0  STANDARD DEVIATION =  3.21741
INFORMATION CRITERIA :    AIC =  209.287          SBIC =  211.217
=== DATA ANALYSIS WITH  35 DATA POINTS BEGINNING AT TIME   2/1981===
MEAN =  .354086     ,T-STATISTIC =    .65        (FOR TESTING ZERO MEAN)
=OUTLIERS <R(OR)S>
  < .01 %  4/1987:  12.95
=SIGNIFICANT AUTOCORRELATIONS (USING BARTLETT LIMITS) <A(OR)S>
=SIGNIFICANT PARTIAL AUTOCORRELATIONS <P(OR)S>
  1 - 5 %       7: -.3795
=LJUNG-BOX PORTMANTEAU TEST STATISTICS ON RESIDUAL AUTOCORRELATIONS <L>
 ORDER D.F. STATISTIC  SIGNIFICANCE
  4      4    3.78        .436
  8      8    9.36        .313
 12     12   12.85        .380
 14     14   13.80        .465
 ---> WRITTEN TO FILE : PIBX.DB , LENGTH =   35
=== FORECASTING FROM   4/1989  WITH FRESH DATA <F>
   DATE   OBSERVATION   FORECAST      ERROR  % ERROR  95% FORECAST INTERVAL
 1/1990         315.12     312.57       2.54       .8       306.27     318.88
 2/1990         321.78     321.85        -.07       .0       312.93     330.77
 3/1990         321.67     321.06        .60        .2       310.14     331.99
 4/1990         353.08     349.03       4.05       1.1       336.41     361.64
 1/1991         341.01     340.33        .67        .2       322.50     358.17
 2/1991         350.48     349.61        .87        .2       327.77     371.46
 3/1991         353.84     348.83       5.01       1.4       323.60     374.05
 4/1991         382.01     376.79       5.22       1.4       348.59     404.99
 CUMULATED ERROR          :     18.91 (=   .7%); MEAN ERROR:       2.36
 MEAN ABSOLUTE ERROR (MAE):      2.38 (=   .7%);
 ROOT MEAN SQUARE ERROR   :      3.10 (=   .9%); MEAN SQUARE ERROR:  9.61
 MEAN ABSOLUTE PERCENTAGE ERROR (MAPE):     .7% .
           0 POINTS BELOW THE LOWER LIMIT (TOTAL:    8 POINTS)
           0 POINTS ABOVE THE UPPER LIMIT (TOTAL:    8 POINTS)
 ---> WRITTEN TO FILE : PIBP.DB , LENGTH =    8
 ---> WRITTEN TO FILE : PIBP.DBM , LENGTH =    8
 ---> WRITTEN TO FILE : PIBP.DBP , LENGTH =    8
```

Annexe D. Détails de l'exemple 6

Tableau D.1. Résultats de spécification pour la série TICD.

```
SERIES READ FROM DISK,          NAME IS TICD.DB                 ,LENGTH   61
       TIME INTERVAL : FROM DEC1974 TO DEC1979.
  /\/\/\/\/\/\/\/\/\/\
=== IDENTIFICATION OF A TIME SERIES MODEL
=== MODEL DESCRIPTION                    FORM      DEGREE/ORD PARAMETERS NUMBER
- DIFFERENCE                          REGULAR             1
=== SUMMARY MEASURES  <V>
SUM OF SQUARES :     COMPUTED =  17.5912      ADJUSTED =  17.5912
VARIANCE ESTIMATES :    BIASED =  .287309      UNBIASED =  .292179
TOTAL  NUMBER  OF  PARAMETERS =  1  STANDARD DEVIATION =  .540535
```

```
INFORMATION CRITERIA :     AIC =  101.098          SBIC =  105.390
=== DATA ANALYSIS WITH  60 DATA POINTS BEGINNING AT TIME  JAN1975===
MEAN =  .766667E-01 ,T-STATISTIC =   1.10        (FOR TESTING ZERO MEAN)
=OUTLIERS <R(OR)S>
 .01-.2 % OCT1979:  1.820
 .2 - 1 % JAN1975: -1.400
  1 - 5 % FEB1975: -.9900     JAN1976: -.9900     NOV1978:  1.310
          SEP1979:  1.150
=SIGNIFICANT AUTOCORRELATIONS (USING BARTLETT LIMITS) <A(OR)S>
 .2 - 1 %       1:  .3989
  1 - 5 %      11:  .3423
=SIGNIFICANT PARTIAL AUTOCORRELATIONS <P(OR)S>
 .2 - 1 %       1:  .3989
 ---> WRITTEN TO FILE : RES.DB , LENGTH =   60
 /\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/
END
```

Tableau D.2. Résultats complets du modèle avec constante sur la série TICD.

