

Projet - Séries temporelles - M2 ISN

A rendre pour le lundi 3 février 2014 par voie électronique. Je demande 3 fichiers :

- Un fichier `.pdf` qui contient les réponses aux questions, illustrées des graphes et sorties statistiques que vous jugerez nécessaires, mais sans code.
- Un fichier `.R` et un fichier `.sas` contenant vos codes R et SAS (commentés et structurés par questions)

Ces fichiers sont à m'envoyer à l'adresse suivante : `chi.tran@math.univ-lille1.fr`

Exercice 1 (Etude d'une série temporelle sous R)

On étudie la série `UKDriverDeaths` qui fait partie des séries de données déjà sur R (faire `help(UKDriverDeaths)`).

Dans cet exercice, je vous demande de prendre des initiatives, au vu de ce que nous avons fait ensemble en cours et TD pour répondre aux questions (volontairement succinctes). Les choix doivent être justifiés le plus rigoureusement possible et tachez d'avoir une rédaction claire, statistiquement précise, et concise lorsqu'il faut commenter.

1. Tracez le graphe de la série et discutez du modèle que vous allez choisir pour la modéliser. En particulier, il faudra détailler le choix de la saisonnalité.
2. Effectuez la désaisonnalisation de la série temporelle et présentez ses différentes composantes. Pour chacune d'elles, commentez.
3. Que peut-on dire de la composante résiduelle ? Comment la modéliseriez-vous (bruit blanc, modèle ARMA etc.) ? Justifiez le modèle retenu et estimez-le.
4. Pouvez-vous prédire la courbe des accidents sur l'année 1986 ?

Exercice 2 (Etude d'une série temporelle sous SAS)

On s'intéresse aux chiffres du chômage, observés trimestriellement depuis le 1er trimestre de 1982 jusqu'au second trimestre de 2013, et disponibles sur la page de l'INSEE (la table `valeurs.xls` est téléchargeable sur <http://labomath.univ-lille1.fr/~tran/enseignements.html>). Plus particulièrement, on s'intéresse à l'évolution des taux de chômage dans la région Nord-Pas de Calais. Les réponses devront être rédigées comme s'il s'agissait d'un rapport technique pour la région Nord-Pas de Calais : justifications rigoureuses statistiquement (destinées à des experts statisticiens), mais aussi, ne pas perdre le but en vue (l'étude du chômage de la région).

Partie A Dans un premier temps, on s'intéresse à la série univariée `NPdC` qui correspond au taux de chômage dans le Nord-Pas de Calais.

1. Tracez le graphe de la série et commentez. Présentez quelques statistiques simples qui vous semblent utiles à présenter.
2. Testez si la série est intégrée ou non, puis estimez un modèle ARIMA qui pourrait décrire les données. On justifiera bien obtenir les différents modèles possibles et comment on choisit entre ces différents modèles.
3. Réalisez des prévisions jusqu'à l'année 2015.

Partie B Dans un second temps, on considère une approche multivariée. En effet, on peut penser que l'évolution du taux de chômage dépend de son environnement national (variable `Tot`) et de celui en Ile de France, qui est une région proche et la plus importante économiquement (variable `IdF`).

1. Tracez sur un même graphique les courbes de chômage dans le Nord-Pas de Calais, en Ile de France et dans toute la France. Commentez.
2. Testez l'intégration des trois séries chronologiques (en les traitant comme des séries univariées). Quel modèle ARMA peut-on mettre pour les séries différenciées ?
3. Lisez l'aide de SAS sur la procédure VARMAX (rappelée en fin d'exercice). En utilisant cette procédure VARMAX, quel modèle VAR retient-on finalement ?
4. (facultative) Testez la cointégration de ces séries et estimez, le cas échéant, la relation de cointégration. (Lisez pour cela la page sur le "Vector Error Correction Modeling" dans l'aide sur la procédure VARMAX).
5. Réalisez des prévisions jusqu'à l'année 2015. Comparez avec le fait de faire des prévisions en utilisant l'analyse univariée de la partie A. Commentez.

Procédure varmax :

```
proc varmax data=... plots=(forecasts)
model NPdC IdF Tot / p=. dify(1) print=covpe(.) dftest coint;
output lead=.;
run;
```

la commande `plots` permet d'obtenir des graphiques incluant les prévisions associées à l'horizon précisé dans `lead`. La commande `model` précise les séries à analyser en donnant le paramètre `p` correspond à l'ordre du VAR. L'option `print=covpe(n)` demande l'affichage des matrices de variance-covariances jusqu'à l'ordre `n`. L'option `dify(1)` précise que le modèle VAR est à être estimé pour $(\Delta NPdC, \Delta IdF, \Delta Tot)$. Les options `dftest` et `coint` permettent de mettre en oeuvre des tests de Dickey Fuller et des tests de co-intégration.