

Méthodes statistiques pour la finance

Ollivier Taramasco

Esimag - 3^e année - Ingénierie financière - Mathématiques et Informatique pour la finance
Master Finance - Finance Quantitative - Mathématiques et Informatique pour la finance

Chapitre 1

Analyse descriptive des rentabilités

1.1 Les données

1.1.1 Échantillon de prix

On note $(s_t, t = 0, 1 \dots T)$ un échantillon des cours à des dates t régulièrement espacées d'un actif financier donné.

Obtenir un tel échantillon pose déjà le problème de la définition du temps à la Bourse :

- Aucune Bourse n'est ouverte 24 h / 24, 365 jours par an. On doit donc en permanence faire face au problème du traitement des jours fériés, des week-ends ou de la fermeture des marchés les jours ouvrés.
- Même sur des marchés financiers en cotation en continu (comme Euronext), il y a des données inexistantes dans les échantillons. Par exemple, si l'unité de temps est "1 jour" et que les prix sont relevés tous les jours ouvrés à 12h00 00 et s'il n'y a pas de transaction un jour donné à cette heure exacte, le prix coté est celui de la précédente transaction effectuée.

En général, lorsqu'on s'intéresse à des données de prix hors "intra-day", on choisit de relever soit le prix d'ouverture, soit le prix de clôture et on saute les jours fériés et les week-ends (le lundi suit donc immédiatement le vendredi). Même si cette "solution" n'est pas entièrement satisfaisante, elle est de loin la plus pratique (et quasiment la seule utilisée).

1.1.2 Correction des données

Versement des dividendes

Si on note d_t la valeur monétaire du dividende versé entre les dates t et $t + 1$, l'investisseur qui détenait à la date t l'actif a reçu en $t + 1$ les d_t €. Son avoir, à cette même date, est donc égal à $s_{t+1} + d_t$ et donc la valeur de détention de l'actif n'est pas s_{t+1} cotée, mais bien $s_{t+1} + d_t$. Comme on s'intéresse justement à cette valeur économique, il est nécessaire de corriger les prix du versement des dividendes.

Si $(s_t, t = 0 \dots T)$ est la série initiale, la série corrigée peut être définie par :

$$\begin{cases} s_0^* = s_0 \\ s_t^* = s_t + \sum_{u=0}^{t-1} d_u \quad t = 1 \dots T \end{cases}$$

ou bien par :

$$\begin{cases} s_T^* = s_T \\ s_t^* = s_t - \sum_{u=t+1}^T d_u \quad t = 0 \dots T-1 \end{cases}$$

Modification du capital

Il arrive que des sociétés fassent appel à de nouveaux actionnaires par une opération appelée *augmentation du capital*. Par exemple, la société X, dont le capital comporte 100 000 actions valant 10 €, désire augmenter celui-ci de 99 000 € par la vente de 11 000 actions à 9 € (le 1 € de différence est une prime pour les nouveaux actionnaires).

Date	t	Émission des nouvelles actions	$t+1$
Nombre d'actions	100 000	11 000	111 000
Prix action	10 €	9 €	s_{t+1}
Capitalisation boursière	1 000 000	1 099 000 €	111 000 s_{t+1}

Au moment de l'émission, la valeur d'une action de la société X est égale à $\frac{1\,099\,000}{111\,000} = 9,90$ €. Il y a transfert de richesse entre les anciens actionnaires (qui perdent 1%) et les nouveaux (qui gagnent 10%). Cette opération est en général, proposée en priorité aux anciens actionnaires. Pour ces derniers, l'opération est "blanche" et donc la valeur de 9,90 € doit être ramenée à 10 €. On applique donc un coefficient correcteur égal à $\frac{10}{9,90}$ à tous les prix à partir de $t+1$.

Il y a un grand nombre de manière de modifier le capital. Dans la majorité des cas, il est possible de calculer un coefficient correcteur c_t .

Comme pour le versement de dividendes, la série initiale est alors transformée en :

$$\begin{cases} s_0^* = s_0 \\ s_t^* = \left(\prod_{u=0}^{t-1} c_u \right) s_t \quad t = 1 \dots T \end{cases}$$

ou bien par :

$$\begin{cases} s_T^* = s_T \\ s_t^* = \frac{s_t}{\prod_{u=t+1}^T c_u} \quad t = 0 \dots T-1 \end{cases}$$

1.2 Processus des prix

On considère que les prix ont été corrigés des versements de dividendes et des modifications du capital. La suite $(s_t, t = 0 \dots T)$ est maintenant la suite des prix corrigés.

Pour étudier cette série, on considère que chaque s_t est une observation d'une variable aléatoire S_t . On définit donc un processus aléatoire $(S_t, t = 0 \dots T)$ sur un certain espace probabilisé (Ω, \mathcal{F}, P) .

Définition 1 (Processus strictement stationnaire) *Le processus aléatoire $(S_t, t \in \mathbb{N})$ est dit strictement stationnaire si et seulement si $\forall n \in \mathbb{N}^*, \forall (t_1, \dots, t_n) \in \mathbb{N}^n, \forall h \in \mathbb{N}$, les deux vecteurs $(S_{t_1}, \dots, S_{t_n})$ et $(S_{t_1+h}, \dots, S_{t_n+h})$ ont la même distribution de probabilité.*

Il faut remarquer que cette définition implique que tous les S_t ont la même loi.

Le cas le plus simple de processus strictement stationnaire est celui où les S_t sont i.i.d.

Définition 2 (Processus stationnaire au sens large ou du second ordre) *Un processus est dit stationnaire au second ordre ou au sens large si et seulement si :*

1. $\forall t \in \mathbb{N}, E(S_t^2) < \infty$
2. $\exists \mu \in \mathbb{R}, \forall t \in \mathbb{N}, E(S_t) = \mu$
3. $\forall t, h \in \mathbb{N}, \text{cov}(S_t, S_{t+h}) = \text{cov}(S_h, S_0)$

Remarque 1 *Un processus strictement stationnaire, est stationnaire au second ordre s'il a des moments d'ordre 2. La réciproque est, en général, fausse.*

Il est souvent très difficile de démontrer qu'un processus est strictement stationnaire et assez facile de démontrer qu'il est stationnaire au second ordre.

Pour un processus stationnaire au second ordre, on ne demande que la stabilité des deux premiers moments.

1.2.1 Stationnarité du processus des prix

La notion de préférence pour le temps implique que les acteurs économiques n'investissent que si l'espérance du prix futur est supérieure au prix d'aujourd'hui (*cf.* Chapitre 1). Le moment d'ordre 1 du processus des prix ne peut donc pas être constant dans le temps, et le processus des prix ne peut pas être stationnaire.

La loi de S_t étant différente de celle de $S_u, u \neq t$, on ne dispose que d'une seule observation s_t de la loi de S_t . Il n'est donc pas possible d'utiliser des méthodes statistiques standards sur les séries de prix.

Pour illustrer ce phénomène, le graphique Fig. 1.1 montre l'évolution de l'indice CAC 40 depuis le 1^{er} janvier 1988.

La trajectoire des prix de l'indice CAC est croissante sur la période d'étude, même s'il apparaît clairement des sous-périodes de fortes hausses ou de fortes baisses (par exemple, le gonflement et le dégonflement de la "bulle internet"). La moyenne empirique de l'indice fluctue fortement d'une période à l'autre : elle vaut par exemple 1616,4 entre 1988 et 1991 et 5196,1 entre 2000 et 2002. Ceci confirme que le processus des prix n'est pas stationnaire.

La première étape de toute étude statistique du comportement des cours boursiers consiste alors à rechercher "la variable naturelle" à laquelle il faut s'intéresser. La théorie financière indique que c'est la rentabilité d'un actif (c'est à dire la variation relative de son prix), puisque c'est la seule variable, calculée à partir des prix, qui rende comparables les investissements.

1.3 Processus des rentabilités (boursières)

1.3.1 Rentabilités d'un titre

Définition 3 *On note :*

$$r_t^a = \frac{s_t - s_{t-1}}{s_{t-1}}, t = 1 \dots T$$

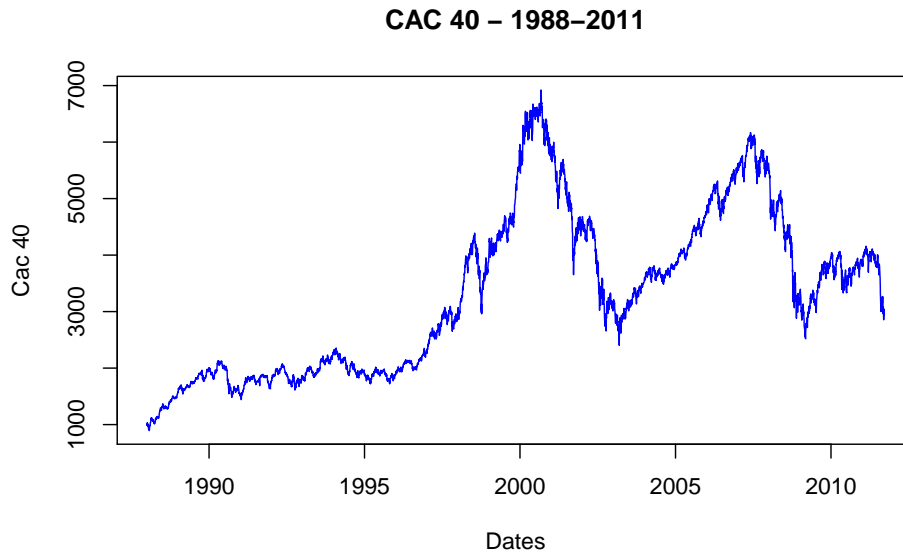


FIG. 1.1 – Évolution du CAC 40

la suite des rentabilités arithmétiques calculées sur une période.

$$r_t = \ln \frac{s_t}{s_{t-1}}, t = 1 \dots T$$

la suite des rentabilités logarithmiques (géométriques) calculées sur une période.

Il est très simple de calculer r_t à partir de r_t^a et inversement puisque :

$$r_t^a = e^{r_t} - 1 \text{ et } r_t = \ln(1 + r_t^a)$$

On considère comme pour les prix, que chaque r_t^a est une observation d'une variable aléatoire R_t^a et on définit alors un processus $(R_t^a, t \in \mathbb{N})$. On définit de même le processus $(R_t, t \in \mathbb{N})$.

Remarque 2

1. $r_t^a \simeq r_t$ si $s_t - s_{t-1}$ est petit devant s_{t-1} .
2. r_t^a est la rentabilité "réelle" de l'opération financière consistant à acheter le titre à la date $t-1$ et à le revendre à la date t .
3. Les prix étant positifs, la loi de R_t^a est à support dans $[-1, +\infty[$, la loi de R_t est à support réel. Il est donc plus facile de rechercher la loi de R_t .

