

## TP3 : modélisation ARMA

Le but de ce TP est d'apprendre à modéliser des séries temporelles par un processus *ARMA* grâce à la procédure *arima*. On étudiera la série des ventes de vin australien disponibles dans un tableau nommé *vin* (cf. premier cours).

### 1 Description

1. Afficher la série ou une partie de la série. A quels intervalles de temps la série est-elle observée ? La décrire succinctement grâce à la proc *means* et grâce à *Analyse/Analyse interactive des données* ...
2. Représenter le chronogramme de la série grâce à l'afficheur de séries temporelles. Présente-t-elle une composante saisonnière ? Une tendance ? Est-elle homoscedastique, à première vue ?
3. Observer le graphe des autocorrélations proposé par l'afficheur de séries temporelles. La série est-elle un bruit blanc ?

### 2 Stabilisation de la variance et élimination de la composante saisonnière

1. Appliquer la fonction *log*, puis *sqrt*, à la série et observer les chronogrammes respectifs. Choisir l'une des transformations en justifiant ce choix. On pourra par la suite étudier l'autre série transformée pour comparer.
2. Appliquer l'opérateur de différentiation avec une période adaptée pour éliminer la composante saisonnière. Afficher le chronogramme. La série obtenue est-elle un bruit blanc ? En examinant les diverses fonctions d'autocorrélation, vers quel(s) modèle(s) peut-on s'orienter ?
3. Que représentent les droites en rouge sur le graphique des corrélations ?

### 3 Modélisation ARMA de la série différenciée

#### 3.1 Identification

1. Appliquer une étape *identify* sur la variable *ventvin* en faisant afficher le tableau des critères d'information BIC grâce à l'option *minic*.
2. Quel est le modèle ARMA retenu à cette étape ?

#### 3.2 Estimation

1. Appliquer une étape d'estimation selon le modèle  $ARMA(p, q)$  retenu à l'étape précédente.
2. Quel est le modèle ARMA\* retenu ? (penser à tenir compte de la moyenne du processus)

\*La notation du modèle ARMA peut varier d'une source à une autre. SAS utilise une convention précisée dans l'aide de SAS : saisir le mot-clé "arima notation", puis sélectionner dans la liste le thème "Notation for ARIMA models". A noter, ce paragraphe concerne les processus  $ARIMA(p, d, q)$ . Un processus  $ARMA(p, q)$  est un processus  $ARIMA(p, 0, q)$ .

### 3.3 Validation du modèle

1. Observer la significativité des paramètres. La sortie de la `proc arima` fournit les p-valeurs d'un test, quelle est l'hypothèse nulle de ce test ? Dans le cas présent, quelle est la conclusion ?
2. On s'intéresse à la blancheur du résidu. Obtenir les résidus grâce à la commande `forecast` et afficher leur chronogramme : ont-ils l'apparence d'un bruit blanc ? Quelle est leur variance ? La sortie SAS présente les p-valeurs obtenues par un test de blancheur proche du test de Portmanteau, le test de Ljung-Box\*. Quelle est la conclusion ?
3. On rappelle que les critères AIC (Akaike information criterion) et SBC (Schwarz information criterion), basés sur une vraisemblance approchée et le nombre de paramètres du modèle, permettent de comparer différents modèles entre eux. Plus le critère est petit, meilleur est le modèle. Ici, prendre note de ces deux critères.

\*Aller consulter l'aide concernant le test de blancheur : dans l'aide de SAS, saisir le mot-clé "arima estimation details", puis sélectionner le titre : "Estimation details" dans la liste. Puis lire le paragraphe "Tests of residuals".

### 3.4 Comparaison avec d'autres modèles

1. Reprendre le tableau en sortie de l'étape d'identification. Quels autres modèles semble-t-il judicieux de tester ?
2. Appliquer la procédure d'estimation pour ces modèles et comparer les résultats. Comparer notamment les critères AIC et SBC.
3. En conclusion, quel est le modèle qu'on peut retenir ? Est-ce satisfaisant ? Peut-on proposer d'autres méthodes pour parvenir à une solution plus satisfaisante ?