```
SERIES READ FROM DISK,         NAME IS TICD.DB                ,LENGTH   61
        TIME INTERVAL : FROM DEC1974 TO DEC1979.
  /\/\/\/\/\/\/\/\/\/\
 1 PARAMETERS WITH STARTING VALUES :
    1     MA   1  .00000
=== ESTIMATION BY MAXIMIZATION OF THE EXACT (LOG)LIKELIHOOD
      (FAST ALGORITHM WITH TOLERANCE  1.0E-05)
=== MODEL DESCRIPTION                 FORM      DEGREE/ORD PARAMETERS NUMBER
- 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
 0    17.24      .000
 1    13.86     -.404
 2    13.72     -.509
 3    13.71     -.494
 4    13.71     -.496
 5    13.71     -.495
 6    13.71     -.495
=ITERATION STOPS - RELATIVE CHANGE IN EACH COEFFICIENT LESS THAN  1.00000E-03
CORRELATION MATRIX
       MA   1
MA   1   1.00
FINAL VALUES OF THE PARAMETERS                    WITH 95% CONFIDENCE LIMITS
             NAME      VALUE         STD ERROR   T-VALUE  LOWER          UPPER
     1        MA   1   -.49532       .11590        -4.3     -.73         -.26
ESTIMATION HAS TAKEN   .0 SEC. FOR  13 EVALUATIONS OF S.S. (MEAN TIME=,  .000)
N.B. QUICK RECURSIONS USED FROM TIME   17
THE FOLLOWING PARAMETERS WERE ESTIMATED SEPARATELY
         MEAN     7.66667E-02
THE FOLLOWING CONSTANTS WERE INVOLVED IN THE LEAST SQUARES ESTIMATION METHOD
         ARMA       .99766
=== SUMMARY MEASURES  <V>
SUM OF SQUARES :     COMPUTED =  13.7127      ADJUSTED =  13.6485
VARIANCE ESTIMATES :   BIASED =  .227475      UNBIASED =  .235319
TOTAL  NUMBER  OF  PARAMETERS =  2  STANDARD DEVIATION =  .485097
INFORMATION CRITERIA :     AIC =  89.1731          SBIC =  95.6113
=== RESIDUAL ANALYSIS WITH  60 RESIDUALS, BEGINNING AT TIME JAN1975===
MEAN = -.872985E-03 ,T-STATISTIC =   -.01        (FOR TESTING ZERO MEAN)
=OUTLIERS <R(OR)S>
 .2 - 1 % JAN1975: -1.323     OCT1979:  1.290
  1 - 5 % NOV1978:  .9776
=SIGNIFICANT AUTOCORRELATIONS (USING BARTLETT LIMITS) <A(OR)S>
  1 - 5 %      11:  .2870
=SIGNIFICANT PARTIAL AUTOCORRELATIONS <P(OR)S>
  1 - 5 %      11:  .2848
=LJUNG-BOX PORTMANTEAU TEST STATISTICS ON RESIDUAL AUTOCORRELATIONS <L>
 ORDER D.F. STATISTIC  SIGNIFICANCE
  6      5     .64        .986
 12     11    8.88        .633
 18     17    9.89        .908
 24     23   11.24        .981
 26     25   11.53        .990
 ---> WRITTEN TO FILE : RESID.DB , LENGTH =   60
=== FITTING INTERVALS AT THE 95% LEVEL, WITH LEAD TIME   1
           0 POINTS BELOW THE LOWER LIMIT (TOTAL:   58 POINTS)
           2 POINTS ABOVE THE UPPER LIMIT (TOTAL:   58 POINTS)
 ---> WRITTEN TO FILE : Fit.DB , LENGTH =   59
 ---> WRITTEN TO FILE : Fit.DBM , LENGTH =   59
 ---> WRITTEN TO FILE : Fit.DBP , LENGTH =   59
=== FORECASTING FROM  DEC1979  WITH FRESH DATA <F>
   DATE    OBSERVATION    FORECAST       ERROR  % ERROR  95% FORECAST INTERVAL
JAN1980                    13.328                          12.377     14.278
FEB1980                    13.404                          11.694     15.115
MAR1980                    13.481                          11.257     15.705
APR1980                    13.558                          10.918     16.197
MAY1980                    13.634                          10.636     16.632
JUN1980                    13.711                          10.393     17.029
JUL1980                    13.788                          10.178     17.397
AUG1980                    13.864                           9.984     17.744
SEP1980                    13.941                           9.809     18.073
OCT1980                    14.018                           9.648     18.387
NOV1980                    14.094                           9.499     18.689
DEC1980                    14.171                           9.361     18.981
 ---> WRITTEN TO FILE : Forc.DB , LENGTH =   12
 ---> WRITTEN TO FILE : Forc.DBM , LENGTH =   12
 ---> WRITTEN TO FILE : Forc.DBP , LENGTH =   12
```

Tableau D.3. Résultats du modèle sans constante pour la série TICD.

