

Ce petit document paraphrase une partie de l'aide du module ETS de SAS, en se concentrant sur la procédure ARIMA mise en œuvre en TP. Cependant, l'étude attentive de l'aide de SAS est comme toujours source de beaucoup d'idées, et en fin de document figurent quelques suggestions de lecture dans cette aide. Une bonne connaissance et compréhension du cours de séries chronologiques est bien entendu indispensable à un travail de modélisation digne de ce nom...

## 1 La procédure ARIMA

**La procédure ARIMA sert à mettre en œuvre la méthode de Box & Jenkins**, de modélisation de séries stationnaires par des processus ARMA. Tout comme la procédure REG, c'est *une procédure interactive*: les requêtes peuvent être soumises les unes après les autres, tant que la commande quit n'a pas été soumise.

Il n'existe qu'une seule option de procédure pour ARIMA, en plus de l'option `data=`. Il s'agit de l'option `OUT=`, qui crée une table contenant les valeurs de la série d'origine, les valeurs des prédictions, les écarts-types estimés et bornes inf et sup des IdC de ces prédictions, et les valeurs des résidus, du modèle ajusté (la requête FORECAST de la procédure ARIMA permet aussi de sauvegarder de telles informations; voir le paragraphe sur cette requête pour plus de précisions). D'autres informations peuvent être sauvegardées dans d'autres tables de sortie, mais comme options des requêtes d'ARIMA.

Il y a 4 étapes dans la méthode Box & Jenkins, correspondant à 3 requêtes SAS de la procédure ARIMA.

### Identification

Elle consiste à observer la nature des fonctions d'autocorrélation, d'autocorrélation partielle, et d'autocorrélation inverse (toutes empiriques), de la série elle-même, pour chercher à identifier *des propositions de modélisation ARMA* en s'aidant de sa propre expertise personnelle, ou d'aides à l'identification (requêtes ESACF, MINIC, SCAN), qu'il faut bien entendu être en mesure de savoir interpréter (l'aide de SAS contient un tout petit laïus sur chacune des méthodes). C'est pendant cette étape que l'on peut différencier la série, *afin d'aboutir à une série stationnaire que le reste de la procédure sera en droit de traiter* : pour ce faire, on peut avoir recours à des tests de stationnarité.

Il ne faut pas oublier que la visualisation directes des données (graphe en lignes brisées de  $(t, X_t)_{t \geq 0}$ ) est indispensable, et potentielle source de suggestion quant au type de non-stationnarité que l'on va devoir traiter. Il est toujours mauvais de se reposer uniquement sur les critères chiffrés que fournit le logiciel, et que l'on ne maîtrise pas forcément, quand d'éventuelles suggestions graphiques sont aussi disponibles<sup>1</sup>.

La syntaxe standard d'une requête IDENTIFY de la procédure ARIMA est la suivante

```
IDENTIFY VAR=nomvar(...) [ NLAG=n STATIONARITY=(ADF ou PP ou RW) OUTCOV=table_sortie
                          SCAN / MINIC / ESACF P=... Q=... CENTER ALPHA=... ] ;
```

- seul le champ VAR= est obligatoire, il permet tout simplement d'indiquer quelle est la variable de la table SAS qui est à étudier en tant que série. On peut **différencier** une fois cette variable (*i.e.* étudier  $Y_n = X_n - X_{n-1}$  au lieu de  $X_n$ , càd  $Y = \nabla X$ ) en écrivant

```
IDENTIFY VAR=nomvar(1) ... ;
```

On peut différencier deux fois de suite (*i.e.*  $Y = \nabla^2 X$ ) en écrivant IDENTIFY VAR=*nomvar*(1,1), et aussi différencier à plus longue distance, par exemple passer de  $X_n$  à  $Y_n = X_n - X_{n-12}$  en soumettant IDENTIFY VAR=*nomvar*(12), puis encore à  $Z_n = Y_n - Y_{n-1}$  avec VAR=*nomvar*(12,1), etc...

Il n'est pas possible de récupérer facilement les nouvelles données ainsi obtenues dans une table SAS au sein de la procédure ARIMA; si l'on veut se faire une idée graphiquement de la stationnarité de la série différenciée, il faut la générer soi-même en utilisant la fonction dif (voir plus loin le paragraphe *Macros et fonctions utiles*).

