ECOLE NATIONALE DE LA STATISTIQUE ET DE L'ADMINISTRATION ECONOMIQUE

Deuxième année
2005-2005

Séries temporelles linéaires
Corrigé des travaux dirigés n°5

Guillaume Lacôte
Bureau **E03**
✉ Guillaume.Lacote@ensae.fr
☞ http://ensae.no-ip.com/SE206/

Version du 20050121-11h02, révisée le 17 février 2005

Corrigé de l'exercice 1

☞ Q1 (a) On suppose ici que le fichier `Donnees1.sd2` est enregistré dans le dossier `W :\Sas`.
Le programme commencera donc par

```
LIBNAME Td_SAS 'W :\Sas\' ;
DATA table ;
SET Td_SAS.donnees1 ;
```

à la suite de quoi les données contenues dans la table `table` enregistrée dans le fichier `W :\Sas\donnees1.sd2` sont accessibles sous le nom `table`.

(b) L'affichage graphique se fait au moyen de la `PROC GPLOT` de la façon suivante :

```
DATA table ;
SET Td_SAS.donnees1 ;
time = _N_ /* Permet de nommer explicitement l'axe des abscisses */ ;
PROC GLPOT ;
PLOT XM * time /* Trace XM en fonction du temps */ ;
SYMBOL I=JOIN ;
RUN ;
```

On observe que la série semble périodique, de période 12.

On définit en conséquence la série désaisonnalisée `DeSaison` au moyen de la fonction retard `LAG` de la façon suivante : [1]

```
DATA table ;
SET Td_SAS.donnees1 ;
DeSaison = XM - LAG12(XM) /* LAG[n](Y)(t) = Y(t-n) */ ;
```

(c) *Publi-information :*

Auto-corrélation	AR(p)	MA(q)	ARMA(p,q)
Directe $\rho(h)$	décroît exponentiellement vers 0	nulle à partir de $q+1$	décroît exponentiellement vers 0
Partielle $r(h)$	nulle à partir de $p+1$	(?)	nulle à partir de $p+1$
Inverse $\rho^i(h)$	nulle à partir de $p+1$	décroît exponentiellement vers 0	décroît exponentiellement vers 0

[1]Un façon habituelle de désaisonnaliser serait d'étudier la série moyennée sur une période $M_{12}XM = (\mathbb{1} + L + \cdots + L^{11})XM$ (voir TD 1, exercice 2) ; cependant sous réserve que l'on parvienne à modéliser $DeSaison = (\mathbb{1} - L^{12})XM$ sous forme ARMA le modèle final en `XM` sera plus simple, et c'est la démarche retenue ici. Cependant le reste de l'étude pourrait porter sur $M_{12}XM$.

Les auto-corrélogrammes direct, partiel et inverse peuvent être visualisés au moyen l'option `IDENTIFY` de la `PROC ARIMA` , de la façon suivante :

```
PROC ARIMA ;
IDENTIFY VAR=DeSaison NLAG=50 /* 0 <= h <= 50 */ ;
RUN ;
```

L'option `NLAG` permet de spécifier le nombre d'auto-covariances (inverses) à calculer.

On observe que les auto-corrélations inverses tendent (exponentiellement) vers zéro ; la série `DeSaison` est donc apparamment stationnaire. Pour s'en assurer, on définit `DesInt` la série de ses différences premières de la façon suivante :

```
DATA table ;
SET Td_SAS.donnees1 ;
DeSaison = XM - LAG12(XM) /* LAG[n](Y)(t) = Y(t-n) */ ;
DesInt = DeSaison - LAG(DeSaison) /* LAG = LAG1 */ ;
```

que l'on étudie à nouveau :

```
PROC ARIMA ;
IDENTIFY VAR=DesInt NLAG=50 /* 0 <= h <= 50 */ ;
RUN ;
```

On observe que les auto-corrélations inverses ne décroissent **pas** exponentiellement (mais plutôt linéairement) vers zéro, ce qui suggère que la série `DesInt` est **sur-différenciée**, ce qui corrobore la stationnarité de `DeSaison` envisagée auparavant.

On se propose donc d'estimer un modèle ARMA pour la série `DeSaison` . On sait qu'une série Y suivant un modèle ARMA(p,q) est telle que son auto-corrélation partielle est nulle à partir de l'ordre $p+1$ et son auto-corrélation directe est non-significative à partir de l'odre $q+1$. On recherche donc les premiers ordres au-delà desquels les auto-corrélations (respectivement partielles et directes) sont toujours en-deçà du fractile à 95% de la loi normale (à savoir 1.96), ce qui conduit à proposer pour la série `DeSaison` des ordres [2]

$$\boxed{d^* = 0,\ p^* = 3 \text{ et } q^* = 2}$$

(d) On cherche tout d'abord à estimer le modèle le plus général ARIMA(3,0,2) pur pour `DeSaison` , au moyen de l'option `ESTIMATE` de la `PROC ARIMA` :

[2] Cette identification préliminaire ne sert qu'à éviter d'estimer ensuite un modèle dont la plupart des cœfficients seront non-significativement non-nuls. En toute rigueur, mieux vaudrait retenir les ordres plus conservatoires $p^* = 7$ et $q^* = 7$ quitte à s'assurer ensuite *par un test statistique* que les cœfficients au-delà de $3, 2$ sont non-significativement non-nuls. Par souci de simplicité on se place directement dans l'hypothèse ou ces tests auraient déjà été effectués.

```
PROC ARIMA ;
IDENTIFY VAR=DesInt ;
ESTIMATE METHOD = ML PLOT P=3 Q=2 ;
RUN ;
```

L'option `METHOD=ML` sélectionne la méthode d'estimation du maximum de vraisemblance ; l'option `PLOT` permet d'obtenir les auto-corrélogrammes des résidus estimés, afin de contrôler la validité du modèle.