```
SERIES READ FROM DISK,         NAME IS TICD.DB                 ,LENGTH   61
        TIME INTERVAL : FROM DEC1974 TO DEC1979.
  /\/\/\/\/\/\/\/\/\/\
 1 PARAMETERS WITH STARTING VALUES :
    1     MA   1  .00000
=== ESTIMATION BY MAXIMIZATION OF THE EXACT (LOG)LIKELIHOOD
      (FAST ALGORITHM WITH TOLERANCE  1.0E-05)
=== MODEL DESCRIPTION                    FORM      DEGREE/ORD PARAMETERS NUMBER
- DIFFERENCE                           REGULAR             1
- 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
 0    17.59     .000
 1    13.94    -.423
 2    13.83    -.514
 3    13.82    -.498
 4    13.82    -.500
 5    13.82    -.500
 6    13.82    -.500
=ITERATION STOPS - RELATIVE CHANGE IN EACH COEFFICIENT LESS THAN  1.00000E-03
CORRELATION MATRIX
       MA   1
MA   1   1.00
FINAL VALUES OF THE PARAMETERS                      WITH 95% CONFIDENCE LIMITS
             NAME      VALUE         STD ERROR   T-VALUE  LOWER        UPPER
     1       MA   1   -.50004       .11453        -4.4    -.73        -.27
ESTIMATION HAS TAKEN   .0 SEC. FOR  13 EVALUATIONS OF S.S. (MEAN TIME=,  .000)
N.B. QUICK RECURSIONS USED FROM TIME   17
*** WARNING-A MEAN LEVEL IS NOT INCLUDED IN THE MODEL
THE FOLLOWING CONSTANTS WERE INVOLVED IN THE LEAST SQUARES ESTIMATION METHOD
         ARMA      .99761
=== SUMMARY MEASURES  <V>
SUM OF SQUARES :     COMPUTED =  13.8249      ADJUSTED =  13.7587
VARIANCE ESTIMATES :   BIASED =  .229312      UNBIASED =  .233199
TOTAL  NUMBER  OF  PARAMETERS =  1  STANDARD DEVIATION =  .482907
INFORMATION CRITERIA :     AIC =  87.6368          SBIC =  91.9290
=== RESIDUAL ANALYSIS WITH  60 RESIDUALS, BEGINNING AT TIME JAN1975===
MEAN =  .504833E-01 ,T-STATISTIC =    .81        (FOR TESTING ZERO MEAN)
=OUTLIERS <R(OR)S>
 .2 - 1 % JAN1975: -1.252     OCT1979:  1.337
  1 - 5 % NOV1978:  1.028     SEP1979:  .9667
=SIGNIFICANT AUTOCORRELATIONS (USING BARTLETT LIMITS) <A(OR)S>
  1 - 5 %       11:  .2965
=SIGNIFICANT PARTIAL AUTOCORRELATIONS <P(OR)S>
  1 - 5 %       11:  .2932
=LJUNG-BOX PORTMANTEAU TEST STATISTICS ON RESIDUAL AUTOCORRELATIONS <L>
 ORDER D.F. STATISTIC  SIGNIFICANCE
  6      5      .49       .992
 12     11     9.20       .604
 18     17    10.27       .892
 24     23    11.55       .977
 26     25    11.95       .987
 ---> WRITTEN TO FILE : RESIDWOC.DB , LENGTH =   60
=== FITTING INTERVALS AT THE 95% LEVEL, WITH LEAD TIME    1
           0 POINTS BELOW THE LOWER LIMIT (TOTAL:   58 POINTS)
           3 POINTS ABOVE THE UPPER LIMIT (TOTAL:   58 POINTS)
 ---> WRITTEN TO FILE : FITWOC.DB , LENGTH =   59
 ---> WRITTEN TO FILE : FITWOC.DBM , LENGTH =   59
 ---> WRITTEN TO FILE : FITWOC.DBP , LENGTH =   59
=== FORECASTING FROM  DEC1979  WITH FRESH DATA <F>
   DATE    OBSERVATION    FORECAST      ERROR  % ERROR  95% FORECAST INTERVAL
JAN1980                    13.277                            12.331     14.224
FEB1980                    13.277                            11.571     14.983
MAR1980                    13.277                            11.057     15.497
APR1980                    13.277                            10.642     15.912
MAY1980                    13.277                            10.284     16.270
JUN1980                    13.277                             9.964     16.590
JUL1980                    13.277                             9.673     16.881
AUG1980                    13.277                             9.403     17.151
SEP1980                    13.277                             9.151     17.403
OCT1980                    13.277                             8.914     17.640
NOV1980                    13.277                             8.689     17.865
DEC1980                    13.277                             8.474     18.080
 ---> WRITTEN TO FILE : FORWOC.DB , LENGTH =   12
 ---> WRITTEN TO FILE : FORWOC.DBM , LENGTH =   12
 ---> WRITTEN TO FILE : FORWOC.DBP , LENGTH =   12
 /\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/\/
END
```

Tableau D.4. Résultats du modèle avec constante sur la série TICD, jusqu'en décembre 1978.

