

Sujet de TP n° 3

Étude et prévision de la série de trafic aérien – comparaison entre différentes approches

1. Méthodologie de Box & Jenkins

Dérouler la méthodologie de Box & Jenkins à l'aide du script fourni et retrouver dans un premier temps le modèle SARIMA suivant pour la série $\{x_t\}$ de trafic aérien :

$$\log(X_t) = (I + \theta_1 B)(I + \theta_{12} B^{12})Z_t, \quad Z_t \sim WN(0, \sigma^2).$$

Ce qui équivaut à :

$$\log(X_t) \sim SARIMA(p=0, d=1, q=1)(P=0, D=1, Q=1)_{s=12}.$$

Dans un second temps, on teste ce modèle pour faire de la prévision : on enlève les 12 dernières valeurs que l'on cherche ensuite à prédire (back-testing). Obtenir la figure (« Travail à faire » dans le script) qui montre, pour la série initiale $\{x_t\}$ (et non pas pour son log), la qualité de la prévision obtenue par back-testing. Calculer l'erreur de prévision ou RMSE (Root Mean Square Error) associée.

2. Modèle MA(13)

Essayer un modèle MA($q = 13$) au lieu de SARIMA(0, 1, 1)(0, 1, 1)₁₂ pour la série $\log(X_t)$. Conclusion.

3. Modèle ARMAX

On considère maintenant un modèle linéaire pour rendre compte de la tendance et saisonnalité, ce qui conduit à un modèle dit ARMAX (Auto Regressive Moving Average with eXogenous inputs) de la forme :

$$y_t = \log(x_t) = \beta_0 + \beta_1 t + \beta_2 \cos\left(\frac{2\pi t}{12}\right) + \beta_3 \sin\left(\frac{2\pi t}{12}\right) + w_t,$$

avec le résidu $\{w_t\}$ de type $SARMA(p, q)_{s=12}$.

En pratique, on procède en deux temps :

- on estime les coefficients de régression par la méthode des moindres carrés (fonction `lm()` de R),
- puis on identifie la structure de dépendance des résidus à l'aide d'un modèle $SARMA(p, q)_{s=12}$.

Dans un deuxième temps, on tient compte du fait que les résidus ne sont pas un bruit blanc pour ré-estimer les paramètres de régression en tenant compte de la structure de dépendance identifiée (fonction `arima()` de R avec l'option `xreg=` pour spécifier les prédicteurs exogènes).

Obtenir la figure qui montre pour la série initiale $\{x_t\}$ la qualité de la prévision obtenue par back-testing sur les 12 dernières valeurs. Calculer le RMSE et comparer avec la première approche.

4. Méthode de lissage exponentiel

Faire le même travail de prévision par une méthode de lissage exponentiel (voir support de cours pages 16 à 20 et fonction `HoltWinters()` de R). Calculer l'erreur RMSE.

5. Améliorations

Essayer d'améliorer l'une ou l'autre de ces approches. Spécifier le modèle pour le meilleur score RMSE obtenu.

À rendre via Campus sous le format : `NOM_TP3.pdf`