- l'option NLAG=*n* stipule qu'on ne veut voir s'afficher que les *n* premiers coefficients d'autocorrélation (simples, partielles, et inverses); par défaut,  $n = \min\{24, nb/4\}$  où *nb* est le nombre de données.

<sup>1</sup>tout comme en régression linéaire simple il ne faut pas se contenter de regarder si le coefficient de corrélation est proche de 1 ou non, mais aussi visualiser le nuage des données.

- l’option `STATIONARITY=` permet de mettre en œuvre un certain test de stationnarité de la série étudiée (ce test étant appliqué après éventuelle différenciation bien sûr). Les tests disponibles sont ceux de Dickey & Fuller “augmenté” (ADF ou `DICKEY`), de Phillips (PP ou `PHILLIPS`), et de la marche aléatoire avec drift (RW ou `RANDOMWALK`), en n’omettant pas les parenthèses (par exemple `identify var=X nlag=8 stationarity=(DICKEY)`).
- l’option `OUTCOV=` permet de sauvegarder dans une table SAS (dont on fournit le nom) les estimations des autocovariances, autocorrélations (simples, partielles, et inverses), et surtout (information non fournie dans l’affichage de la procédure `ARIMA`) les écarts-types estimés des coefficients d’autocorrélation<sup>2</sup>.  
Les observations de cette table sont indexées par le numéro du *lag* : la première observation contient donc les coefficients à *lag* 1, la seconde à *lag* 2, etc
- l’option `ALPHA=` indique le niveau du test de bruit blanc figurant en fin d’étape d’identification.
- l’option `CENTER` fait centrer la série initiale par la moyenne empirique.
- pour les autres options, se référer à la documentation SAS; notamment dans la section *The ARIMA Procedure* → *Details*, figurent une description succincte et des références sur les méthodes *Smallest Canonical Correlation Method (SCAN)*, *Minimisation de Critères d’information, AIC et BIC (MINIC)*, et *Extended Sample Autocorrelation Function (ESACF)*. Celles-ci sont une source de suggestions de valeurs de paramètres de modèles ARMA, à soumettre à l’estimation dans l’étape suivante.

## Estimation (et validation)

Une fois les présomptions faites sur le modèle, on peut proposer des modèles ARMA particuliers et demander l’estimation de leurs coefficients. Cette étape n’est pas indépendante de la précédente (car SAS a besoin de savoir quelle est la variable étudiée, dont le nom est fourni en paramètre de `var=` de la requête `identify`).

**Il faut soumettre une étape ESTIMATE par modèle que l’on souhaite essayer.** Le résultat inclut au minimum les informations suivantes : estimateurs des paramètres du modèle soumis et *t*-tests associés (tests de leur nullité<sup>3</sup>), estimation de la variance des résidus, test de bruit blanc de Ljung-Box pour la série des résidus obtenue (variante de la statistique de Box-Pierce), valeurs des critères d’information AIC et BIC pour le modèle soumis (BIC est noté SBC, pour *Schwartz Bayesian Criterion*).

La syntaxe de base est très simple, mais peuvent s’y greffer de nombreuses options, notamment de contrôle du processus itératif d’estimation des paramètres, et bien entendu de sauvegarde des résultats dans des tables SAS en sortie :

```
ESTIMATE p=... q=... PLOT [ METHOD=... NOCONSTANT GRID OUTEST=... OUTSTAT=... ]
```

Spécifier le modèle à estimer s’écrit comme suit : “`p=1`” correspond à un AR(1), “`q=1`” à un MA(1), “`p=2 q=3`” à un ARMA(2,3), “`p=(2 3 4)`” à un AR(4) avec coefficient nul aux *lags* 1 et 3 (*i.e.*  $\Phi(B)X_t = Z_t$  où  $\Phi(z) = 1 - \phi_2 z^2 - \phi_4 z^4$ ), “`p=(1) (2)`” à un AR(2) de la forme  $\Phi(B)X_t = Z_t$  où  $\Phi(z) = (1 - \phi_1 z)(1 - \phi'_1 z - \phi'_2 z^2)$ , et ainsi de suite... Ici, on entend que la série soit centrée, ce qui est toujours le cas après différenciation ou usage de l’option `NOCONSTANT`.

