En conséquence de la proposition ci-dessus, pour identifier un $GARCH(p,q)$, on identifiera d'abord le processus $ARMA(p,q)$ qui modélise (X_t^2). Pour identifier p dans le cas $m=q$ (c'est à dire $p \leq q$), il faut effectuer des tests sur les coefficients α_1, ..., α_q du processus $ARMA(m,q)$ (sont-ils significativement non nuls ?).

6.3. Feuille d'exercices numéro 7 (durée : 3h)

Préliminaires. Créer un fichier texte dans lequel vous répondrez clairement aux questions ci-dessous, en incluant vos codes R, les résultats obtenus sous R (graphique y compris), vos interprétations, remarques ... Une fois ce TP fini, vous metterez en forme votre compte-rendu et l'exporterez au format pdf.

6.3.1. `EuStockMarket` (valeurs à la fermeture de divers indices des marchés européens). Charger les logarithmes des acroissement de l'indice DAX en éxécutant :
`dax <- diff(log(EuStockMarkets))[,"DAX"]`. Soit `dax2=dax^2` (les carrés des composantes de `dax`).

(1) On cherche à ajuster un modèle $ARMA_{m,q}$ avec $m \geq q$ sur `dax2`. Choisir (m,q) dans $\{(1,0),(1,1),(2,0),(2,1),(2,2)\}$ de manière à minimiser la quantité AIC (utiliser les instructions du TP 5).

(2) Proposer un modèle $GARCH_{p,q}$ modélisant `dax`. Tester les résidus.

6.3.2. `NYSE` (New-York stock exchange daily return). Charger le fichier `nyse.dat` à l'adresse
`http://math.unice.fr/~rubentha/enseignement/nyse.dat`

(1) On suppose que `nyse` suit un modèle $GARCH(1,1)$. On notera (X_t) le processus. Estimer les paramètres du modèle.

(2) Tracer dans la même fenêtre : `nyse` entre les temps 900 et 1000 et, pour chaque temps t dans $\{900,\dots,1000\}$, les deux extrémités d'un intervalle $[-m;m]$ tel que
$$\mathbb{P}(X_t \in [-m;m] | X_{t-1}, X_{t-2}, \dots) = 0,32\,.$$

Mise en œuvre sous `R`. Pour utiliser les fonctions spécifiques à l'étude des modèles $ARCH$ et $GARCH$, il faut avant tout charger le package `tseries` à l'aide de la commande `library(tseries)` (et/ou en cherchant dans le menu « packages » de `R`). La fonction `garch` permet d'estimer un $GARCH_{p,q}$: `serie<-garch(data,order=c(p,q))` Parmi les sorties de cette fonction : `coef, residuals, fitted.values`. La sortie `serie$fitted.values` contient pour chaque temps t les valeurs $-\sigma_t$ et σ_t (avec la notation standard : $X_t = \sigma_t e_t$ ($e_t \sim \mathcal{N}(0,1)$ et σ_t^2 est la variance calculée à partir de $X_{t-1}, X_{t-2}, \dots, \sigma_{t-1}, \sigma_{t-2}, \dots$. On rappelle que si $Z \sim \mathcal{N}(0,1)$, $\mathbb{P}(Z \leq 1) = 0,8413$ (approximativement).

Instructions utile : si `A` est un tableau à deux colonnes alors `A[,j]` est la j-ème colonne de `A`. Pour fixer la taille de la fenêtre graphique en ordonnée : `plot(x,ylim=c(-0.7,0.9))` (c'est un exemple).

6.4. Feuille d'exercices numéro 8 (révisions)

(1) Soit X un $AR(1)$ (stationnaire) défini par :
$$\forall t \in \mathbb{Z}\,,\; X_t = \phi X_{t-1} + Z_t\,,$$
avec Z un bruit blanc centré de variance σ^2 et $|\phi| < 1$. On pose :
$$\forall t \in \mathbb{Z}\,,\; Y_t = X_t - X_{t-1}\,.$$
   (a) Montrer que Y est un processus stationnaire centré, puis que c'est un $ARMA(p,q)$. Préciser les valeurs de p et q et donner l'équation de récurrence vérifiée par Y.

   (b) Que vaut la variance de Y_t (pour tout t) ?

