

* Devoir *

Problème #1 (10 points)

Le fichier Excel « Taux de change US/Euro » contient les données mensuelles du taux de change du dollar américain par rapport à la monnaie Euro, de janvier 1999 à octobre 2017 inclusivement. Ces données sont importantes pour les investisseurs, tant américains qu'Européens.

Il serait donc utile d'examiner si ces données pourraient être modélisées adéquatement par une série chronologique et si des prédictions suffisamment fiables pourraient être ainsi établies concernant les taux de change sur un certain nombre de mois.

Votre étude consiste donc ...

- (1) À identifier un (des) modèle(s) de série(s) chronologique(s) représentant les données mensuelles des années 1999 à 2016 inclusivement, les mois de la dernière année (2017) étant mis en réserve pour vérifier une partie des prévisions que vous ferez plus tard.
- (2) À estimer les paramètres du (des) modèle(s) de série(s) chronologique(s) que vous aurez retenu(s) comme candidats.
- (3) À vérifier la qualité du (des) modèle(s) dont vous aurez estimé les paramètres.
- (4) À établir des prévisions à partir du “meilleur” modèle de série chronologique choisi.

••••

Suggestions ...

- a) Dresser d'abord un graphique de votre série temporelle (du 01/1999 au 12/ 2016).
- b) Déterminer s'il y a stationnarité. Si oui, passez en e).
- c) S'il n'y a pas stationnarité, déterminer s'il serait approprié d'utiliser la transformée Box–Cox. Si ce n'est pas approprié, appliquer une (ou plusieurs différences) sur la série temporelle et passer en e). Si c'est approprié, choisir la fonction qui sera pertinente pour transformer vos données et passer en d).
- d) Dresser le graphique de votre série transformée et examiner une première différence. Dresser son graphique et vérifier, avec le test de Dickey-Fuller, s'il y a suffisamment d'évidence quant à la stationnarité de cette différence (à un niveau de confiance de 95%). S'il n'y a pas stationnarité, vous passez à une deuxième différence, sinon vous pouvez passer à l'identification du modèle stationnaire en e).

e) Vous identifiez le (les) modèle(s) ARMA(p,q) le(s) plus approprié(s) en utilisant les fonctions d'autocorrélation échantillonnale ACF, PACF et EACF. Dresser les graphiques pertinents et vérifier la fiabilité des valeurs obtenues. Vous pourriez aussi avoir besoin du test BIC pour compléter votre (vos) sélection(s).

f) Vous estimez ensuite les paramètres du (des) modèle(s) choisi(s) par la méthode du maximum de vraisemblance (MLE) et vous utilisez les propriétés statistiques de ces estimateurs pour valider les valeurs des paramètres obtenus.

g) Pour raffiner votre sélection, vous faites une analyse des résidus de votre (vos) modèle(s). Pour la question de “normalité” des résidus, utiliser un (des) histogramme(s), un (des) graphique(s) QQ et un (des) test(s) de Shapiro-Wilks. Pour la question de “l'indépendance des résidus”, vous dresser un (des) graphique(s) des résidus et un (des) “run test”.

h) Pour finaliser votre sélection finale du « meilleur modèle », vous utilisez le “Ljung-Box test” et vous examinez la question de “l'overfitting”.

i) À partir du modèle de série chronologique que vous avez finalement choisi, vous utiliserez la méthode du minimum de l'erreur quadratique moyenne (MMSE) pour effectuer vos prévisions pour les 10 premiers mois de l'année 2017. Dresser un tableau et un graphique de ces prédictions. Établir l'intervalle de prédiction, pour chaque mois additionnel (à partir de 01/2017), correspondant à un niveau de confiance de 95%.

j) Comparer vos prédictions avec les valeurs données dans votre fichier Excel pour les mois correspondants de l'année 2017. Dresser un tableau.

k) Considérez-vous que vos résultats sont suffisamment probants? Justifier votre réponse.

Problème #2 (10 points)

Le fichier Excel « SAAQ-2015 » contient des statistiques sur le nombre d'accidents (automobiles) avec dommages corporels, le nombre de personnes accidentés, le nombre de demandes d'indemnités et le coût total de l'indemnisation (en dollars constants 2015) pour les années 1978 à 2015 inclusivement.

Vous aimeriez pouvoir établir des prévisions « fiables » pour les années 2016 à 2020 afin d'évaluer si chacune de ces rubriques continuera à diminuer sensiblement. Vous aurez donc à modéliser ces données par des séries chronologiques et à établir des prévisions utiles à l'étude du risque inhérent à ce type d'assurance.

D'une façon complètement « autonome » cette fois-ci, votre étude de ce problème consistera donc ...

- (1) À identifier un (des) modèle(s) de série(s) chronologique(s) représentant les données des années 1978 à 2015, pour chacune de ces quatre rubriques.
- (2) À estimer les paramètres des modèles de séries chronologiques que vous aurez retenus.
- (3) À vérifier la qualité des modèles dont vous aurez estimé les paramètres.
- (4) À établir des prévisions à partir des modèles de séries chronologiques choisis.

* Remarques finales :

- Vous utiliser le logiciel R.
- Soyez très clair dans votre démarche, à toutes les étapes.
- Vos tableaux et graphiques doivent être très bien identifiés. Vous pouvez les incorporer dans votre texte ou les mettre en appendice.
- Vos analyses doivent être pertinentes, rigoureuses et nous amener logiquement au choix de votre modèle de série chronologique.
- Ne pas oublier ... votre modèle doit être simple, mais pas trop simple!
- Les codes R utilisés devront figurer en appendice.