D’autres options de la requête `ESTIMATE` sont les suivantes :

- l’option `METHOD=` peut prendre 3 valeurs, correspondant à 3 méthodes d’estimation des paramètres : `ML` signifie *Maximum Likelihood*, `ULS` signifie *Unconditional Least-Squares*, et la valeur par défaut est `CLS`, *Conditional Least-Squares*. Il n’est pas question de détailler ici ce en quoi chacune des ces méthodes consiste, on peut trouver quelques explications (et beaucoup d’autres informations) dans la page de l’aide *Details* → *Estimation Details*.
- l’option `PLOT` fait afficher les graphes en bâton des fonctions d’autocorrélation (simple, partielle, et inverse) non pas de la série elle-même comme dans l’étape d’identification, mais de la série des résidus obtenue à partir du modèle considéré et estimé.
- l’option `NOCONSTANT` impose un modèle sans intercepte (*i.e.* une série centrée). Un synonyme est parfois `NOINT` (suivant les versions).

<sup>2</sup>la documentation de SAS n’est pas très explicite à ce sujet: “*The standard error estimate is based on the hypothesis that the process generating the time series is a pure moving-average process of order LAG-1*” (dixit paragraphe *Details* → *The OUTCOV= data set*)

<sup>3</sup>il faut prendre garde à l’interprétation des *p*-valeurs de ces tests de nullité des coefficients du modèle, car leur validité est conditionnée par le fait que le modèle proposé n’est pas forcément le bon : comme ces *p*-valeurs sont calculées sur la base de ce modèle, il faut prendre garde à ne pas conclure trop vite en face d’une *p*-valeur faible

- l’option `GRID` fait afficher une ou plusieurs grilles  $3 \times 3$  indiquant la valeur de la SCR (SSE en anglais, somme des carrés des résidus) en la valeur estimée des paramètres, et dans un voisinage de cette valeur. La SCR est la quantité qui entre en compte dans les critères de choix d’ordre type AIC et BIC. Il va de soi que si le nombre de paramètres à estimer est  $> 3$ , non seulement ces grilles deviennent difficiles à lire mais leur affichage devient volumineux. On peut paramétrer cette option avec l’option `GRIDVAL=`.
- l’option `OUTEST=nom` crée une table en sortie contenant notamment les valeurs des estimations des paramètres et de leurs écarts-types (sous l’hypothèse de validité du modèle, cf note 3 de bas de page). On peut également inclure dans cette table les corrélations ou covariances empiriques de ces estimateurs (càd le contenu de l’onglet *Correlations of Parameter Estimates* de la sortie de la requête `ESTIMATE`) à l’aide des options supplémentaires `OUTCORR` et `OUTCOV` (à ne pas confondre avec l’option `OUTCOV=` de la requête `IDENTIFY`).
- l’option `OUTSTAT=nom` crée une table en sortie contenant les résultats des diagnostics de modélisation tels que les valeurs des critères d’information AIC et BIC. Voir les détails au paragraphe de l’aide *Details* → *The OUTSTAT= data set*.
- D’autres options existent, et sont essentiellement de deux types. Dans un premier type, on peut imposer des valeurs aux paramètres du modèle étudié, qui ne seront donc pas estimés, ou alors proposer des valeurs de départ pour ces paramètres, dans le cadre du processus itératif d’estimation. Le second type concerne le contrôle de ce processus itératif : précision des calculs, nombre maximum d’itérations, critère d’arrêt.

## Prévision (forecasting)

La requête `FORECAST` va permettre de calculer les prévisions (fourchettes de confiance comprises) pour les valeurs de la série, en lui appliquant **le dernier modèle** qui a été soumis à une requête `ESTIMATE`. Les prévisions sont directement celles pour la série *initiale*, autrement dit, si par exemple il y a eu différenciation, le logiciel calcule les prévisions pour la série différenciée puis repasse *automatiquement* à la série initiale par simples additions : c’est ce qui figure dans la table en sortie de l’option `OUT=` de la procédure `ARIMA`. On notera que cette table ne contiendra aucune observation si aucune requête `FORECAST` n’est soumise.