```
SERIES READ FROM DISK,         NAME IS TICD.DB                 ,LENGTH   61
        TIME INTERVAL : FROM DEC1974 TO DEC1979.
  /\/\/\/\/\/\/\/\/\/\
WARNING *** MODEL FITTING IS PERFORMED WITH ONLY  49 DATA, ENDING
AT TIME DEC1978. 12 FRESH DATA ARE RESERVED FOR EX-POST VALIDATION
 1 PARAMETERS WITH STARTING VALUES :
    1     MA   1  .00000
=== ESTIMATION BY MAXIMIZATION OF THE EXACT (LOG)LIKELIHOOD
      (FAST ALGORITHM WITH TOLERANCE  1.0E-05)
=== MODEL DESCRIPTION                    FORM      DEGREE/ORD PARAMETERS NUMBER
- 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
 0    11.53      .000
 1    9.799     -.349
 2    9.682     -.456
 3    9.681     -.459
 4    9.681     -.459
 5    9.681     -.459
 6    9.681     -.459
=ITERATION STOPS - RELATIVE CHANGE IN EACH COEFFICIENT LESS THAN  1.00000E-03
CORRELATION MATRIX
       MA   1
MA   1   1.00
FINAL VALUES OF THE PARAMETERS                        WITH 95% CONFIDENCE LIMITS
            NAME      VALUE         STD ERROR   T-VALUE  LOWER          UPPER
     1      MA   1   -.45937        .13824        -3.3    -.74         -.18
ESTIMATION HAS TAKEN   .0 SEC. FOR  13 EVALUATIONS OF S.S. (MEAN TIME=,  .000)
N.B. QUICK RECURSIONS USED FROM TIME   15
THE FOLLOWING PARAMETERS WERE ESTIMATED SEPARATELY
        MEAN     5.12500E-02
THE FOLLOWING CONSTANTS WERE INVOLVED IN THE LEAST SQUARES ESTIMATION METHOD
        ARMA      .99753
=== SUMMARY MEASURES  <V>
SUM OF SQUARES :     COMPUTED =  9.68149      ADJUSTED =  9.63380
VARIANCE ESTIMATES :   BIASED =  .200704      UNBIASED =  .209430
TOTAL  NUMBER  OF  PARAMETERS =  2  STANDARD DEVIATION =  .457636
INFORMATION CRITERIA :    AIC =  66.7327          SBIC =  72.5264
=== RESIDUAL ANALYSIS WITH  48 RESIDUALS, BEGINNING AT TIME JAN1975===
MEAN = -.133310E-02 ,T-STATISTIC =   -.02        (FOR TESTING ZERO MEAN)
=OUTLIERS <R(OR)S>
 .2 - 1 % JAN1975: -1.319
  1 - 5 % JAN1976: -.9070     NOV1978:  1.001
=SIGNIFICANT AUTOCORRELATIONS (USING BARTLETT LIMITS) <A(OR)S>
=SIGNIFICANT PARTIAL AUTOCORRELATIONS <P(OR)S>
=LJUNG-BOX PORTMANTEAU TEST STATISTICS ON RESIDUAL AUTOCORRELATIONS <L>
 ORDER D.F. STATISTIC  SIGNIFICANCE
  6      5     .76        .980
 12     11    5.14        .924
 14     13    6.01        .946
 ---> WRITTEN TO FILE : RESIDWC8.DB , LENGTH =   48
=== FITTING INTERVALS AT THE 95% LEVEL, WITH LEAD TIME   1
          1 POINTS BELOW THE LOWER LIMIT (TOTAL:   47 POINTS)
          1 POINTS ABOVE THE UPPER LIMIT (TOTAL:   47 POINTS)
 ---> WRITTEN TO FILE : FITWC8.DB , LENGTH =   47
 ---> WRITTEN TO FILE : FITWC8.DBM , LENGTH =   47
 ---> WRITTEN TO FILE : FITWC8.DBP , LENGTH =   47
=== FORECASTING FROM  DEC1978  WITH FRESH DATA <F>
   DATE   OBSERVATION   FORECAST      ERROR  % ERROR  95% FORECAST INTERVAL
JAN1979        11.090     11.156      -.066       .6      10.259     12.053
FEB1979        10.620     11.207      -.587      5.5       9.621     12.794
