Séries temporelles
TP4 : Processus Sarima

Données

Le fichier dataST-TP5.sav et/ou dataST-TP5.txt, qui vous a été envoyé, représente la masse salariale trimestrielle de 1990 à 2010 (soit 84 semestres) des employés (MSE) d'une maison d'assurance. L'objectif est de prédire la masse salariale des quatres trimestres de la dernière année, en utilisant les données des années 1990 à 2009.

Partie A : modélisation

Dans cette partie, on cherche à modéliser la série observée pour les 80 premiers trimestres (1990 à 2009).

1. Importer le jeu de données, représenter graphiquement les 80 premiers trimestres de la série, et donner une interprétation qualitative. (Tendance, saisonnalité,. . .).
2. Rappeler la définition théorique d'une série stationnaire du second ordre. La série étudiée est-elle stationnaire ? Pourquoi ? Dites comment peut-on la rendre stationnaire ? (il ne s'agit pas de faire de calculs mais d'expliquer votre démarche).
3. Dresser le tableau théorique d'identification d'un modèle ARMA à l'aide des fonctions ACF et PACF.
4. Désaisonnaliser la série, représenter graphiquement la tendance, la composante saisonnière et l'erreur. Analyser les graphes.
5. Tracer les fonctions d'auto-corrélations (ACF) et d'auto-corrélations partielles (PACF) empiriques de la série stationnarisée (Erreur issue de la désaisonnalisation). En déduire qu'il s'agit d'un $ARIMA(3, 0, 1)(0, 0, 0)$ (voir Annexe pour une définition de cette notation).
6. Estimer les paramètres de ce modèle et tester si les résidus obtenus sont indépendants et identiquement distribués.
7. On suppose maintenant qu'il s'agit d'un modèle de type $ARIMA(0, 0, 1)(1, 0, 1)$. Estimer les paramètres de ce modèle et comparer avec les résultats avec la question précédente. Proposer des critères pour choisir entre les deux modèles.
8. Estimer avec une régression linéaire (on utilisera la colonne année comme variable explicative) les paramètres de la tendance et analyser la pertinence de vos résultats.
9. Modélisation de la MSE

(a) D'après les résultats précédents, on propose de modéliser la MSE avec un modèle saisonnier $ARIMA(0,1,1)(1,0,1)$. Quelle est la valeur de la saisonnalité ? Quels sont les coefficients saisonniers ? Estimer les paramètres de ce modèle.

(b) On suppose qu'il s'agit maintenant d'un modèle saisonnier $ARIMA(0,1,1)(0,1,1)$. Estimer les paramètres du modèle et comparer au précédent.

(c) Modéliser la série par un lissage multiplicatif de Winters. Estimer et analyser les paramètres associés.

Partie B : Prévision

L'objectif est de prédire la masse salariale des quatre trimestres de la dernière année, en utilisant les données des années 1990 à 2009.

1. Effectuer la prévision à l'aide du lissage multiplicatif de Winters.
2. Effectuer la prévision à l'aide du modèle saisonnier $ARIMA(0,1,1)(0,1,1)$.
3. Représenter alors sur un même graphique la série complète ainsi que les prédictions de la dernière année par lissage exponentiel et $SARIMA$.
4. Comment comparer les deux résultats.
5. Conclure.

ANNEXE TECHNIQUE

Exemple : ARIMA(3,0,1)(0,0,0) signifie que la partie non saisonnière (3,0,1) comporte trois coefficients AR et un coefficient MA et la partie saisonnier (0,0,0) ne comporte ni de coefficient AR, ni de différenciation, ni de coefficient MA. Par exemple un ARIMA(0,0,1)(1,0,1) veut dire que la partie non saisonnier (0,0,1) comporte un coefficient MA et la partie saisonnier (1,0,1) comporte un coefficient AR et un coefficient MA.