(2) Soit X un processus $ARMA$ (stationnaire) vérifiant

$$\forall t \in \mathbb{Z}\,,\; X_t - \frac{7}{6}X_{t-1} + \frac{1}{3}X_{t-2} = Z_t - \frac{1}{4}Z_{t-1} - \frac{1}{8}Z_{t-2}\,,$$

avec Z un bruit blanc (centré). Montrer que :

$$\forall t \in \mathbb{Z}\,,\; X_t - \frac{2}{3}X_{t-1} = Z_t + \frac{1}{4}Z_{t-1}\,.$$

(3) La série `sncf.rda` contient le trafic mensuel sur les lignes SNCF de janvier 1963 à décembre 1980 en millions de passagers par kilomètres. Rappel : Δ_T est l'opérateur des différences avec un décalage de T.

(a) Notons X la série à étudier. Tracer la densité spectrale de X et les autocorrélations de X pour trouver la période T de la composante saisonnière de X.

(b) Notons $Y_1 = \Delta_1 X$, $Y_2 = \Delta_T X$, $Y_3 = \Delta_T \circ \Delta_1 X$. Pour chacune de ces séries $Y_{\dots}$,

(i) tracer les autocorrélations et les autocorrélations partielles,

(ii) en déduire un modèle $ARMA$ pour $Y_{\dots}$ (on doit pouvoir voir le nombre de coefficients non nuls sur les graphiques précédents) et estimer les coefficients,

(iii) tracer la série des résidus et tester la blancheur des résidus.
(On doit trouver que seul Y_3 est un $ARMA$.)

(iv) Trouver, parmi les $(p,q) \in \{0,1,2\}^2$, le couple minimisant le critère AIC pour Y_3.

(c) Prédire le trafic de l'année 1981.

(4) On étudie la série `serie1.rda` (disponible sur `http://math.unice.fr/~rubentha/enseignement/serie1.rda`, sauver le fichier sur l'ordinateur puis taper `load('chemin_complet/serie1.rda')`, la série `serie1` sera chargé dans `R`). Le modèle choisi pour cette série est

$$\forall t \in \{1,2,\dots,n\}\,,\; x_t = m(t) + s(t) + \epsilon_t\,, \tag{6.1}$$

où m est une tendance polynômiale, s une saisonnalité et ϵ un processus $ARMA$ (stationnaire) centré (condtui par un bruit blanc gaussien).

(a) Tracer la série `serie1.rda` et son périodogramme. Quelle est la période p de la saisonnalité s ?

(b) Estimer une tendance polynômiale et une saisonnalité de période p par moindres carrés ordinaires sur la série `serie1.rda`.

(c) Sur le résidu de cette estimation, proposer un modèle $ARMA$.

(d) Une fois identifié l'ordre du bruit $ARMA$, ajuster directement avec `arima` un modèle de type (6.1) sur `serie1.rda`. Pensez à valider votre modèle et donnez les valeurs estimées des paramètres de l'$ARMA$.

(e) Représenter l'estimation de $m + s$.

(f) En utilisant la fonction `predict`, prédire les 12 prochaines valeurs de la série `serie1.rda`.

(5) Dans cet exercice, on étudie la série `serie2.rda` (disponible sur `http://math.unice.fr/~rubentha/enseignement/serie2.rda`, à charger dans `R` à l'aide de l'instruction `load`).

(a) Tracer la série `serie2.rda`. Est-elle stationnaire ?

(b) Ajuster un modèle $ARMA$ sur cette série. Pensez à valider votre modèle.

(c) Peut-on considérer cet $ARMA$ comme étant de moyenne nulle ?

6.5. Corrigé de la feuille d'exercices numéro 8

(1) (a) Pour tout t, $\mathbb{E}(X_t) = \phi\mathbb{E}(X_{t-1}) + 0$. Comme X est stationnaire, $\mathbb{E}(X_t) = \mathbb{E}(X_{t-1})$, et donc, puisque $|\phi| < 1$, $\mathbb{E}(X_t) = 0$. Donc, pour tout t, $\mathbb{E}(Y_t) = \mathbb{E}(X_t) - \mathbb{E}(X_{t-1}) = 0$. Pour tout t, $\mathbb{E}(Y_t^2) = \mathrm{Var}(X_t) + \mathrm{Var}(X_{t-1}) - 2\,\mathrm{Cov}(X_t, X_{t-1})$, et puisque X est stationnaire, cette quantité ne dépend pas de t. Pour tout t et $h > 0$,