Même si la seule variable qui ait du sens en finance est la rentabilité arithmétique, on va préférer dans la modélisation la rentabilité logarithmique. Il y a, de plus, deux arguments supplémentaires qui nous confortent dans ce choix :

1. Rentabilité calculée sur T périodes consécutives $T \geq 1$

$$\begin{cases} r^a(t, T) &= \frac{s_{t+T-1} - s_{t-1}}{s_{t-1}} &= \prod_{u=1}^T (1 + r_{t+u-1}^a) - 1 \\ r(t, T) &= \ln \left(\frac{s_{t+T-1}}{s_{t-1}} \right) &= \sum_{u=1}^T r_{t+u-1} \end{cases}$$

Le calcul est donc bien plus simple dans le cas de la rentabilité logarithmique.

2. Cotation en continu : Sur des marchés ouverts en continu, il est possible d'acheter ou de vendre un titre à n'importe quel instant. Il est donc naturel de définir la rentabilité instantanée par :

$$r_t^c dt = \frac{s_{t+dt} - s_t}{s_t} = \frac{ds_t}{s_t}$$

et donc

$$\int_t^T r_s^c ds = r(t, T)$$

La rentabilité logarithmique est donc la rentabilité réelle d'un actif acheté/vendu continuellement sur un intervalle de temps donné. On donne donc un sens "financier" à ce calcul.

1.3.2 Rentabilités d'un portefeuille

Un portefeuille W_t de N titres numérotés de 1 à N est détenu à la date t à raison de $a_i, i = 1 \dots N$ action(s) par titre (a_i peut être négatif si la vente à découvert est permise).

On note $w_i, i = 1 \dots N$ la proportion de la fortune initiale investie dans l'actif i . Par définition $w_i = \frac{a_i s_{it}}{\sum_{j=1}^N a_j s_{jt}}$.

On obtient alors très facilement les formules suivantes :

$$\begin{aligned} r_W^a(t, t+h) &= \sum_{i=1}^N w_i r_i^a(t, t+h) \\ r_W(t, t+h) &= \ln \left(\sum_{i=1}^N w_i e^{r_i(t, t+h)} \right) \end{aligned}$$

On remarque que si pour tous les titres $r_i(t, t+h)$ est petit, $r_W(t, t+h) \approx \sum_{i=1}^N w_i r_i(t, t+h)$.

1.4 Hypothèse statique

L'hypothèse statique suppose que le vecteur Y_t des rentabilités de $N \geq 1$ titres calculées sur un même pas de temps sont des vecteurs i.i.d. La suite des observations (de dimension N) $(y_t, t = 1 \dots T)$ de ce processus forme alors un échantillon empirique, au sens de la statistique.

Avant d'entrer dans la modélisation, on peut se demander si l'hypothèse statique est conforme à la théorie financière.

Choisissons un titre et notons S_t son prix à la date t et R_t sa rentabilité sur la période $[t, t+1]$. Il vient $\forall t \in \mathbb{N}^*$:

$$\begin{aligned} \ln S_t &= \ln S_{t-1} + R_t \\ &= \ln S_0 + \sum_{u=1}^t R_u \end{aligned}$$

1. Si le portefeuille est autofinancé, on ne peut pas définir cette proportion

On note $\mu = E(R_t)$ et $\sigma^2 = Var(R_t)$ (ce qui implicitement suppose que le processus des rentabilités admet des moments au moins jusqu'à l'ordre 2). L'hypothèse de préférence pour le temps des individus impose que $\mu > 0$. Il vient alors :

$$E[\ln S_t | S_{t-1}] = \ln S_{t-1} + \mu > \ln S_{t-1}$$

Le processus $(\ln S_t)$ est une sous-martingale par rapport à lui-même, ce qui est un cas particulier de l'hypothèse d'efficience des marchés selon Fama.

De plus, $E[\ln S_t | S_0] = t\mu + \ln S_0$, il y a augmentation moyenne du prix et $Var[\ln S_t | S_0] = t\sigma^2$ qui s'accompagne d'une variabilité des prix croissante.

Si on choisissait n'importe quel portefeuille d'espérance de rentabilité $\mu_W > 0$, on obtiendrait le même résultat. Comme aucun agent économique est sensé choisir un portefeuille d'espérance de rentabilité négative, on peut dire que l'hypothèse statique est conforme à la notion d'efficience des marchés. Il n'est donc pas absurde, dans un premier temps, d'en étudier les conséquences empiriques.

1.4.1 Hypothèse statique forte

L'hypothèse statique forte consiste à considérer que le processus $(Y_t, t = 1 \dots)$ (vectoriel) des rentabilités de N titres ($N \geq 1$) du marché, calculées sur un même intervalle de temps (1 s, 1 jour, ...) est i.i.d. de loi normale multivariée $\mathcal{N}_{\mathbb{R}^N}(m, \Lambda)$. On note $(Y_t) \rightsquigarrow \text{NIID}(m, \Lambda)$.

1.4.2 Interprétation des paramètres

Nous savons que les acteurs économiques sont sensibles à (au moins) deux dimensions d'un investissement : l'espérance de gain et le risque.

Pour un titre i donné, le processus $(R_{it}, t = 1 \dots) \rightsquigarrow N(m_i, \sigma_i^2)$. Si un acteur économique choisit d'investir dans ce titre, la meilleure prévision qu'il puisse faire de la rentabilité de cet investissement est $E(R_{it}) = m_i$. L'espérance mathématique de la variable R_{it} est donc la mesure naturelle de l'espérance de rentabilité.

Comment peut-on alors caractériser le risque de l'investissement ? Si la variance σ_i^2 est nulle, la variable R_{it} devient certaine et égale à son espérance. Pour l'investisseur, le titre est sans risque. Par contre, plus la variance est importante, moins il devient probable d'être à une distance "raisonnable" de la prévision (*i.e.* m_i). Pour 2 titres ayant la même espérance de rentabilité, celui des deux qui a la plus grande variance sera perçu comme plus risqué. La variance peut donc servir de *mesure* du risque de l'investissement.

1.4.3 Différentes mesures de risque

Le risque étant nécessairement fonction de la variance de R_{it} , celle-ci peut servir de mesure du risque. Les financiers préfèrent l'écart-type, car il a la même unité que la rentabilité (c'est à dire un pourcentage). Les acteurs économiques peuvent définir autrement le risque qu'ils encourent. En particulier,

1. Le risque est mesuré par la probabilité que la rentabilité d'un investissement soit inférieure à un taux cible. Sous l'hypothèse statique forte :

$$mr_1(m, \sigma) = P(R_t < \tau) = P\left(\frac{R_t - m}{\sigma} < \frac{\tau - m}{\sigma}\right) = N\left(\frac{\tau - m}{\sigma}\right)$$

où N est la fonction de répartition de la loi $\mathcal{N}(0, 1)$. Si le taux cible $\tau < m$, alors cette mesure du risque est une fonction croissante de σ (car $\frac{\partial mr_1(m, \sigma)}{\partial \sigma} = \frac{m-\tau}{\sigma^2} \phi\left(\frac{m-\tau}{\sigma}\right)$ avec $\phi(x)$ densité de $\mathcal{N}(0, 1)$). L'investisseur sensible à cette mesure du risque choisit toujours un taux cible inférieur à m . Par exemple, $\tau = 0$: risque de perte, ou $\tau = R_f$: risque d'une rentabilité inférieure au taux sans risque. Un simple changement d'échelle de mesure permet de définir le risque par $\frac{\tau-m}{\sigma}$ et lorsque $\tau = R_f$, on obtient la mesure de Sharpe : $mr_2(m, \sigma) = \frac{m-R_f}{\sigma}$.

2. Le risque est mesuré par la *Value at Risk (VaR)* : Elle se définit comme le premier quantile à $\alpha\%$ de la variation de la valeur de la richesse. Plus cette valeur est grande, plus l'actif est risqué. Si l'investisseur a une fortune initiale égale à W_0 , la *VaR* est définie par :

$$P(W_f - W_0 < -VaR(\alpha)) = P(W_0 e^{R_t} < W_0 - VaR(\alpha)) = \alpha\%$$

Sous l'hypothèse statique forte, il vient :

$$mr_3(m, \sigma) = VaR(\alpha) = W_0 \left(1 - e^m e^{N^{-1}(\alpha)\sigma}\right)$$

De plus, $\frac{\partial VaR(\alpha)}{\partial \sigma} = -W_0 N^{-1}(\alpha) \sigma e^m e^{N^{-1}(\alpha)\sigma} > 0$ dès que $\alpha < 50\%$.

3. Le risque est défini par la perte moyenne en cas de perte. Plus cette dernière est grande, plus l'investissement est perçu comme risqué. Mathématiquement on écrit :

$$mr_4(m, \sigma) = -E[R_t u(R_t)]$$

avec $u(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$. Sous l'hypothèse statique forte,

$$mr_4(m, \sigma) = \sigma \phi\left(\frac{m}{\sigma}\right) + m - m N\left(\frac{m}{\sigma}\right)$$

et $\frac{\partial mr_4(m, \sigma)}{\partial \sigma} = \phi\left(\frac{m}{\sigma}\right) > 0$.

4. etc

Sous l'hypothèse statique forte, toutes ces mesures de risque sont des fonctions croissantes de la variance de la rentabilité. Si deux titres ont la même espérance de rentabilité, celui qui est perçu comme le plus risqué est donc celui dont la variance de la rentabilité est la plus grande.

1.4.4 Diversification temporelle?

Sous l'hypothèse statique forte, le processus des rentabilités des N titres calculées sur H périodes consécutives $Y(t, t+H) \rightsquigarrow \text{NIID}(m \times H, H \times \Lambda)$. On définit alors la rentabilité par unité de temps par $\frac{1}{H} Y(t, t+H) \rightsquigarrow \text{NIID}(m, \frac{1}{H} \Lambda)$.

Ce calcul montre qu'à long terme ($H \rightarrow +\infty$), la rentabilité par unité de temps devient déterministe, ce qui fait dire à certains auteurs qu'il y a disparition complète du risque et de conclure qu'un investisseur ayant un horizon de placement long doit détenir une proportion importante d'actions dans son portefeuille, et que par contre, un investisseur ayant un horizon court doit détenir une proportion importante d'obligations sûres (d'état) à court terme.

Le tableau (1.1) donne la limite quand l'horizon de placement tend vers l'infini des différentes mesures du risque. On peut remarquer que suivant le choix de la mesure, il y a ou non réellement diversification temporelle.