La manière de calculer les prévisions diffère suivant la méthode d’estimation sélectionnée.

Les principales options de la requête `FORECAST` sont les suivantes :

- l’option `LEAD=k` indique que l’on demande le calcul des prévisions jusqu’à distance (*lag*)  $k$  après la dernière donnée de la série. Par défaut  $k = 24$ .
- l’option `OUT=nom` va sauvegarder les données et les prévisions dans la table *nom*.
- l’option `ALPHA=` permet de fixer le niveau définissant les intervalles de prévision (*forecast limiting values*); soit dit en passant, il ne faut pas parler d’intervalles de confiance dans ce cadre, car on n’estime pas des prédictions, ce ne sont pas des paramètres statistiques !
- l’option `ID=` permet d’ajouter à la table en paramètre de l’option `OUT=`, la variable qui fait office de temps dans la table initiale, c’est-à-dire qui date les observations. La présence de cette variable dans cette table des prévisions est indispensable si l’on veut pouvoir tracer la courbe des prévisions en fonction du temps.
- l’option `NOPRINT` annule l’affichage des résultats de cette étape, tout en maintenant bien entendu le calcul des prévisions et leur éventuelle sauvegarde via l’option `OUT=`.

## 2 L'analyse des séries (sans saisonnalité) en bref

1. Etudier la non-stationnarité de la série initiale à l'aide :
  - ◇ du graphe en lignes brisées des données (on cherchera à visualiser une tendance ou des changements de régime)
  - ◇ du diagramme en bâtons de la fonction d'autocorrélation empirique (la non-stationnarité se traduit souvent par une décroissance très lente de cette fonction)
  - ◇ du test de Dickey & Fuller, d'hypothèse nulle "existence d'une racine unité (modèle non-causal)" (attention : le mot "stationary" signifie parfois "causal" en anglais, ce qui porte à confusion).

Si l'un de ces critères suggère la non-stationnarité, différencier une fois la série (en tapant `identify var=var(1)`) et répéter cette étape. Si nécessaire, différencier de nouveau.

*Attention* ceci ne concerne que les séries sans composante saisonnière, voir en bas de page pour des préliminaires dans le cas contraire.

2. Observer le résultat du test de bruit blanc indiqué en fin d'étape d'identification : si la statistique du  $\chi^2$  ainsi calculée est faible, et que les premiers coefficients d'autocorrélation le sont aussi, il est raisonnable de penser que la série initiale (ou la série différenciée suivant le cas) n'est qu'un bruit blanc, ce qui met un terme à l'analyse !
3. Visualiser les fonctions d'autocorrélation, simple et partielle, pour tenter d'identifier le modèle : en vertu des propriétés établies en cours, si l'ACF s'"écrase" après le rang  $q$ , cela suggère une composante MA( $q$ ). Si la PACF ou IACF s'"écrase" après le rang  $p$ , cela suggère une composante AR( $p$ ). On notera que si la série initiale a été différenciée, il est possible qu'une composante MA(1) ou MA(2) soit alors présente. Consulter le cours de statistique des séries chronologiques pour des justifications de ces "principes".
4. Il est aussi possible, ensuite, de recourir aux méthodes SCAN, MINIC, et ESACF pour obtenir des suggestions de valeurs pour les paramètres des modèles ARMA qu'on tentera d'ajuster.
5. Par des requêtes `estimate`, soumettre plusieurs modèles ARMA. Utiliser l'option `NOCONSTANT` si la série étudiée est centrée, ou à été différenciée (cela supprime l'estimation de l'intercepte, qui est considéré comme nul).
6. Pour chacun de ces modèles, noter la valeur de la statistique du test de Ljung-Box qui teste l'hypothèse de bruit blanc sur la série des résidus. Si elle a une  $p$ -valeur élevée, le modèle peut être conservé pour continuer l'analyse. Sinon, retourner aux étapes précédentes ou passer à un autre modèle. On peut aussi visualiser le graphe des résidus en fonction du temps, mais ce n'est pas indispensable dans un premier temps.  
  