$$\begin{aligned}\mathbb{E}(Y_tY_{t+h}) &= \mathbb{E}((X_t - X_{t-1})(X_{t+h} - X_{t+h-1}))\\ &= \mathrm{Cov}(X_t, X_{t+h}) - \mathrm{Cov}(X_t, X_{t+h-1}) - \mathrm{Cov}(X_{t-1}, X_{t+h}) + \mathrm{Cov}(X_{t-1}, X_{t+h-1})\,,\end{aligned}$$

qui ne dépend pas de t puisque X est stationnaire. Donc Y est stationnaire.
On calcule :

$$\begin{aligned}Y_t &= \phi X_{t-1} + Z_t - (\phi X_{t-2} + Z_{t-1})\\ &= \phi Y_{t-1} + Z_t - Z_{t-1}\,.\end{aligned}$$

Donc Y est un $ARMA(1,1)$.

(b) On calcule (pour t quelconque) :

$$\mathrm{Var}(X_t) = \phi^2\,\mathrm{Var}(X_{t-1}) + \sigma^2\,.$$

Donc

$$\mathrm{Var}(X_t) = \frac{\sigma^2}{1-\phi^2}\,.$$

Nous avons pour tout t :

$$\begin{aligned}\mathbb{E}(X_tX_{t-1}) &= \phi\,\mathrm{Var}(X_{t-1}) + 0\\ &= \frac{\phi\sigma^2}{1-\phi^2}\,.\end{aligned}$$

D'où :

$$\begin{aligned}\mathrm{Var}(Y_t) &= \mathrm{Var}(X_t) + \mathrm{Var}(X_{t-1}) - 2\,\mathrm{Cov}(X_t, X_{t-1})\\ &= \frac{2\sigma^2}{1-\phi^2} - 2\left(\frac{\phi\sigma^2}{1-\phi^2}\right)\\ &= \frac{2\sigma^2(1-\phi)}{1-\phi^2} = \frac{2\sigma^2}{1+\phi}\,.\end{aligned}$$

(2) Nous voulons éliminer les X dans l'expression $X_t - (2/3)X_{t-1}$ (t quelconque fixé). Nous commençons donc le calcul par

$$\begin{aligned}X_t - \frac{2}{3}X_{t-1} &= X_t - \frac{7}{6}X_{t-1} + \frac{1}{3}X_{t-2} + \frac{X_{t-1}}{2} - \frac{1}{3}X_{t-2}\\ &= \left(Z_t - \frac{Z_{t-1}}{4} - \frac{Z_{t-2}}{8}\right) + \frac{1}{2}\left(X_{t-1} - \frac{7}{6}X_{t-2} + \frac{X_{t-3}}{3}\right)\\ &\quad + \frac{1}{4}X_{t-2} - \frac{1}{6}X_{t-3}\\ &= \left(Z_t - \frac{Z_{t-1}}{4} - \frac{Z_{t-2}}{8}\right) + \frac{1}{2}\left(X_{t-1} - \frac{7}{6}X_{t-2} + \frac{X_{t-3}}{3}\right)\\ &\quad + \frac{1}{4}\left(X_{t-2} - \frac{7}{6}X_{t-3} + \frac{X_{t-4}}{3}\right) + \frac{1}{8}X_{t-3} - \frac{1}{12}X_{t-4}\,.\end{aligned}$$

Ce qui nous donne l'idée de calculer $\sum_{k\geq 0}\left[\frac{1}{2^k}\left(X_{t-k} - \frac{7}{6}X_{t-k-1} + \frac{1}{3}X_{t-3}\right)\right]$. Nous commençons par montrer que cette suite est bien définie (en utilisant le fait que les Z_t sont centrés et tous de même variance)

$$\mathbb{E}\left(\left(\sum_{k\geq 0}\left[\frac{1}{2^k}\left(X_{t-k} - \frac{7}{6}X_{t-k-1} + \frac{1}{3}X_{t-k-2}\right)\right]\right)^2\right)$$

$$= \mathbb{E}\left(\left(\sum_{k\geq 0}\left[\frac{1}{2^k}\left(Z_{t-k}-\frac{1}{4}Z_{t-k-1}-\frac{1}{8}Z_{t-k-2}\right)\right]\right)^2\right)$$
$$= \sum_{k\geq 0}\mathbb{E}(Z_{t-k}^2)\times\left(\frac{1}{2^k}-\frac{1}{4}\frac{1}{2^{k-1}}\mathbb{1}_{k\geq 1}-\frac{1}{8}\frac{1}{2^{k-2}}\mathbb{1}_{k\geq 2}\right)^2$$
$$< +\infty \text{ car série géométrique}\,.$$