Mesure de risque	Horizon H	Limite	Diversification
σ	$\sqrt{H} \sigma$	$+\infty$	Non
σ^2	$H \sigma^2$	$+\infty$	Non
$mr_1(m, \sigma)$	$N\left(\sqrt{H} \frac{\tau-m}{\sigma}\right)$	0	Oui
$mr_2(m, \sigma)$	$\sqrt{H} \frac{m-R_f}{\sigma} \quad (a)$	$+\infty$	Oui
$VaR(\alpha)$	$W_0 \left(1 - e^{mH + N^{-1}(\alpha) \sigma \sqrt{H}}\right)$	$-\infty$	Oui
$mr_4(m, \sigma)$	$\sigma \sqrt{H} \phi\left(\sqrt{H} \frac{m}{\sigma}\right) + mH \left(1 - N\left(\sqrt{H} \frac{m}{\sigma}\right)\right)$	$+\infty$	Non

TAB. 1.1 – Mesures de risque et diversification temporelle

^aCe calcul suppose que le taux sans risque “continu” sur H périodes est égal à H fois le taux sans risque sur 1 période.

1.4.5 Échantillonnage sous l’hypothèse statique forte

On dispose de T observations $(y_1 \dots y_T)$ de ce vecteur des rentabilités et il est naturel d’estimer les deux paramètres par leurs résumés empiriques :

$$\begin{aligned}\hat{m}_T = \bar{y}_T &= \frac{1}{T} \sum_{t=1}^T y_t \\ \hat{\Lambda} &= \frac{1}{T} \sum_{t=1}^T (y_t - \bar{y}_T)^t (y_t - \bar{y}_T)\end{aligned}$$

Proposition 1 *Sous l’hypothèse statique forte, les deux estimateurs \hat{m}_T et $\hat{\Lambda}$ sont convergents, asymptotiquement indépendants, normaux et sans biais. Leurs variance et covariances asymptotiques sont caractérisées par :*

$$\begin{aligned}V_{as} \left[\sqrt{T} (\hat{m}_T - m) \right] &= \Lambda \\ \text{cov}_{as} \left[\sqrt{T} (\hat{\Lambda}_T - \Lambda) a, (\hat{\Lambda}_T - \Lambda) a \right] &= {}^t a \Lambda b \Lambda + \Lambda b {}^t a \Lambda\end{aligned}$$

où a et b sont deux vecteurs quelconques de \mathbb{R}^N .

On déduit de la dernière formule, en prenant $a = b = {}^t(0 \dots 0 \ 1 \ 0 \dots 0)$, où le 1 est en i^e position, que

$$V_{as} \left[\sqrt{T} (\hat{\sigma}_{iT}^2 - \sigma_i^2) \right] = 2 \sigma_i^4$$

1.4.6 Moments d’ordre 3 et 4

Pour les titres individuels, on définit :

1. le coefficient d’asymétrie (skewness): $\gamma_3 = \frac{E[(R_t - E(R_t))^3]}{\sigma^3(R_t)}$. Sous l’hypothèse statique forte $\gamma_3 = 0$.
2. le coefficient d’aplatissement (kurtosis): $\gamma_4 = \frac{E[(R_t - E(R_t))^4]}{\sigma^4(R_t)}$. Sous l’hypothèse statique forte $\gamma_4 = 3$.

Les deux estimateurs “sans biais” de ces grandeurs sont calculés de la manière suivante :

1. $\hat{\gamma}_{3T} = \frac{1}{(T-1)(T-2)} \sum_{t=1}^T \frac{(R_t - \bar{R}_T)^3}{\hat{\sigma}_T^3}$
2. $\hat{\gamma}_{4T} - 3 = \frac{T(T+1)}{(T-1)(T-2)(T-3)} \sum_{t=1}^T \frac{(R_t - \bar{R}_T)^4}{\hat{\sigma}_T^4} - \frac{3(T-1)^2}{(T-2)(T-3)}$

Proposition 2 *Sous l'hypothèse statique forte, les deux estimateurs sans biais précédents sont convergents et asymptotiquement normaux. Leur variance asymptotique est donnée par :*

- $V_{as} \left[\sqrt{\frac{(T-2)(T+1)(T+3)}{6T(T-1)}} (\hat{\gamma}_{3T} - \gamma_3) \right] = 1$
- $V_{as} \left[\sqrt{\frac{(T-3)(T-2)(T+3)(T+5)}{24T(T-1)^2}} (\hat{\gamma}_{4T} - \gamma_4) \right] = 1$

Ce qui permet d'effectuer les tests d'hypothèses $\gamma_3 = 0$ contre $\gamma_3 \neq 0$ et $\gamma_4 = 3$ contre $\gamma_4 \neq 3$.

1.4.7 Résultats empiriques

Dans un premier temps, nous regardons la trajectoire du processus des rentabilités journalières de l'indice CAC 40 (cf. Fig. 1.2).

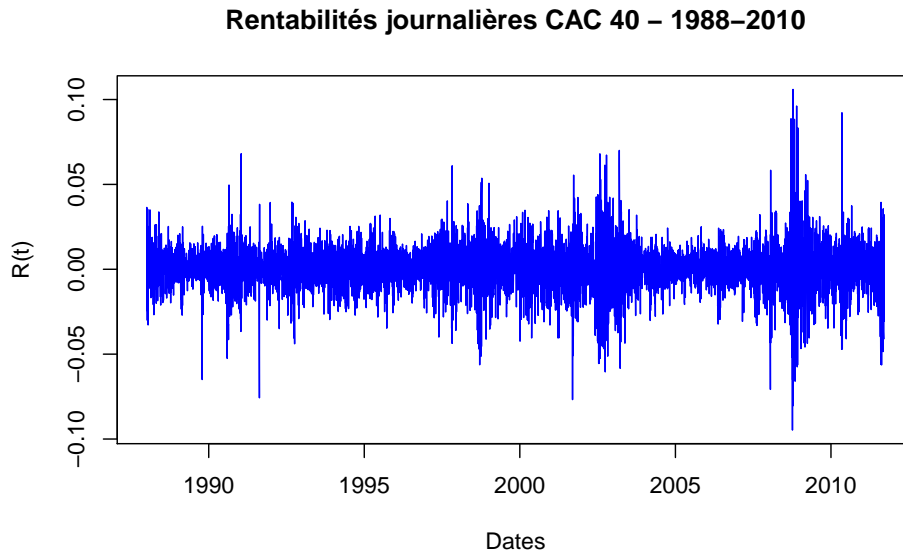


FIG. 1.2 – *Rentabilités journalières du CAC 40*

Un coup d’œil (rapide!) à la figure (Fig. 1.2) permet, dans un premier temps, de considérer la suite des rentabilités boursières comme un échantillon empirique (dans un premier temps, l’hypothèse statique n’est pas totalement absurde). Les tableaux (1.2), (1.3), (1.4) donnent les résultats des calculs pour les rentabilités journalières, hebdomadaires et mensuelles de l’indice CAC. Les valeurs entre parenthèses sont les *P-valeurs* des tests d’hypothèses respectifs $m = 0$ contre $m \neq 0$, $\gamma_3 = 0$ contre $\gamma_3 \neq 0$ et $\gamma_4 = 3$ contre $\gamma_4 \neq 3$. Les valeurs “annualisées” consiste à multiplier \bar{R}_T par respectivement 250, 52 ou 12 et $\hat{\sigma}_T$ par $\sqrt{250}$, $\sqrt{52}$ ou $\sqrt{12}$ si les calculs ont été faits sur des rentabilités respectivement journalières, hebdomadaires ou mensuelles. Pour étudier

Période	\bar{R}_T	\bar{R}_T annualisé	$\bar{\sigma}_T^2$	$\bar{\sigma}_T$	$\bar{\sigma}_T$ annualisé	$\bar{\gamma}_{3T}$	$\bar{\gamma}_{4T}$	T
1988-1994	0.055% (0.0651)	13.677%	0.00013	1.147%	18.132%	-0.252 (0.0001)	6.746 (0.0000)	1497
1994-1999	0.044% (0.2091)	11.093%	0.00016	1.246%	19.704%	-0.130 (0.0600)	4.936 (0.0000)	1246
1999-2005	-0.002% (0.9591)	-0.512%	0.00024	1.561%	24.675%	-0.036 (0.5655)	5.120 (0.0000)	1529
2005-2012	-0.017% (0.6447)	-4.248%	0.00023	1.526%	24.128%	0.047 (0.4235)	10.500 (0.0000)	1716
1988-1999	0.050% (0.0282)	12.503%	0.00014	1.193%	18.859%	-0.191 (0.0000)	5.806 (0.0000)	2743
1999-2012	-0.010% (0.7133)	-2.487%	0.00024	1.542%	24.384%	0.007 (0.8708)	7.839 (0.0000)	3245
1988-2012	0.018% (0.3306)	4.380%	0.00019	1.393%	22.029%	-0.065 (0.0386)	7.796 (0.0000)	5988

TAB. 1.2 – Moments des rentabilités journalières de l'indice CAC 40

la stabilité dans le temps des résultats obtenus, ces calculs ont été faits sur des sous périodes de tailles diverses.

On peut faire plusieurs remarques qualitatives :

Premièrement, il n'y a pas de grande stabilité de la valeur des moments sur les diverses sous-périodes. Par exemple, l'estimation de la volatilité (annualisée) des rentabilités journalières de l'indice CAC passe 17,3% sur la période 1993-1998 à 23,8% sur la période 1998-2002. Cependant, on sait qu'une seule observation "anormale" (dans le sens où sa valeur absolue est à plusieurs écarts-types de la moyenne des rentabilités) a un effet important sur la valeur de ces estimateurs, surtout sur celle du coefficient d'aplatissement (²)

Deuxièmement, le coefficient d'asymétrie est significativement négatif (sauf sur les périodes 1993-1998 et 2002-2008) au niveau 5% pour les rentabilités journalières. Cette asymétrie (à droite) a tendance à disparaître pour les rentabilités hebdomadaires et mensuelles.

Troisièmement, la valeur du coefficient d'aplatissement journalier et hebdomadaire est très grand (et significativement différent de 3 au niveau 5%) , ce qui caractérise les distributions à queues épaisses ou *leptokurtiques*, et qui est très classique pour les séries économiques et financières.

Quatrièmement, le rapport "signal sur bruit" $\frac{m}{\sigma} \approx 2,7\%$ pour les rentabilités journalières sur l'ensemble de la période. Ce rapport est très petit et rend très difficile la prévision de la rentabilité journalière future. Par exemple, sous l'hypothèse statique forte, l'intervalle de confiance à 95% des rentabilités journalières est $[-2.6\%, 2.6\%]$.