Il faut en premier lieu s'assurer que le modèle estimé est bien un ARMA, ce qui revient à s'assurer que le résidu estimé est compatible avec l'hypothèse de bruit blanc : SAS pratique à cet effet le test de Porte-Manteau dont le résultat est indiqué dans la section `Autocorrelation Check for White Noise`.[3]

Dans le cas présent le test de Porte-Manteau conduit à accepter l'hypothèse selon laquelle le résidu est un bruit blanc, de sorte que la modélisation ARIMA(3,0,2) est **statistiquement** légitime. En outre, les tests individuels de nullitié des cœfficients du modèle sont tous rejetés au seuil 5%, de sorte que le modèle estimé est également valide.

(e) Sachant que le modèle ARIMA(3,0,2) est valide, on cherche enfin $p \leq 3$ et $q \leq 2$ tels que le modèle ARIMA(p,0,q) soit valide. Pour ce faire on estime successivement

– un modèle AR :
On cherche à modéliser `DeSaison` par un modèle AR pur, dont on sait que l'ordre éventuel est au plus 3. On calcule donc

```
PROC ARIMA ;
IDENTIFY VAR=DesInt ;
ESTIMATE METHOD = ML PLOT P=3 /* Implicitement, Q=0 */ ;
RUN ;
```

L'hypothèse selon laquelle les résidus ainsi estimés suivent un bruit blanc est acceptée, donc on peut légitimement admettre que `DeSaison` suit un modèle AR d'ordre au plus 3. Cependant le test individuel de nullité du cœfficient associé au troisième retard est **rejeté** au seuil 5%.
On réestime donc un modèle AR(2) :

```
PROC ARIMA ;
IDENTIFY VAR=DesInt ;
ESTIMATE METHOD = ML PLOT P=2 ;
RUN ;
```

dont le résidu estimé est toujours un bruit blanc (c'est heureux !), mais dont le second retard n'est toujours pas significativement non-nul.

[3] La colonne `Prob` donne la $p-value$ associée : l'hypothèse que la variable est significativement non-nulle est acceptée à $1 - pvalue$ %. Par exemple une valeur de 0.023 indique que la variable est significativement non-nulle "à 97%", tandis qu'une valeur de 0.452 laisse entendre que la variable n'est pas significative à 54%.

On estime donc finalement un modèle AR(1) :

```
PROC ARIMA ;
IDENTIFY VAR=DesInt ;
ESTIMATE METHOD = ML PLOT P=1 ;
RUN ;
```

qui est toujours valide et dont tous les cœfficients sont, cette fois, significativement non-nuls.
Ainsi la série `DeSaison` peut légitimement être modélisée par un modèle AR(1).

– un modèle MA :
On cherche à modéliser `DeSaison` par un modèle MA pur, dont on sait que l'ordre éventuel est au plus 2. On calcule donc

```
40 PROC ARIMA ;
IDENTIFY VAR=DesInt ;
ESTIMATE METHOD = ML PLOT Q=2 ;
RUN ;
```

L'hypothèse selon laquelle les résidus ainsi estimés suivent un bruit blanc est acceptée, donc on peut légitimement admettre que `DeSaison` suit un modèle MA d'ordre au plus 2 ; en outre tous les cœfficients sont significativement non-nuls (au seuil 5%).
Ainsi la série `DeSaison` peut légitimement être modélisée par un modèle MA(2).

En définitive, la série `DeSaison` peut être modélisée par trois modèles statistiquement valides, à savoir ARIMA(3,0,2), ARIMA(1,0,0) et ARIMA(0,0,2).

Le critère de parcimonie, selon lequel un modèle comprenant strictement moins de cœfficients est toujours préférable, conduit nénanmoins à rejeter le modèle ARIMA(3,0,2) qui est un sur-modèle strict à la fois du modèle AR(1) et du modèle MA(2).

Pour arbitrer enfin entre ces deux derniers modèles, on peut recourir à un critère informationnel, qui met en rapport la vraisemblance du modèle estimé et le nombre de paramètres nécessaires pour l'estimer. Le critère d'Akaike (`AIC`) arbitre en l'occurence en faveur du modèle MA(2), que l'on rentiendra finalement.

(f) En conclusion, on peut raisonnablement proposer pour la série d'origine le modèle

$$\boxed{(\mathbb{1} - L^{12})(\mathtt{XM} - 0,377) = (\mathbb{1} - 0,286L + 0,231L^2)\epsilon}$$

où ϵ est un bruit blanc de variance $\sigma_\epsilon^2 \simeq 0,233$.

☞ Q2

☞ Q3

⋆

⋆ ⋆