Donc la somme qui nous intéresse est finie presque sûrement. Nous calculons alors

$$\sum_{k\geq 0}\left[\frac{1}{2^k}\left(X_{t-k}-\frac{7}{6}X_{t-k-1}+\frac{1}{3}X_{t-k-2}\right)\right]$$
$$= X_t + X_{t-1}\left(-\frac{7}{6}+\frac{1}{2}\right)+\sum_{k\geq 2}\frac{X_{t-k}}{2^k}\left(1-\frac{7}{6}\times 2+\frac{1}{3}\times 2^2\right)$$
$$= X_t + X_{t-1}\times\frac{2}{3}\,.$$

Nous avons par ailleurs

$$\sum_{k\geq 0}\left[\frac{1}{2^k}\left(X_{t-k}-\frac{7}{6}X_{t-k-1}+\frac{1}{3}X_{t-k-2}\right)\right]$$
$$= \sum_{k\geq 0}\frac{1}{2^k}\left(Z_t-\frac{Z_{t-1}}{4}-\frac{1}{8}Z_{t-2}\right)$$
$$= Z_t + Z_{t-1}\left(-\frac{1}{4}+\frac{1}{2}\right)+\sum_{k\geq 2}\frac{Z_{t-k}}{2^k}\left(1-\frac{1}{4}\times 2-\frac{1}{8}\times 2^2\right)$$
$$= Z_t + Z_{t-1}\left(-\frac{1}{4}+\frac{1}{2}\right)\,.$$

Ce qui finit la démonstration.

(3) (a) Pour tracer le périodogramme : `k=kernel("daniell",4);` `spec.pgram(sncf,k,taper=0,log='no')`. On obtient la figure 6.5.1. Rappel : les abscisses vont de 0 à π (l'échelle affichée n'est pas correcte). Le pic le plus important se trouve donc en $\pi/6$, ce qui correspond à une période $2\pi/(\pi/6) = 12$ (ce qui est bien cohérent avec le graphique de la série et les autocorrélations (`png(filename="blabla/tp8/5.1.png") ; par(mfrow=c(2,1)) ; plot(sncf) ;` `acf(sncf,lag.max=50) ; dev.off()`, figure 6.5.2)

(b) Calculons les séries demandées : `y1=diff.ts(sncf,lag=1); y2=diff.ts(sncf,lag=12);` `y3=diff.ts(y1,lag=12)`.

— On commence par `y1` (figure 6.5.3). On se dit que Y_1 pourrait être un $AR(3)$, et donc que X pourrait être un $ARIMA(3,1,0)$. On estime les coefficients par : `out<-arima(sncf,order=c(3,1,0))`. On fait un test de niveau $0,05$ pour savoir si les résidus forment un bruit blanc. L'instruction `Box.test(out$resid,lag=10)` renvoie une p-valeur de $0,002345$ (le test calcule une statitique et la p-valeur est la probabilité d'observer une statistique aussi grande sous l'hypothèse « bruit blanc »). Comme la p-valeur est plus petite que $0,05$, on rejette l'hypothèse « bruit blanc ». Notre processus ne peut pas être un $ARIMA(3,1,0)$.

— On continue avec Y_2 (figure 6.5.4). On se dit que Y_2 pourrait être un $AM(1)$ et donc que X pourrait être un $SARIMA(0,0,1,12)$. On estime les coefficients par `out<-arima(sncf,order=c(0,0,1),seasonal=list(order=c(0,0,1),period=12))`. On fait un test de niveau $0,05$ pour savoir si les résidus forment un bruit blanc. L'instruction `Box.test(out$resid,lag=10)` renvoie une p-valeur de $2,2\times 10^{-16}$,