Ces diverses constatations rendent peu crédible l'hypothèse statique forte.

2. Par exemple, si on calcule le coefficient d'aplatissement des rentabilités journalières de l'indice CAC 40 sur la période 26/09/2000 - 26/09/2001, cette valeur passe de 5,11 à 3,61 lorsqu'on supprime le 11 septembre 2001

Période	\bar{R}_T	\bar{R}_T annualisé	$\bar{\sigma}_T^2$	$\bar{\sigma}_T$	$\bar{\sigma}_T$ annualisé	$\bar{\gamma}_{3T}$	$\bar{\gamma}_{4T}$	T
1988-1994	0.263% (0.0747)	13.701%	0.00068	2.606%	18.796%	-0.161 (0.2426)	4.523 (0.0000)	313
1994-1999	0.210% (0.2113)	10.901%	0.00073	2.702%	19.488%	-0.106 (0.4799)	5.031 (0.0000)	261
1999-2005	-0.009% (0.9569)	-0.491%	0.00095	3.087%	22.262%	-0.055 (0.6886)	5.175 (0.0000)	313
2005-2012	-0.062% (0.6881)	-3.211%	0.00082	2.872%	20.707%	-0.969 (0.0000)	5.698 (0.0000)	349
1988-1999	0.239% (0.0310)	12.428%	0.00070	2.648%	19.098%	-0.136 (0.1837)	4.751 (0.0000)	574
1999-2012	-0.037% (0.7488)	-1.925%	0.00088	2.973%	21.441%	-0.486 (0.0000)	5.427 (0.0000)	662
1988-2012	0.091% (0.2575)	4.741%	0.00080	2.829%	20.402%	-0.369 (0.0000)	5.276 (0.0000)	1236

TAB. 1.3 – Moments des rentabilités hebdomadaires de l'indice CAC 40

Période	\bar{R}_T	\bar{R}_T annualisé	$\bar{\sigma}_T^2$	$\bar{\sigma}_T$	$\bar{\sigma}_T$ annualisé	$\bar{\gamma}_{3T}$	$\bar{\gamma}_{4T}$	T
1988-1994	1.066% (0.1301)	12.789%	0.00349	5.905%	20.457%	-0.022 (0.9375)	4.123 (0.0444)	72
1994-1999	0.896% (0.2737)	10.757%	0.00395	6.284%	21.770%	-0.344 (0.2647)	3.764 (0.2095)	60
1999-2005	0.040% (0.9561)	0.483%	0.00381	6.174%	21.387%	-0.494 (0.0805)	3.432 (0.4399)	72
2005-2012	-0.186% (0.7777)	-2.231%	0.00349	5.908%	20.465%	-0.527 (0.0488)	3.081 (0.8775)	81
1988-1999	0.989% (0.0630)	11.865%	0.00367	6.058%	20.984%	-0.184 (0.3841)	3.880 (0.0356)	132
1999-2012	-0.080% (0.8703)	-0.954%	0.00362	6.015%	20.838%	-0.503 (0.0103)	3.217 (0.5783)	153
1988-2012	0.415% (0.2474)	4.983%	0.00366	6.048%	20.951%	-0.346 (0.0166)	3.544 (0.0585)	285

TAB. 1.4 – Moments des rentabilités mensuelles de l'indice CAC 40

1.4.8 Échantillonnage sans hypothèse de normalité

L'hypothèse statique forte étant remise en cause par les constatations empiriques, nous revenons à l'hypothèse statique sans la normalité.

Proposition 3 *Sous l'hypothèse statique, les estimateurs \hat{m}_T de m et $\hat{\Lambda}_T$ de Λ sont convergents et asymptotiquement normaux.*

1. $V_{as} \left[\sqrt{T} (\hat{m}_T - m) \right] = \Lambda$
2. Les deux estimateurs \hat{m}_T et $\hat{\Lambda}_T$ sont asymptotiquement indépendants si la loi de Y_t est symétrique.
3. La variance asymptotique de $\hat{\Lambda}_T$ dépend de Λ mais aussi de l'importance des queues de distributions. C'est une fonction croissante des coefficients d'aplatissement.

La première question qu'on doit se poser est celle de la distribution de Y_t . On estime alors les densités empiriques des rentabilités journalières, hebdomadaires et mensuelles de différents indices. Ces densités ont été estimées par la méthode du noyau (le noyau étant gaussien et le paramètre de lissage h est fixé à $\frac{\hat{\sigma}}{T^{1/5}}$).

Sur le même graphique, on trouve la densité empirique des rentabilités (tracée en gras) et la densité de la loi normale d'espérance la moyenne empirique des rentabilités et de variance, la variance empirique de ces mêmes rentabilités (trait fin).

On se rend compte que les distributions sont unimodales, qu'elles semblent relativement symétriques et qu'au plus les rentabilités ont été calculées sur une période longue, au plus leur densité empirique est proche de celle de la loi normale (cf. graphiques 1.3, 1.4 et 1.5). Si les rentabilités journalières sont i.i.d. et admettent des moments d'ordre 2, alors ce qu'on remarque ici n'est qu'une application du théorème central-limite. Ces remarques confirment les résultats sur les moments : si on admet l'hypothèse statique de distribution des rentabilités, la distribution (au moins des rentabilités journalières) doit être cherchée parmi celles qui sont leptokurtiques.

Dans la panoplie des lois classiques, la plus utilisée (en dehors de la loi normale) est la famille des *Loi de Lévy*, initialement proposée par Mandelbrot (1961). Cette famille permet de modéliser les queues de distributions épaisses et possède, comme les lois normales, la propriété de stabilité par combinaison linéaire de variables.

Les lois de Lévy sont définies à l'aide de la fonction caractéristique :

$$\begin{aligned} \ln \Phi(z) &= \ln E \left[e^{iz R_t} \right] \\ &= i \delta z - \gamma |z|^\alpha \left\{ 1 + i \beta \left(\frac{z}{|z|} \right) \omega(z, \alpha) \right\} \end{aligned}$$

avec :

$$\omega(z, \alpha) = \begin{cases} \frac{\alpha \pi}{2} & \text{pour } \alpha \neq 1, \\ \frac{2}{\pi} \ln |z| & \text{pour } \alpha = 1. \end{cases}$$

Les quatre paramètres de cette loi peuvent être interprétés de la manière suivante :

- $\alpha \in [0, 2]$ donne une idée de l'épaisseur des queues de la distributions dans le sens où ces lois sont de types parétien : $P(R_t > x) \sim c x^{-\alpha}$ pour $x \rightarrow +\infty$. Seules les lois de Lévy avec $\alpha = 2$ admettent un moment d'ordre 2.

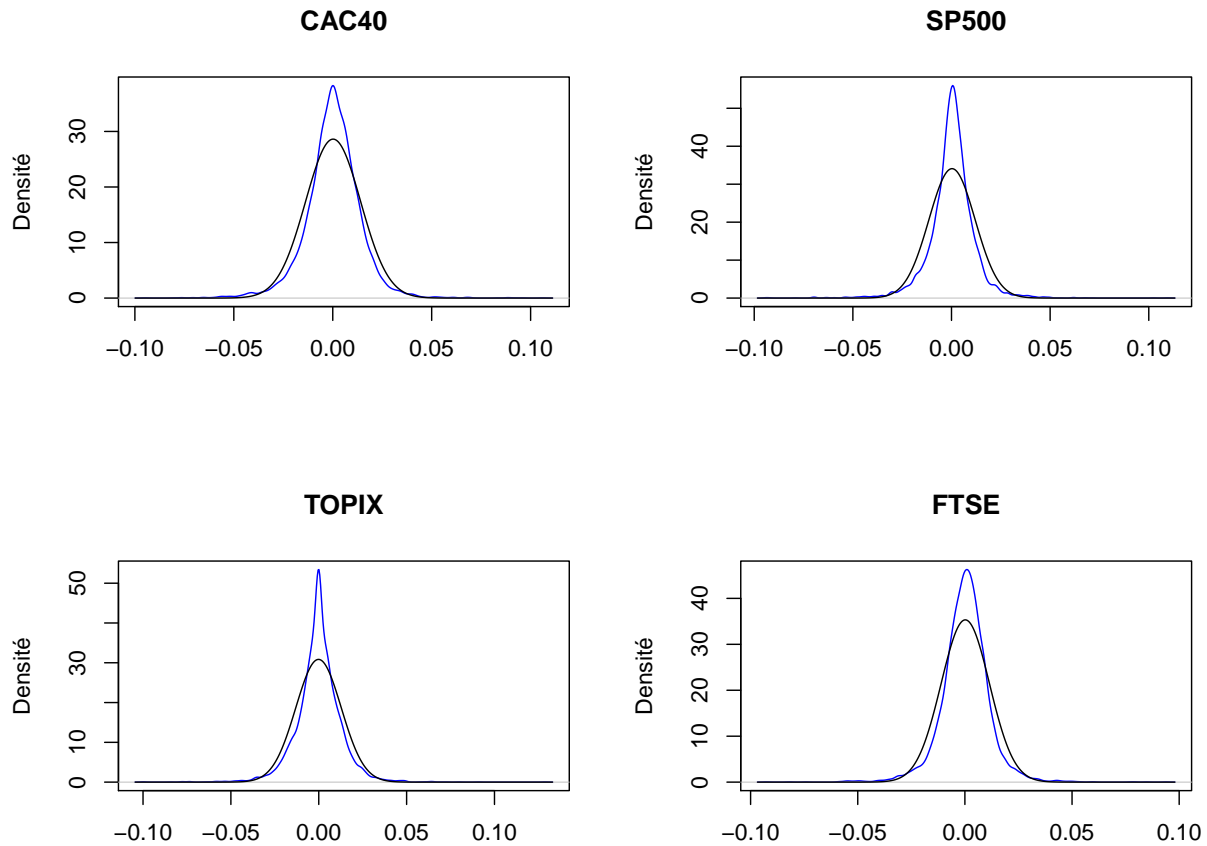


FIG. 1.3 – Densités des rentabilités journalières

- $\beta \in [-1, 1]$ est un coefficient d'asymétrie.
- $\gamma \in \mathbb{R}_+^*$ est un paramètre d'échelle.
- δ est un paramètre de localisation.

La famille des lois normales correspond au cas $\alpha = 2$ et $\beta = 0$. La famille des lois de Cauchy à $\alpha = 1$ et $\beta = 0$.