MAR1979        10.470     11.259      -.789      7.5       9.202     13.316
APR1979        10.340     11.310      -.970      9.4       8.872     13.748
MAY1979        10.440     11.361      -.921      8.8       8.594     14.128
JUN1979         9.980     11.412     -1.432     14.4       8.351     14.474
JUL1979        10.230     11.464     -1.234     12.1       8.134     14.793
AUG1979        10.860     11.515      -.655      6.0       7.937     15.092
SEP1979        12.010     11.566       .444      3.7       7.757     15.376
OCT1979        13.830     11.617      2.213     16.0       7.589     15.645
NOV1979        13.970     11.669      2.301     16.5       7.433     15.904
DEC1979        13.420     11.720      1.700     12.7       7.287     16.153
JAN1980                   11.771                           7.149     16.393
…
DEC1980                   12.335                           5.993     18.676
 CUMULATED ERROR          :      .004 (=   .0%); MEAN ERROR:       .000
 MEAN ABSOLUTE ERROR (MAE):     1.109 (=  9.7%);
 ROOT MEAN SQUARE ERROR   :     1.292 (= 11.3%); MEAN SQUARE ERROR:  1.67
 MEAN ABSOLUTE PERCENTAGE ERROR (MAPE):    9.4% .
          0 POINTS BELOW THE LOWER LIMIT (TOTAL:   12 POINTS)
          0 POINTS ABOVE THE UPPER LIMIT (TOTAL:   12 POINTS)
 ---> WRITTEN TO FILE : FORWC8.DB , LENGTH =   24
 ---> WRITTEN TO FILE : FORWC8.DBM , LENGTH =   24
 ---> WRITTEN TO FILE : FORWC8.DBP , LENGTH =   24
```

Tableau D.5. Résultats du modèle sans constante sur la série TICD, jusqu'en décembre 1978.

```
SERIES READ FROM DISK,         NAME IS TICD.DB                ,LENGTH   61
        TIME INTERVAL : FROM DEC1974 TO DEC1979.
  /\/\/\/\/\/\/\/\/\
WARNING *** MODEL FITTING IS PERFORMED WITH ONLY  49 DATA, ENDING
AT TIME DEC1978. 12 FRESH DATA ARE RESERVED FOR EX-POST VALIDATION
 1 PARAMETERS WITH STARTING VALUES :
    1     MA   1  .00000
=== ESTIMATION BY MAXIMIZATION OF THE EXACT (LOG)LIKELIHOOD
     (FAST ALGORITHM WITH TOLERANCE  1.0E-05)
=== MODEL DESCRIPTION                  FORM      DEGREE/ORD PARAMETERS NUMBER
- DIFFERENCE                         REGULAR           1
- 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
 0    11.66      .000
 1    9.815     -.362
 2    9.710     -.463
 3    9.710     -.465
 4    9.710     -.465
 5    9.710     -.465
 6    9.710     -.465
=ITERATION STOPS - RELATIVE CHANGE IN EACH COEFFICIENT LESS THAN  1.00000E-03
```

```
CORRELATION MATRIX
       MA    1
MA   1   1.00
FINAL VALUES OF THE PARAMETERS                        WITH 95% CONFIDENCE LIMITS
          NAME      VALUE         STD ERROR   T-VALUE  LOWER          UPPER
     1      MA   1   -.46472        .13607       -3.4    -.74         -.19
ESTIMATION HAS TAKEN   .0 SEC. FOR  13 EVALUATIONS OF S.S. (MEAN TIME=,  .000)
N.B. QUICK RECURSIONS USED FROM TIME   15
*** WARNING-A MEAN LEVEL IS NOT INCLUDED IN THE MODEL
THE FOLLOWING CONSTANTS WERE INVOLVED IN THE LEAST SQUARES ESTIMATION METHOD
        ARMA      .99747