Il faut prendre garde à ne pas sur-paramétrer en même temps la composante MA et la composante AR. Considérer un AR pur d'ordre élevé, ou bien un MA pur d'ordre élevé, ne pose par contre pas de problème de stabilité dans l'estimation.
7. Utiliser la requête `forecast` avec options `back=k` et `lead=m` pour calculer les prévisions des rangs allant de  $n - k$  à  $n + m$  (l'option `back=` est un outil de "validation", qui n'a d'intérêt que pour visualiser si le modèle choisi produit des prévisions proches de la série de départ).
8. Examiner le graphe des résidus en appliquant éventuellement à cette série des résidus (sauvegardée dans la table en sortie de l'option de procédure `out=`) la procédure `univariate` ou `spectra` afin de tester une dernière fois l'hypothèse de normalité ou de bruit blanc des résidus.

### Traitement des saisonnalités avec proc ARIMA

Pour se débarrasser d'une composante saisonnière, il faut différencier de manière adéquate au sein de la requête `identify`. Par exemple, si l'on pense qu'il y a une périodicité annuelle et que l'on dispose de données trimestrielles, il faudra différencier en passant de  $X_n$  à  $Y_n = X_n - X_{n-4}$  à l'aide de la requête

```
identify var=variable(4)
```

et pour passer directement de  $X_n$  à  $Z_n = Y_n - Y_{n-1}$ , ce sera

```
identify var=variable(4,1)
```

alors que `identify var=variable(1,4)` fera passer de  $X_n$  à  $Z_n = Y_n - Y_{n-4}$  où  $Y_n = X_n - X_{n-1}$ .

## 3 Quelques mots sur le module SAS/ETS

### Aide à l'utilisation du module ETS

Il est toujours bon de rappeler qu'un excellent moyen d'améliorer la connaissance d'un logiciel en l'absence d'un tuteur, est de lire le *manuel de référence qui est livré avec*. Celui-ci est toujours accompagné de nombreux exemples, d'un paragraphe de prise en main, et celui de SAS est de ce point de vue là plutôt bien fichu. En ce qui concerne le module SAS/ETS, si une utilisation un tant soit peu sérieuse doit en être faite, dans le cadre d'un stage par exemple, il est plus que conseillé de lire le chapitre suivant de l'aide de SAS :

*SAS/ETS* → *General Information* → *Working with Time Series Data*

Le lecteur curieux et soucieux de ne pas peiner lors de l'utilisation des procédures ayant trait à l'analyse de séries chronologiques, y trouvera des informations pratiques très utiles concernant les sujets suivants : rappels de gestion des dates sous SAS, gestion des valeurs manquantes dans des séries temporelles (interpolation à l'aide de la procédure EXPAND), graphiques utiles pour l'analyse (série + prévisions et fourchettes de prévision), fonctions lag et dif (cf ci-dessous), transformation log et Box-Cox des séries, rappels sur la fusion et la transposition de tables SAS.

### Macros et fonctions utiles

Le module SAS/ETS est fourni avec une poignée de macros, dont la liste et la description se trouve dans l'onglet de l'aide *SAS/ETS* → *General Information* → *SAS Macros and Functions*. On y trouve notamment des macros permettant de : voir quelle est la "meilleure" transformation Box-Cox applicable à une série donnée (%BOXCOXAR), tester la pertinence de modéliser le log d'une série plutôt que la série elle-même (%LOGTEST), ou encore réaliser le test et calculer la  $p$ -valeur du test de Dickey-Fuller d'existence d'une racine unité (%DFTEST et %DFPVALUE).

Par ailleurs, la fonction dif permet de calculer les valeurs  $\nabla Y_n = Y_n - Y_{n-1}$  lorsque  $(Y_n)$  est une variable d'une table SAS. Par exemple, pour visualiser le graphe de la série différenciée obtenue à partir d'une table SAS don contenant la série Y et la variable temporelle date, on écrit

```
data don2; set don; dY=dif(Y);  
proc gplot data=don2; plot dY*date; run;
```

Dans le même esprit, la fonction lag permet d'accéder à la valeur précédente d'une variable d'une table SAS. Mais il faut prendre garde à bien comprendre ce que ces fonctions font réellement; les détails et mises en garde se trouve à l'onglet de l'aide *SAS/ETS* → *General Information* → *Working with Time Series Data* → *Lags, Leads, Differences, and Summations*; dans cet onglet on trouvera également d'autres conseils, et des rappels concernant les sommes implicites (requêtes de type "var+1;" dans une étape DATA).