En dehors de ces deux sous-familles, les lois de Lévy n'ont de forme analytique simple pour leur densité, ce qui rend très délicats l'estimation des quatre paramètres qui les caractérisent. De plus, seule la sous-famille des lois normales admet des moments d'ordre 2 et il est donc impossible d'utiliser les résumés empiriques classiques que sont la variance ou l'écart-type. Malgré ces inconvénients, l'utilisation des lois stables de Lévy en finance revient périodiquement à la mode !

Les paramètres de ce modèle ont été estimés dans R (package *fBasics*, fonction *stableFit*). Le tableau (1.5) présente les résultats. Sans surprise, le paramètre $\alpha < 2$.

De nombreuses autres familles de lois ont été testées pour caractériser la distribution des rentabilités boursières. En particulier les lois de Student ou les lois de Pearson. Ces lois ont toutefois

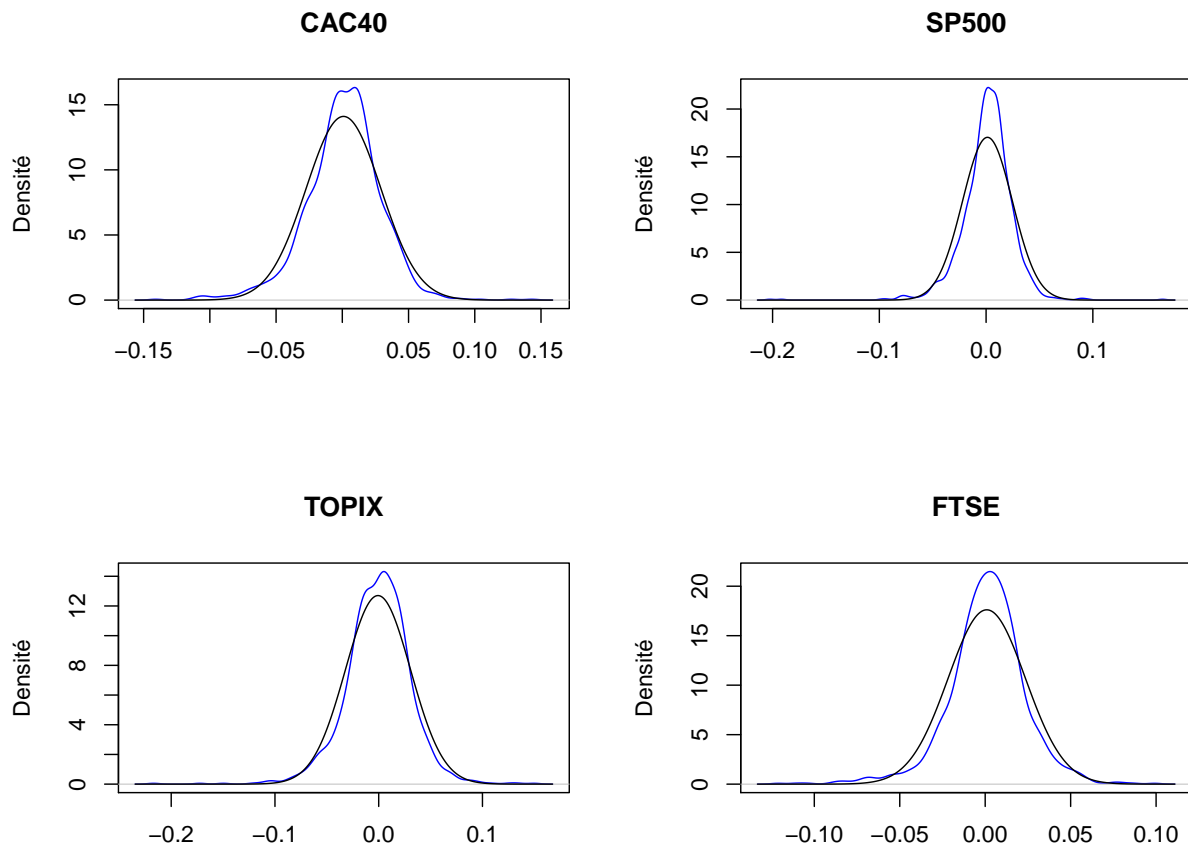


FIG. 1.4 – Densités des rentabilités hebdomadaires

l'inconvénient de ne pas être stables par convolution : si ce sont les rentabilités journalières qui suivent une loi de Student (ou de Pearson), les rentabilités hebdomadaires, mensuelles etc n'ont pas le même type de distribution.

1.4.9 Un autre modèle possible : loi normale contaminée

Pour rester dans le cadre statique et garder les distributions normales, on peut supposer que les rentabilités boursières sont soumises à deux régimes différents : un régime "normal" et un régime "de crise". Un modèle possible est alors :

$$R_t = \varepsilon_t X_t + (1 - \varepsilon_t) Y_t$$

α	β	γ	δ
1.60200	-0.12900	0.00754	0.00060

TAB. 1.5 – Paramètres de la loi de Lévy pour les rentabilités journalières du CAC 40

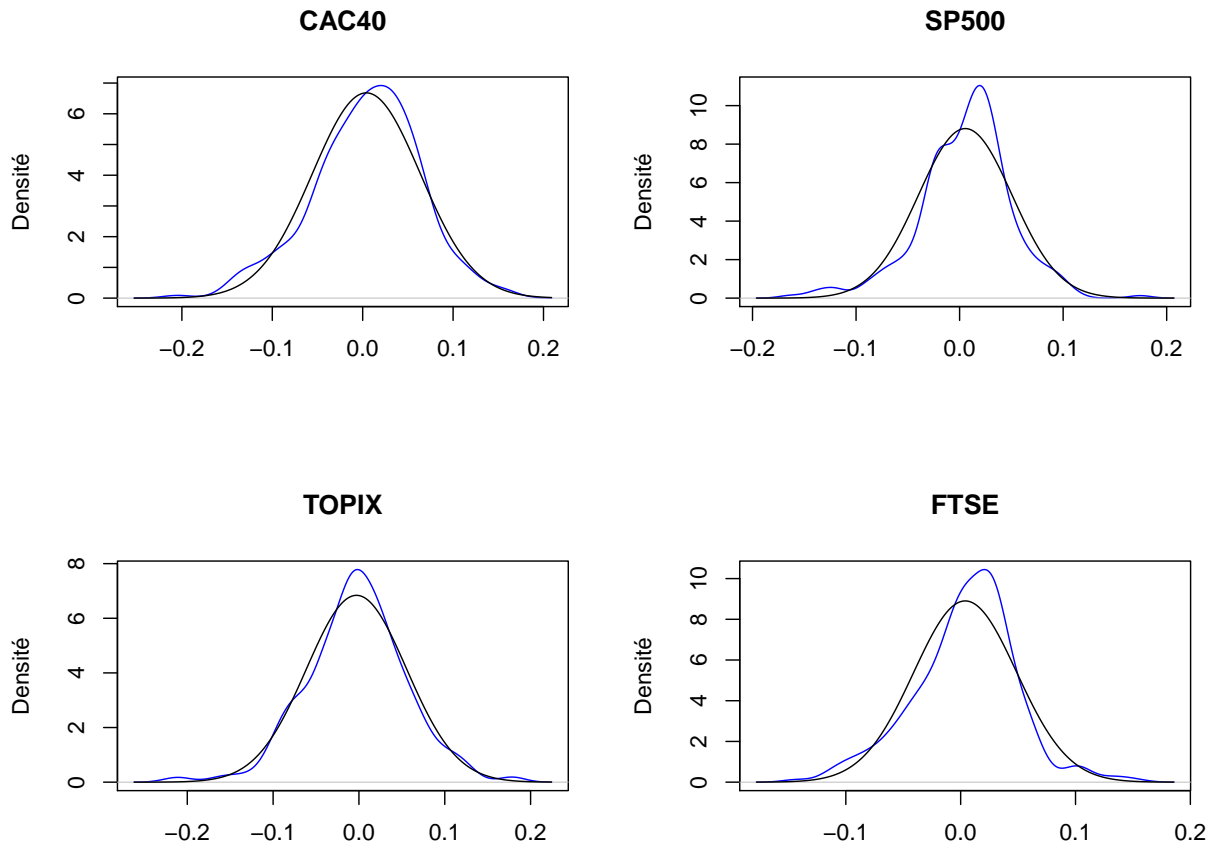


FIG. 1.5 – Densités des rentabilités mensuelles

où les variables X_t , Y_t et ε_t sont i.i.d. indépendantes entre elles et de lois respectives $\mathcal{N}(m_1, \sigma_1^2)$, $\mathcal{N}(m_2, \sigma_2^2)$ et $\mathcal{B}(1, p)$. ε_t représente le régime, X_t et Y_t les distributions des rentabilités dans les deux états.

Résultats probabilistes

On obtient alors les résultats suivants :

$$\begin{aligned}
 E[R_t] &= p m_1 + (1 - p) m_2 \\
 E[(R_t - E(R_t))^2] &= p(1 - p)(m_1 - m_2)^2 + p\sigma_1^2 + (1 - p)\sigma_2^2 \\
 E[(R_t - E(R_t))^3] &= (m_1 - m_2)p(1 - p)\left[(1 - 2p)(m_1 - m_2)^2 - 3(\sigma_1^2 - \sigma_2^2)\right] \\
 E[(R_t - E(R_t))^4] &= (m_1 - m_2)^2 p(1 - p)\left[(m_1 - m_2)^2(1 - 3p + 3p^2) + 6[(1 - p)\sigma_1^2 + p\sigma_2^2]\right] \\
 &\quad + 3(p\sigma_1^4 + (1 - p)\sigma_2^4)
 \end{aligned}$$

La distribution est symétrique quand $m_1 = m_2$ et dans ces conditions :

$$E \left[(R_t - E(R_t))^4 \right] - 3 E \left[(R_t - E(R_t))^2 \right]^2 = 3p(1-p) (\sigma_1^2 - \sigma_2^2)^2 > 0$$

La distribution est alors leptokurtique.

Dans le cas général, $m_1 \neq m_2$, on trouve :

$$\begin{aligned} E \left[(R_t - E(R_t))^4 \right] - 3 E \left[(R_t - E(R_t))^2 \right]^2 &= 6p(1-p)(1-2p)(\sigma_1^2 - \sigma_2^2)(m_1 - m_2)^2 \\ &\quad + 3p(1-p)(\sigma_1^2 - \sigma_2^2)^2 \\ &\quad + p(1-p)(6p^2 + 6p - 1)(m_1 - m_2)^4 \end{aligned}$$

Le premier terme de cette somme peut être négatif (il suffit de prendre $p > 0,5$ et $\sigma_1 > \sigma_2$). Il y a donc des cas où la distribution de R_t est *playkurtique*. Le graphique (FIG. 1.6) montre la densité de cette famille de lois.