=== SUMMARY MEASURES  <V>
SUM OF SQUARES :     COMPUTED =  9.70954      ADJUSTED =  9.66045
VARIANCE ESTIMATES :   BIASED =  .201259      UNBIASED =  .205542
TOTAL  NUMBER  OF  PARAMETERS =  1  STANDARD DEVIATION =  .453367
INFORMATION CRITERIA :    AIC =  64.8328          SBIC =  68.6953
=== RESIDUAL ANALYSIS WITH  48 RESIDUALS, BEGINNING AT TIME JAN1975===
MEAN =  .338027E-01 ,T-STATISTIC =    .52        (FOR TESTING ZERO MEAN)
=OUTLIERS <R(OR)S>
 .2 - 1 % JAN1975: -1.270
  1 - 5 % NOV1978:  1.035
=SIGNIFICANT AUTOCORRELATIONS (USING BARTLETT LIMITS) <A(OR)S>
=SIGNIFICANT PARTIAL AUTOCORRELATIONS <P(OR)S>
=LJUNG-BOX PORTMANTEAU TEST STATISTICS ON RESIDUAL AUTOCORRELATIONS <L>
 ORDER D.F. STATISTIC  SIGNIFICANCE
  6      5     .73        .981
 12     11    5.10        .926
 14     13    6.01        .946
 ---> WRITTEN TO FILE : RESIWOC8.DB , LENGTH =   48
=== FITTING INTERVALS AT THE 95% LEVEL, WITH LEAD TIME   1
           0 POINTS BELOW THE LOWER LIMIT (TOTAL:   47 POINTS)
           1 POINTS ABOVE THE UPPER LIMIT (TOTAL:   47 POINTS)
 ---> WRITTEN TO FILE : FITWOC8.DB , LENGTH =   47
 ---> WRITTEN TO FILE : FITWOC8.DBM , LENGTH =   47
 ---> WRITTEN TO FILE : FITWOC8.DBP , LENGTH =   47
=== FORECASTING FROM  DEC1978  WITH FRESH DATA <F>
   DATE   OBSERVATION   FORECAST      ERROR  % ERROR  95% FORECAST INTERVAL
JAN1979        11.090     11.117       -.027       .2      10.228     12.005
FEB1979        10.620     11.117       -.497      4.7       9.541     12.693
MAR1979        10.470     11.117       -.647      6.2       9.073     13.161
APR1979        10.340     11.117       -.777      7.5       8.694     13.540
MAY1979        10.440     11.117       -.677      6.5       8.366     13.867
JUN1979         9.980     11.117      -1.137     11.4       8.074     14.160
JUL1979        10.230     11.117       -.887      8.7       7.807     14.426
AUG1979        10.860     11.117       -.257      2.4       7.561     14.673
SEP1979        12.010     11.117        .893      7.4       7.330     14.904
OCT1979        13.830     11.117       2.713     19.6       7.112     15.121
NOV1979        13.970     11.117       2.853     20.4       6.906     15.327
DEC1979        13.420     11.117       2.303     17.2       6.710     15.524
JAN1980                   11.117                            6.521     15.712
…
DEC1980                   11.117                            4.812     17.422
 CUMULATED ERROR          :     3.858 (=  2.8%); MEAN ERROR:       .322
 MEAN ABSOLUTE ERROR (MAE):     1.139 (= 10.0%);
 ROOT MEAN SQUARE ERROR   :     1.457 (= 12.7%); MEAN SQUARE ERROR:  2.12
 MEAN ABSOLUTE PERCENTAGE ERROR (MAPE):    9.3% .
           0 POINTS BELOW THE LOWER LIMIT (TOTAL:   12 POINTS)
           0 POINTS ABOVE THE UPPER LIMIT (TOTAL:   12 POINTS)
 ---> WRITTEN TO FILE : FORWOC8.DB , LENGTH =   24
 ---> WRITTEN TO FILE : FORWOC8.DBM , LENGTH =   24
 ---> WRITTEN TO FILE : FORWOC8.DBP , LENGTH =   24
```