TP Séries Chronologiques n^0 2 : tendance d'une série chronologique

On commence par simuler les séries X_2 et X_3 (notez qu'à moins de les réaffecter, ces séries sont enregistrées sous forme de fichiers dans votre $.Data$; ceci est important car on les utilisera plusieurs fois à titre d'exercices). On pourra notamment utiliser les commandes rep et for décrites respectivement p.4 et p.38 du guide Splus.

Exercices

1. Simuler une série X_2 trimestrielle commençant le deuxième trimestre 1980 et finissant le troisième trimestre 2000, telle que $X_2(k) = a(k) + S(k) + \varepsilon_k$ (k étant le numéro du trimestre), où:

   - La tendance est $a(k) = 0.1k + 3$.
   - Le saisonnier S est une fonction périodique de période 1 an, telle que $S = 3$ au premier trimestre, $S = 5$ au second, $S = -1$ au troisième et $S = -7$ au quatrième.
   - ε est un bruit blanc gaussien de variance 2 (c'est-à-dire une suite de variables aléatoires indépendantes gaussiennes centrées de variance 2).

   Quelle est la partie déterministe de X_2 ? Les variables $X_2(k)$ sont-elles indépendantes?

2. Simuler une série X_3 telle que $X_3(k+2) = 0.6X_3(k+1) - 0.2X_3(k) + \varepsilon_k$ où ε est un bruit blanc de loi uniforme sur $[-1, 1]$ (c'est-à-dire une suite de variables aléatoires indépendantes) en fixant $X_3(0)$ et $X_3(1)$ à des valeurs quelconques. Représenter plusieurs réalisation de la série chronologique, puis regarder son évolution (on génèrera d'abord plusieurs trajectoires de taille 100, puis d'autres de taille plus longue...). Les variables $X_3(k)$ sont-elles indépendantes?

Regression

L'objectif de ce TP est de mettre en pratique différentes procédures de régression permettant d'estimer la tendance éventuelle d'une série chronologique. Se connecter sur le répertoire $TPSeries$. Cette fois-ci, et contrairement à ce que l'on avait fait avec les séries simulées, on travaille avec une série issue de données réelles et dont on ne sait rien d'une possible modélisation.

Splus -e	Lancer Splus (avec l'option éditeur de ligne)
help.start()	Ouvre la fenêtre d'aide.
x_hstart	On travaille sur une série chronologique "réelle" interne à Splus.
	Cette série représente le nombre de déménagements par mois.
is.ts(x)	x est-il une "time series" (série chronologique)?
tp_time(x)	On associe le vecteur tp constitué des temps associés à x.
motif()	Ouvre une fenêtre graphique.
tsplot(x)	Tracé de la série chronologique.

	Régression linéaire
res_lm(x ~ tp)	Régression linéaire de x par tp.
res	Pour voir.
names(res)	Différents composants associés à *res*.
res$coef	Coefficients de *res*.
abline(res)	Droite de régression s'ajoutant à la série chronologique.
summary(res)	Résumé des résultats de la régression.
segments(tp,res$fitted,tp,x)	Visualisation des résidus.
res$res	Valeurs des résidus.
ltext("R2=...")	Remplacer les "..." par la valeur lue lors de la commande précédente. Affiche la valeur du R2 Positionement à la souris : rendre d'abord la fenêtre active en cliquant dans sa barre supérieure, puis cliquer à l'endroit où on veut écrire la valeur.
R2_1-var(res$res)/var(x)	Calcul manuel du coefficient R2.
	Régression polynomiale (degre 2)
res2_lm(x ~ poly(tp,2))	Ajustement polynomial (degré 2)
plot(tp,x)	diagramme de dispersion
lines(tp,res2$fitted)	Graphe du polynome
R2_1-var(res2$res)/var(x)	Calcul direct du R2 (éditer à partir de `history(var,ed=T)`).
round(R2,4)	Lecture de sa valeur (`round(x,n)` arrondi x à n décimales).
summary(res2)	Vérification.
segments(tp,res2$fitted,tp,x)	Visualisation des résidus.
	Régression polynomiale (degre 9)
res3_lm(x~ poly(tp,9))	Ajustement polynomial (degré 9.)
plot(tp,x)	Diagramme de dispersion
lines(tp,res3$fitted)	Graphe du polynome
summary(res3)	Résumé du résultat.

Exercice

Sur cette série chronologique, représenter dans une même fenêtre graphique (découpée...) les résultats des différentes regression en faisant croître le degré du polynôme. A partir de quel degré a-t-on à peu près toujours la même chose? En enlevant la tendance obtenue avec la regression polynomiale pour ce degré, peut-on dégager une saisonnalité?

Regressions locales

Pour obtenir une meilleure adaptation de la régression aux variations locales de la série, on peut procéder avec des régressions locales (loess et lowess : "lo" comme local, "ss" comme scatterplot smoothing) qui fonctionnent avec une fenêtre (dont on peut directement choisir la longueur dans le cas lowess), "coulissant" successivement le long des temps de la série. Dans chaque fenêtre, on fait une régression polynômiale de degré en général 0, 1 ou 2, par moindres carrés pondérés, sur les données de la série, et on obtient ainsi un point lissé pris en général au milieu de la fenêtre. Les poids de la pondération sont obtenus suivant la règle suivante: les poids sont plus élévés plus on est proche du point à lisser. Dans le cas de la régression lowess, on opère de manière itérative, une nouvelle pondération (d'où le "w" pour weighted), en fonction des écarts aux valeurs initiales.

———————————————————-

Régression loess

———————————————————-

resl_loess(x~tp)	régression "locale".
names(resl)	Différents résultats de la régression loess.
resl.preplot_preplot(resl)	Préparation du graphique.
plot(tp,x)	Diagramme de dispersion (ter).
plot(resl.preplot)	Graphique associé.
title("Résultats de loess(x~ tp)")	
plot(resl.preplot,new=F)	(bis)
points(tp,x)	Représentation des points (remarquez que ce graphique s'ajoute à un graphique existant).
title("Résultats de loess(x~ tp) avec points")	
R2_1-var(resl$res)/var(x)	Calcul direct du R2. Conclusion?

————————————————

Lissage lowess

————————————————

low1_lowess(tp,x,f=0.5)	Lissage par lowess de la série x, avec une taille de fenêtre fixée à 0.5*(longueur de x).
names(low1)	Différents résultats de la régression lowess. Noter que lowess n'a pas la syntaxe et les options identiques aux autres régressions.
low1_ts(low1$y,1966,12)	On forme à nouveau une série chronologique.
tsplot(x,low1)	Tracé de la série x et de la série lissée.
motif()	
low2_lowess(tp,x,f=0.1)	Lissage par lowess de la série x, avec une taille de fenêtre fixée à 0.1*(longueur de x).
low2_ts(low2$y,1966,12)	On forme la série chronologique.
tsplot(x,low2)	Tracé de la série x et de la série lissée.
tsplot(low1,low2)	Pour comparer les deux lissages; expliquer la différence entre les deux lissages. Changer à nouveau la taille de la fenêtre afin d'obtenir un saisonnier + bruit satisfaisant (d'après vous, qu'est-ce donc qu'un saisonnier + bruit satisfaisant?). Quelle est la taille de la fenêtre choisie?

Exercice

1. Reprener les séries du TP précédent et estimer leur tendance à l'aide des diffèrentes méthodes proposées.