Rappel : estimation par le maximum de vraisemblance

Définition 4 On appelle (*fonction de*) *vraisemblance* d'un paramètre $\theta \in \mathbb{R}^p, p \in \mathbb{N}^*$, pour les observations x_1, \dots, x_n , ($x_i \in \mathbb{R}^q, q \in \mathbb{N}^*$) la fonction

$$\mathcal{L}(x_1, \dots, x_n; \theta) : \theta \mapsto \mathcal{L}(x_1, \dots, x_n; \theta) = f_{(X_1, \dots, X_n)}(x_1, \dots, x_n; \theta)$$

où $f_{(X_1, \dots, X_n)}(x_1, \dots, x_n; \theta)$ est la densité du vecteur aléatoire (X_1, \dots, X_n) (qui dépend du paramètre θ).

On appelle *fonction score* d'un paramètre $\theta \in \mathbb{R}^p, p \in \mathbb{N}^*$, pour les observations x_1, \dots, x_n , le gradient du logarithme de la vraisemblance (ou log-vraisemblance) :

$$U : \theta \mapsto U(x_1, \dots, x_n; \theta) = \text{grad}_{\theta} \ln \mathcal{L}(x_1, \dots, x_n; \theta) = \frac{\partial \mathcal{L}}{\partial \theta}(x_1, \dots, x_n; \theta)$$

Si les vecteurs X_1, \dots, X_n sont deux à deux indépendants, la densité $f_{(X_1, \dots, X_n)}(x_1, \dots, x_n; \theta)$ et donc la vraisemblance s'écrivent comme un produit de densités, la fonction score comme une somme de vecteurs.

Définition 5 L'*estimateur de maximum de vraisemblance EMV* ou *MLE* (*maximum likelihood estimator*) est le vecteur de \mathbb{R}^p défini (lorsqu'il existe) par

$$\theta_n = \arg \max_{\theta \in \Theta} \mathcal{L}(x_1, \dots, x_n; \theta) = \arg \max_{\theta \in \Theta} \ln \mathcal{L}(x_1, \dots, x_n; \theta) \quad (1.1)$$

où $\Theta \subset \mathbb{R}^p$ est l'ensemble des valeurs admissibles du paramètre θ .

La condition du premier ordre implique que $U(x_1, \dots, x_n; \theta_n) = \frac{\partial \mathcal{L}}{\partial \theta}(x_1, \dots, x_n; \theta_n) = 0_{\mathbb{R}^p}$.

Le vecteur $\theta_n \in \mathbb{R}^p$ peut être considéré comme la réalisation du vecteur aléatoire $\hat{\theta}_n$ lorsque chacune des observations x_i est la réalisation du vecteur aléatoire X_i .

Par exemple, si (x_1, \dots, x_n) est un échantillon empirique provenant d'une loi $\mathcal{N}(m, \sigma^2)$, on a $\theta = (m, \sigma^2)$, $\theta_n = (\bar{x}_n, s_n^2)$ et $\hat{\theta}_n = (\bar{X}_n, S_n^2)$ avec

$$\begin{aligned}\bar{x}_n &= \frac{1}{n} \sum_{i=1}^n x_i \\ s_n^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \\ \bar{X}_n &= \frac{1}{n} \sum_{i=1}^n X_i \\ S_n^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2\end{aligned}$$

Remarque 3 *Attention : la plupart du temps, on appelle indifféremment estimateur de maximum de vraisemblance (EMV) le vecteur de nombres réels θ_n et le vecteur aléatoire $\hat{\theta}_n$.*

Définition 6 *On appelle **matrice d'information de Fisher** \mathcal{I}_n , la matrice de covariance du vecteur $U(X_1, \dots, X_n; \theta)$*

on obtient :

$$\mathcal{I}_n = E \left(\frac{\partial \ln \mathcal{L}}{\partial \theta} (X_1, \dots, X_n; \theta) \frac{\partial \ln \mathcal{L}}{\partial \theta'} (X_1, \dots, X_n; \theta) \right)$$

et sous certaines conditions de régularités :

$$\mathcal{I}_n = -E [\text{grad}_{\theta}(U)(X_1, \dots, X_n; \theta)] = -E \left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta \partial \theta'} (X_1, \dots, X_n; \theta) \right)$$

la notation $\frac{\partial \ln \mathcal{L}}{\partial \theta} \frac{\partial \ln \mathcal{L}}{\partial \theta'}$ signifie qu'on calcule la matrice de terme général $\left(\frac{\partial \ln \mathcal{L}}{\partial \theta_i} \frac{\partial \ln \mathcal{L}}{\partial \theta_j} \right)_{i,j=1,\dots,p}$.

De même, la notation $\frac{\partial^2 \ln \mathcal{L}}{\partial \theta \partial \theta'}$ signifie qu'on calcule la matrice de terme général $\left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta_i \partial \theta_j} \right)_{i,j=1,\dots,p}$.

Théorème 1 *L'estimateur de maximum de vraisemblance est asymptotiquement gaussien, sans biais et sa matrice de covariance asymptotique atteint la borne de Crámer - Rao : $K_n = \mathcal{I}_n^{-1}$.*

$$\mathcal{I}_n^{1/2} (\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}_{\mathbb{R}^p} (0, I_p)$$

Dans le cas d'un échantillon gaussien, on a

$$U(X_1, \dots, X_n, \theta) = \sum_{i=1}^n \left(\frac{X_i - m}{\sigma^2} - \frac{1}{2\sigma^2} \right)$$

et donc

$$\text{grad}_{\theta} U((X_1, \dots, X_n, \theta)) = - \sum_{i=1}^n \left(\begin{array}{cc} \frac{1}{\sigma^2} & \frac{X_i - m}{\sigma^4} \\ \frac{X_i - m}{\sigma^4} & \frac{(X_i - m)^2}{\sigma^6} - \frac{1}{2\sigma^4} \end{array} \right)$$

Et par conséquent :

$$\mathcal{I}_n = n \left(\begin{array}{cc} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{array} \right)$$

et

$$K_n = \frac{1}{n} \left(\begin{array}{cc} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{array} \right)$$

Estimation des paramètres

On dispose d'un échantillon (r_1, \dots, r_T) de T rentabilités. On ne sait pas *a priori* si l'observation t provient du premier ou du second régime. La log-vraisemblance de l'échantillon est donnée par :

$$L_{r_1, \dots, r_T}(m_1, \sigma_1, m_2, \sigma_2, p) = \sum_{t=1}^T \ln [p \phi_{m_1, \sigma_1}(r_t) + (1-p) \phi_{m_2, \sigma_2}(r_t)] \quad (1.2)$$

où $\phi_{m, \sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$ est la densité de la loi $\mathcal{N}(m, \sigma^2)$.

Grâce au théorème de Bayes, on obtient les résultats suivant :

$$P(\varepsilon_t = 1 | R_t = r_t; \theta, p) = \frac{p \phi_{m_1, \sigma_1}(r_t)}{p \phi_{m_1, \sigma_1}(r_t) + (1-p) \phi_{m_2, \sigma_2}(r_t)} \quad (1.3)$$

$$P(\varepsilon_t = 0 | R_t = r_t; \theta, p) = \frac{(1-p) \phi_{m_2, \sigma_2}(r_t)}{p \phi_{m_1, \sigma_1}(r_t) + (1-p) \phi_{m_2, \sigma_2}(r_t)} \quad (1.4)$$

où θ est le vecteur $^t(m_1, \sigma_1, m_2, \sigma_2)$.

On peut montrer que lorsque le maximum $(\hat{\theta}, \hat{p})$ de la log-vraisemblance (1.2) est atteint, ces estimateurs sont solution des équations :

$$\hat{p} = \frac{1}{T} \sum_{t=1}^T P(\varepsilon_t = 1 | R_t = r_t; \hat{\theta}, \hat{p}) \quad (1.5)$$

$$\hat{m}_1 = \frac{\sum_{t=1}^T P(\varepsilon_t = 1 | R_t = r_t; \hat{\theta}, \hat{p}) r_t}{\sum_{t=1}^T P(\varepsilon_t = 1 | R_t = r_t; \hat{\theta}, \hat{p})} \quad (1.6)$$

$$\hat{\sigma}_1^2 = \frac{\sum_{t=1}^T P(\varepsilon_t = 1 | R_t = r_t; \hat{\theta}, \hat{p}) r_t^2}{\sum_{t=1}^T P(\varepsilon_t = 1 | R_t = r_t; \hat{\theta}, \hat{p})} - \hat{m}_1^2 \quad (1.7)$$

$$\hat{m}_2 = \frac{\sum_{t=1}^T P(\varepsilon_t = 0 | R_t = r_t; \hat{\theta}, \hat{p}) r_t}{\sum_{t=1}^T P(\varepsilon_t = 0 | R_t = r_t; \hat{\theta}, \hat{p})} \quad (1.8)$$

$$\hat{\sigma}_2^2 = \frac{\sum_{t=1}^T P(\varepsilon_t = 0 | R_t = r_t; \hat{\theta}, \hat{p}) r_t^2}{\sum_{t=1}^T P(\varepsilon_t = 0 | R_t = r_t; \hat{\theta}, \hat{p})} - \hat{m}_2^2 \quad (1.9)$$

Ceci conduit naturellement à un algorithme de type "point fixe" pour résoudre les équations (1.5 - 1.9). Cette méthode est l'application de la méthode "algorithme EM (Expectation-Maximization)", développée par Dempster, Laird, et Rubin (1977), pour les lois normales contaminées.