### Quelques autres procédures du module ETS

#### a. La procédure AUTOREG

Les résultats (IdC, tests,...) du modèle de régression linéaire classique ( $Y_n = \sum_{j=1}^p \theta_j X_n^{(j)} + \epsilon_n$ ) reposent en grande partie sur le fait que la suite  $(\epsilon_n)$  des innovations est un bruit blanc gaussien. Si cette suite est autocorrélée, ou que la variance de  $\epsilon_n$  dépend de l'état à l'instant  $n$  (i.e. l'innovation est  $\sigma(X_n)\epsilon_n$  au lieu de  $\epsilon_n$ , modèle *hétéroscédastique*), alors tous les résultats théoriques (lois approchées) sont fortement affectés, ils ne sont plus fiables.

La procédure AUTOREG réalise estimations et prévisions pour des modèles de régression linéaire dans le cadre de séries chronologiques dont les innovations sont autocorrélées ou bien hétéroscédastiques. Dans le premier cas, on peut modéliser des erreurs AR de tout ordre, voire même utiliser une procédure de sélection *stepwise* pour sélectionner l'ordre "automatiquement"; le diagnostic de l'autocorrélation repose essentiellement sur le test de Durbin-Watson. Dans le second cas, la procédure permet de modéliser par un modèle GARCH (*generalized autoregressive conditional heteroscedasticity*) ou une de ses variantes, et de tenir compte de cette hétéroscédasticité.

#### b. La procédure FORECAST

La procédure FORECAST permet de générer rapidement et de manière "automatique" des prévisions pour séries chronologiques, en appliquant des méthodologies assez classiques de type lissage, mais sans différenciation, pour enlever la tendance et un éventuel facteur saisonnier.

La première méthode est celle du *lissage exponentiel* (option `method=expo trend=n`), permettant de traiter les tendances constantes ( $n = 1$ , lissage simple), affines ( $n = 2$ , lissage double) ou quadratiques ( $n = 3$ , lissage triple). Les coefficients de l'opérateur de lissage sont de la forme  $w(1-w)^k$ , où  $w$  est le

paramètre de lissage à indiquer lors du lancement de la procédure (option `weight=w`). Par défaut, ce paramètre de lissage est pris égal à  $w = 1 - 0.8^n$  où  $n$  est la valeur de l'option `trend=` (qui vaut 3 par défaut dans ce cas). Plus  $w$  est faible, plus le lissage est sévère (*i.e.* prend en compte un plus grand nombre de valeurs environnantes).

La seconde méthode est celle de Holt-Winters, permettant de gérer des modèles à composante saisonnière de type multiplicatif (`method=winters`) ou additif (`method=addwinters`), c'ad respectivement de la forme

$$X_t = (b_0 + b_1 t)s(t) + \epsilon_t \quad \text{ou} \quad X_t = b_0 + b_1 t + s(t) + \epsilon_t.$$

Il faut alors spécifier la nature de la saisonnalité avec l'option `seasons=`, des valeurs possibles étant `qtr` (trimestriel), `qtr2` (semestriel), `month` (mensuel), `day`, `hour`, `hour2`, etc. On peut aussi utiliser la syntaxe plus simple `seasons=k`, ce qui signifie qu'une période correspond à  $k$  valeurs successives dans la série des données ( $k = 12$  correspondant par exemple au cas d'une périodicité annuelle avec des données mensuelles).

Enfin, la méthode par défaut est `method=stepar` (pour *stepwise autoregressive*) : elle permet de gérer les modèles avec tendance polynômiale et erreurs AR, avec sélection automatique de l'ordre de l'AR suivant une procédure *stepwise*. Pour cette méthode, la valeur par défaut de `trend` est 2.