Si on note $l(R_t; m_1, \sigma_1^2, m_2, \sigma_2^2, p) = \ln [p \phi_{m_1, \sigma_1}(R_t) + (1-p) \phi_{m_2, \sigma_2}(R_t)]$, on obtient :

$$\text{grad} l(R_t; m_1, \sigma_1^2, m_2, \sigma_2^2, p) = \frac{1}{p \phi_{m_1, \sigma_1}(R_t) + (1-p) \phi_{m_2, \sigma_2}(R_t)} \begin{pmatrix} \phi_{m_1, \sigma_1}(R_t) - \phi_{m_2, \sigma_2}(R_t) \\ p \frac{R_t - m_1}{\sigma_1^2} \phi_{m_1, \sigma_1}(R_t) \\ p \frac{(R_t - m_1)^2 - \sigma_1^2}{2 \sigma_1^4} \phi_{m_1, \sigma_1}(R_t) \\ (1-p) \frac{R_t - m_2}{\sigma_2^2} \phi_{m_2, \sigma_2}(R_t) \\ (1-p) \frac{(R_t - m_2)^2 - \sigma_2^2}{2 \sigma_2^4} \phi_{m_2, \sigma_2}(R_t) \end{pmatrix}$$

m_1	0.069%	m_1 annualisé	17.316%
IC(99%)	[0.069%, 0.070%]		[17.206%, 17.426%]
σ_1	1.005%	σ_1 annualisé	15.891%
IC(99%)	[0.959%, 1.049%]		[15.170%, 16.581%]
m_2	-0.250%	m_2 annualisé	-62.475%
IC(99%)	[-0.253%, -0.247%]		[-63.144%, -61.805%]
σ_2	2.595%	σ_2 annualisé	41.031%
IC(99%)	[2.391%, 2.784%]		[37.806%, 44.021%]
p	83.936%		
IC(99%)	[80.072%, 87.801%]		

TAB. 1.6 – Résultats de l'algorithme EM pour les rentabilités journalières du CAC 40

La matrice de Fisher \mathcal{I}_T vaut alors

$$\mathcal{I} = T E \left[\text{gradl}(R_t; m_1, \sigma_1^2, m_2, \sigma_2^2, p) \text{gradl}(R_t; m_1, \sigma_1^2, m_2, \sigma_2^2, p)' \right]$$

et est estimée par

$$\bar{\mathcal{I}}_T = \sum_{t=1}^T \text{gradl}(r_t; \hat{m}_1, \hat{\sigma}_1^2, \hat{m}_2, \hat{\sigma}_2^2, \hat{p}) \text{gradl}(r_t; \hat{m}_1, \hat{\sigma}_1^2, \hat{m}_2, \hat{\sigma}_2^2, \hat{p})'$$

Une fois obtenus les paramètres $(\hat{p}, \hat{\theta})$, il est naturel de décider si l'observation t provient du premier ou du second régime, c'est à dire d'estimer $\hat{\varepsilon}_t$ par :

$$\hat{\varepsilon}_t = \begin{cases} 1 & \text{si } \hat{p} \phi_{\hat{m}_1, \hat{\sigma}_1}(r_t) > (1 - \hat{p}) \phi_{\hat{m}_2, \hat{\sigma}_2}(r_t) \\ 0 & \text{sinon} \end{cases}$$

Ceci permet de dessiner les trajectoires des rentabilités boursières pour les deux régimes.

Résultats pour les rentabilités journalières de l'indice CAC 40

Le tableau (TAB. 1.6) donne la valeur des estimateurs EM pour les rentabilités journalières de l'indice CAC 40. On peut remarquer que les deux régimes sont caractérisés d'une part par un état "normal" (84% du temps, avec une moyenne positive de l'ordre de 17% annuelle et une "petite" volatilité de l'ordre de 16% annuelle) et d'autre part par état "de crise" où la moyenne des rentabilités est très négative (-62% par an) et où la volatilité est très forte (41% par an).

Cette interprétation (état normal - état de crise) est confirmée par le graphique (FIG. 1.7) qui montre qu'effectivement les périodes troublées correspondent au régime 2 et les périodes calmes au régime 1.

Pour juger (grossièrement) de la qualité de l'estimation, on peut voir sur le graphique (FIG. 1.8) que les densités empiriques (estimée par un noyau gaussien) et la densité de la loi contaminée dont les paramètres sont ceux qui ont été calculés par l'algorithme EM sont extrêmement proches.

Cependant, un examen plus attentif du graphique (FIG. 1.7) conduit à une nouvelle remarque : les périodes "de crise" (ou celles "normale") ont tendance à perdurer. Ceci signifie qu'il doit y

avoir **dépendance** entre les rentabilités successives. En se rappelant que les moments estimés (*cf.* tableaux 1.2, 1.3, 1.4) varient fortement d’une période à l’autre, et que rien, dans la théorie financière (elle dit plutôt le contraire), interdit aux investisseurs de réviser leur jugement sur les performances futures des sociétés, l’hypothèse statique est fortement remise en cause.

1.5 Série chronologique des rentabilités boursières

La conclusion du paragraphe précédent impose de traiter les rentabilités boursières comme un processus stochastique. La distribution de $R(t)$ va donc dépendre de t .

Quelles sont alors les “contraintes” qu’il faut s’imposer pour espérer modéliser “correctement” ce processus des rentabilités boursières?

- *Contraintes liées à la théorie financière :*
 - Le paradigme *espérance de gain - risque* est incontournable. Il faut donc être capable, à partir du modèle, de définir comment les investisseurs vont se positionner dans cet espace “espérance de gain - risque”.
 - Il ne doit pas être possible de faire des profits d’arbitrage purs, ni même de construire des stratégies, basées sur les données du passé, durablement gagnantes (efficience des marchés).
- *Contraintes “statistiques :”*
 - Le modèle doit ‘expliquer ce qui est observé. Par exemple, il doit conduire à une distribution leptokurtique des rentabilités.
 - Il faut que le modèle soit *bien spécifié*, c’est à dire qu’on puisse l’estimer.
 - Les paramètres du modèle (s’il y en a) doivent être stables dans le temps. On peut imaginer des “chocs” qui feraient varier brutalement ces paramètres (décisions financières propres à l’entreprise, phénomènes économiques généraux, tensions politiques ...) mais pas une révision importante en continue de ces paramètres.

Pour rester dans un cadre strict d’efficience des marchés à la Fama, supposons que les rentabilités soit indépendantes. Une manière assez simple d’envisager la “non distribution identique” des rentabilités est d’écrire le modèle suivant :

$$R(t) = m(t) + \sigma(t) \varepsilon(t) \quad (1.10)$$

où $m(t), \sigma(t)$ sont deux fonctions déterministes du temps et $(\varepsilon(t), t = 0 \dots)$ un bruit blanc de variance égale à 1.

Dans ces conditions, on a évidemment $E[R(t)] = m(t)$ et $Var[R(t)] = \sigma^2(t)$.

L’objet de ce paragraphe est d’étudier le processus des rentabilités boursières comme un processus ARMA. Les résultats théoriques sur ces processus sont résumés dans le paragraphe 1.6.

Les deux graphiques montrent que seuls les coefficients d’autocorrélation ou d’autocorrélation partielle 3, 5 et 15 sont significativement non nuls.

Les différentes méthodes d’identification conduisent à choisir un modèle ARMA(1,3) pour lequel le coefficient MA(2) est nul. Traité dans R, on obtient le résultat suivant :

```
>library(tseries)
```

```

> armaCac<-arma(Cac40$Rent, lag=list(ar=c(1), ma=c(1,3)))
Warning message:
order is ignored in: arma(Cac40$Rent, lag = list(ar = c(1), ma = c(1, 3)))
> summary(armaCac)

Call:
arma(x = Cac40$Rent, lag = list(ar = c(1), ma = c(1, 3)))

Model:
ARMA(1,3)

Residuals:
      Min       1Q   Median       3Q      Max
-0.0769813 -0.0068277  0.0001378  0.0073786  0.0694881

Coefficient(s):
      Estimate Std. Error t value Pr(>|t|)
ar1      -0.6491047   0.1560967  -4.158 3.21e-05 ***
ma1       0.6748511   0.1543271   4.373 1.23e-05 ***
ma3      -0.0486495   0.0122435  -3.974 7.08e-05 ***
intercept 0.0005634   0.0003106   1.814  0.0697 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Fit:
sigma^2 estimated as 0.0001673, Conditional Sum-of-Squares = 0.79, AIC = -27623.4

> R2Cac<-1-var(armaCac$residuals[!is.na(armaCac$residuals)])/var(Cac40$Rent)
> R2Cac
[1] 0.004925224

```

$$(Id - \{-0.649\} B)(X_t - 0.0006) = (Id + \{0.675\} B + \{-0.049\} B^3) \varepsilon_t$$

Le coefficient de détermination estimé est de l'ordre de $R^2 = 0.0049$, ce qui signifie que ce modèle est inutile pour la prévision (*cf.* 1.11)

En fait, ce type de modèle est très mal adapté pour représenter les séries de rentabilités boursières. Les résultats du paragraphe (1.4.9) avaient montré que les rentabilités de grande amplitude (respectivement de petite amplitude) étaient souvent suivies de rentabilités de grande amplitude (respectivement de petite amplitude). Ceci nous conduit à penser que ce ne sont pas les rentabilités elles-mêmes qui sont autocorrélées, mais leurs valeurs absolues (ou leurs carrées ou autres transformations non linéaires). Il n'est donc pas surprenant d'obtenir de très mauvais résultats lorsqu'on cherche un modèle linéaire (même infini !) pour représenter ces séries.

Le chapitre suivant s'intéresse donc à une modélisation plus conforme aux observations, des séries boursières.

1.6 Séries chronologiques classiques

1.6.1 Généralités sur les processus du second ordre

Proposition 4 Si $(X_t, t \in \mathbb{Z})$ est un processus stationnaire (au second ordre) et si $(a_i)_{i \in \mathbb{Z}}$ est une suite de nombres réels absolument convergente ($\sum_{i=-\infty}^{+\infty} |a_i| < +\infty$) alors le processus $(Y_t, t \in \mathbb{Z})$ défini par

$$\forall t \in \mathbb{Z}, Y_t = \sum_{i=-\infty}^{+\infty} a_i X_{t-i}$$

est un processus stationnaire.

Définition 7 (Opérateur retard) L'opérateur retard B associe à un processus $(X_t, t \in \mathbb{Z})$, le processus $(Y_t, t \in \mathbb{Z})$ défini par

$$\forall t \in \mathbb{Z}, Y_t = B X_t = X_{t-1}$$

Si on note $\mathcal{S}(\Omega, \mathcal{F}, P)$ l'espace vectoriel des processus stationnaires (au second ordre), alors B est un opérateur linéaire sur $\mathcal{S}(\Omega, \mathcal{F}, P)$, inversible et son inverse est l'opérateur avant défini par

$$\forall t \in \mathbb{Z}, F X_t = X_{t+1}$$

Définition 8 (Séries en B) Soit $(a_i)_{i \in \mathbb{Z}}$ une suite de nombres réels absolument convergente. On définit l'opérateur $\sum_{i=-\infty}^{+\infty} a_i B^i$ par

$$\forall X = (X_t, t \in \mathbb{Z}) \in \mathcal{S}(\Omega, \mathcal{F}, P), \left(\sum_{i=-\infty}^{+\infty} a_i B^i \right) X = Y \in \mathcal{S}(\Omega, \mathcal{F}, P)$$

défini par

$$\forall t \in \mathbb{Z}, Y_t = \sum_{i=-\infty}^{+\infty} a_i X_{t-i}$$

Ces opérateurs séries en B ont des propriétés qui vont permettre de les manier comme des séries entières habituelles. En particulier, on peut en faire des combinaisons linéaires, la composée de deux séries en B est aussi une série en B et leur produit est commutatif.