Les options `out=` et `outest=` permettent de sauvegarder un certain nombre de résultats en sortie de procédure, en particulier les prévisions elles-mêmes, les estimations des coefficients de la tendance et de la composante saisonnière, et d'autres statistiques encore, certaines devant être spécifiées à la main car non produites par défaut. On notera que la procédure `forecast` ne produit aucune sortie à l'écran (dans la fenêtre *Sortie/Output* de SAS). Voir le manuel pour plus de détails sur cette procédure.

*Remarque 1* : cette procédure `forecast` n'est pas la panacée, car elle ne permet pas de gérer n'importe quel modèle pour estimer la tendance ou la composante saisonnière (par exemple les tendances quadratiques ne sont pas autorisées pour le modèle saisonnier géré par la méthode `winters`).

*Remarque 2* : contrairement à la procédure ARIMA par exemple, elle ne permet pas de prendre en compte d'autres variables explicatives, et utilise des méthodes d'extrapolation, de sorte que certains résultats statistiques (tels que la validité des intervalles de prévision) ne sont valables que de manière approchée; par contre, ces mêmes raccourcis calculatoires, source de ce défaut, permettent d'obtenir des résultats avec un temps de calcul moindre qu'avec les procédures classiques (ce qui ne peut se ressentir que sur des séries volumineuses).

### c. La procédure SPECTRA

La procédure SPECTRA réalise l'analyse spectrale des séries chronologiques univariées, et l'analyse spectrale croisée pour des séries chronologiques multivariées. Les techniques d'analyse spectrale permettent d'étudier les périodicités ou motifs cycliques dans de telles données. La procédure permet d'estimer la densité spectrale, en calculant le périodogramme et en le lissant. Elle permet aussi de tester si la série considérée est issue d'un bruit blanc (option de procédure `whitetest`; cf onglet de l'aide SAS/ETS → *Procedure Reference* → *The SPECTRA Procedure* → *Details* → *White Noise Test*).

### d. Autres procédures d'intérêt

Parmi les autres procédures du module SAS/ETS, on peut aussi parler de STATESPACE (modèle *state-space*, analyse de séries chronologiques multivariées, c'est-à-dire analyse simultanée de plusieurs séries scalaires sur la même échelle de temps), VARMAX (modèles VARMAX, Vector Autoregressive Moving-Average processes with exogenous regressors), MDC (modèle logit conditionnel et associés), et MODEL (estimation, simulation, et prévision, pour modèles définis par des équations non-linéaires).

## Le Système de Prévision “*Time Series Forecasting System*”

Il s'agit d'une application SAS, avec fenêtre unique, menus, boutons à cliquer et tout le toutim, qui permet de réaliser de manière assez conviviale des études de séries chronologiques (notamment l'“après étude” : sauvegarde, présentation, impression, catalogage...), et inclut une partie de ce que les principales procédures de SAS/ETS peuvent réaliser. Mais il ne faut pas se réjouir trop vite, car les résultats que cette application va fournir, *il va falloir les interpréter* de la même manière que par le procédé classique : **il s'agit essentiellement d'un emballage des procédures existantes**, même si l'application apporte son lot d'outils (par exemple un système de sélection automatique de modèle, qui vaut ce qu'il vaut mais est configurable). Les capacités de cette application ne se limitent pas à la seule méthode de Box & Jenkins<sup>4</sup>, elle traite aussi des méthodes de lissage exponentiel classiques, de la méthode de Winters...

<sup>4</sup>il est intéressant de noter qu'il est possible de visualiser les codes SAS des procédures ARIMA générées par cette application, ce qui est très utile en phase d'apprentissage.

Sous Windows, ce module est accessible via l'onglet

*Solutions* → *Analyse* → *Application de Prévision de Séries Temporelles*

de la barre de menu de la fenêtre principale de SAS. Une fois lancée, cette application bénéficie elle aussi d'un menu, dans la barre de menu principale (comme les autres fenêtres de SAS). Le manuel de référence sur cette application se trouve dans l'aide de SAS/ETS (onglet *The Time Series Forecasting System*), où tous les détails sont fournis, notamment un chapitre de prise en main (sous-onglet *Getting Started with Time Series Forecasting*).

(*J. Worms, UVSQ, Février 2007*)