Proposition 5 (inversion de $\text{Id} - \lambda B$) L'opérateur $\text{Id} - \lambda B, \lambda \in \mathbb{R}$ est inversible si et seulement si $|\lambda| \neq 1$. Son inverse est :

$$\begin{aligned} & \sum_{i=0}^{+\infty} \lambda^i B^i \quad \text{si } |\lambda| < 1 \\ & - \sum_{i=-\infty}^{-1} \lambda^i B^i \quad \text{si } |\lambda| > 1 \end{aligned}$$

On a le même résultat en remplaçant l'opérateur retard B par l'opérateur avance F .

Si $|\lambda| < 1$, le processus $Y = (Y_t = (Id - \lambda B)^{-1} X_t, t \in \mathbb{Z})$ n'est défini qu'à l'aide du passé du processus $X = (X_t, t \in \mathbb{Z})$. Si $|\lambda| > 1$, Y est défini à l'aide des valeurs futures de X .

Proposition 6 (Inversion d'un polynôme) Soit Φ le polynôme défini par $\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ dont les racines (complexes) $z_j = \frac{1}{\lambda_j}$ triées par module décroissant, sont toutes de module différentes de 1. Alors l'opérateur $\Phi(B) = Id - \phi_1 B - \dots - \phi_p B^p = \prod_{j=1}^p (Id - \lambda_j B)$ est inversible.

Supposons que les $r \neq 0$ première(s) racine(s) soient de module strictement supérieur à 1, et les $p - r$ restante(s) de module strictement inférieur à 1. On peut écrire

$$\Phi(B) = \prod_{i=1}^r (Id - \lambda_i B) \prod_{i=r+1}^p \left(Id - \frac{1}{\lambda_i} B \right) \prod_{i=r+1}^p -\lambda_i B$$

Chaque terme de ce produit est inversible et son inverse est donné par la proposition 5.

1.6.2 Processus autorégressifs

Définition 9 (processus autorégressifs) On appelle processus autorégressif d'ordre $p \in \mathbb{N}^*$ ($AR(p)$), un processus stationnaire (au second ordre) qui vérifie :

$$\forall t \in \mathbb{Z}, X_t = \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t \quad (1.11)$$

où $(\phi_i) \in \mathbb{R}^p$, et $(\varepsilon_t, t \in \mathbb{Z})$ est un bruit blanc.

Si on note $\Phi(z)$ le polynôme $\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$, on obtient :

$$\Phi(B) X_t = \varepsilon_t$$

Propriétés des processus autorégressifs

1. On peut ajouter une constante $\phi_0 \in \mathbb{R}$ à l'équation (1.11):

$$\forall t \in \mathbb{Z}, X_t = \phi_0 + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t \quad (1.12)$$

Dans ces conditions, la stationnarité impose : $E[X_t] = \phi_0 + \sum_{i=1}^p \phi_i E[X_t]$ et donc le processus $Y_t, t \in \mathbb{Z}$ défini par $\forall t \in \mathbb{Z}, Y_t = X_t - E[X_t]$ est un processus stationnaire vérifiant l'équation (1.11). On peut donc, quitte à le centrer, considérer que le processus autorégressif est centré.

2. Si toutes les racines du polynôme Φ sont de module strictement supérieur à 1 alors on peut écrire $\Psi(B) = \Phi^{-1}(B) = \sum_{i=0}^{+\infty} \psi_i B^i$ et il vient donc :

$$X_t = \sum_{i=0}^{+\infty} \psi_i \varepsilon_{t-i}$$

Le processus (ε_t) est alors appelé *innovation du processus* X_t car alors les ε_t sont non corrélés avec les X_s pour $s \leq t - 1$.

3. Quitte à changer de bruit blanc, on peut toujours supposer que les racines de Φ sont de module strictement supérieur à 1. Cela implique qu'on peut avoir de multiples représentations du processus X et qu'il faut choisir entre elles. La représentation qui ne fait intervenir que les racines de module strictement supérieur à 1 s'impose et s'appelle *représentation canonique du processus*.
4. Si au moins une des racines de Φ est de module égal à 1, le processus n'est pas stationnaire.

1.6.3 Processus moyenne mobile

Définition 10 (Processus $\text{MA}(\mathbf{q})$) On appelle processus moyenne mobile d'ordre $q > 0$ ($\text{MA}(q)$) un processus $X = (X_t, t \in \mathbb{Z})$ qui vérifie une relation du type

$$\forall t \in \mathbb{Z}, X_t = \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}$$

où $(\theta_j) \in \mathbb{R}^q$ et $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ est un bruit blanc.

Si on note Θ le polynôme $\Theta(z) = 1 - \theta_1 z - \dots - \theta_q z^q$, on peut écrire :

$$\forall t \in \mathbb{Z}, X_t = \Theta(B) \varepsilon_t \quad (1.13)$$

Contrairement aux processus $\text{AR}(p)$, la définition du processus $\text{MA}(q)$ est explicite. Le processus X est parfaitement déterminé par la relation (1.13) et est automatiquement stationnaire.

Propriétés des processus $\text{MA}(q)$

1. Si les racines de Θ sont de modules strictement supérieur à 1, alors

$$\forall t \in \mathbb{Z}, \varepsilon_t = \Pi(B) X_t = \sum_{j=0}^{+\infty} \pi_j X_{t-j}$$

avec $\pi_0 = 1$ et $\sum_{j=0}^{+\infty} |\pi_j| < +\infty$. ε_t est alors l'innovation du processus X .

2. Quitte à changer de bruit blanc et si Θ n'a pas de racine de module égal à 1, on peut toujours supposer que ses sont de modules strictement supérieur à 1. On obtient alors la *représentation canonique* du processus $\text{MA}(q)$.

1.6.4 Processus $\text{ARMA}(p, q)$

Définition 11 On appelle processus $\text{ARMA}(p, q)$ un processus stationnaire $X = (X_t, t \in \mathbb{Z})$ qui vérifie une relation du type

$$\forall t \in \mathbb{Z}, X_t - \sum_{i=1}^p \phi_i X_{t-i} = \theta_0 + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}$$

où $\theta_0 \in \mathbb{R}$, $(\theta_j) \in \mathbb{R}^q$, $(\phi_i) \in \mathbb{R}^p$ et $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ est un bruit blanc.

En posant $\Theta(z) = 1 - \sum_{j=1}^q \theta_j z^j$ et $\Phi(z) = 1 - \sum_{i=1}^p \phi_i z^i$, il vient :

$$\forall t \in \mathbb{Z}, \Phi(B) X_t = \theta_0 + \Theta(B) \varepsilon_t \quad (1.14)$$

Quitte à centrer le processus, on peut toujours se ramener au cas où $\theta_0 = 0$.

Propriétés des processus ARMA

1. Si les racines de Φ sont de modules strictement supérieur à 1, il existe une suite absolument convergente (Ψ_i) de nombres réels tels que :

$$\forall t \in \mathbb{Z}, X_t = \varepsilon_t + \sum_{i=1}^{+\infty} \Psi_i \varepsilon_{t-i}$$

2. Si les racines de Θ sont de modules strictement supérieur à 1, il existe une suite absolument convergente (Π_i) de nombres réels tels que :

$$\forall t \in \mathbb{Z}, X_t = \varepsilon_t + \sum_{i=1}^{+\infty} \Pi_i X_{t-i}$$

3. Si les racines de $\Phi(z)$ et de $\Theta(z)$ sont de modules strictement supérieur à 1, alors (ε_t) est l'innovation du processus.
4. Le processus $(X_t, t \in \mathbb{Z})$ admet une *représentation* ARMA(p, q) *minimale* si on peut écrire $\Phi(B) X_t = \Theta(B) \varepsilon_t$ où Φ est un polynôme de degré p , Θ un polynôme de degré q , les racines de ces deux polynômes sont de module strictement supérieur à 1, ils n'ont pas de racine(s) commune(s) et (ε_t) est un bruit blanc.

1.6.5 Processus ARIMA(p, d, q)

Pour de nombreuses séries, dont la plupart des séries de prix, l'hypothèse de stationnarité n'est pas réaliste (voire contraire à la théorie économique). Par contre, si on considère les différences premières (ou plus généralement les différences d'ordre $d \geq 1$) d'une série alors cette hypothèse devient plus vraisemblable.

Définition 12 (processus ARIMA(p, d, q)) Un processus $X = (X_t, t \in \mathbb{Z})$ admet une *représentation* ARIMA(p, d, q) si le processus $Y = \Delta^d X = (Id - B)^d X$ admet une *représentation* ARMA(p, q).

La notation “polynomiale” d'un processus ARIMA(p, d, q) est donc

$$\Phi(B) \Delta^d X_t = \Theta(B) \varepsilon_t$$

Comme pour les processus AR(p), MA(q) ou ARMA(p, q), quitte à changer de bruit blanc et à condition que les deux polynômes $\Theta(z)$ et $\Phi(z)$ n'aient pas de racines de module égal à 1, on a la représentation canonique du processus lorsque ces mêmes racines sont toutes de module strictement supérieur à 1.

1.6.6 Autocorrélation, autocorrélation partielle

Définition 13 (fonction d'autocorrélation) Soit $(X_t, t \in \mathbb{Z})$ un processus stationnaire (au second ordre). On appelle fonction d'autocorrélation ρ du processus :

$$\rho : \begin{array}{ll} \mathbb{Z} & \rightarrow [-1, 1] \\ h & \mapsto \frac{\text{cov}(X_t, X_{t+h})}{\text{var}(X_t)} = \frac{\gamma(h)}{\gamma(0)} \end{array}$$

Les fonctions d'autocovariance $\gamma(h)$ et d'autocorrélation sont paires et on peut donc restreindre leurs définitions à \mathbb{N} .

Définition 14 (fonction d'autocorrélation partielle) Soit $(X_t, t \in \mathbb{Z})$ un processus stationnaire (au second ordre). On appelle fonction d'autocorrélation partielle τ du processus :

$$\tau : \begin{array}{ll} \mathbb{Z} & \rightarrow [-1, 1] \\ h & \mapsto \frac{\text{cov}(X_t - X_t^*, X_{t+h} - X_{t+h}^*)}{\text{var}(X_t - X_t^*)} \end{array}$$

où X_t^* (respectivement X_{t+h}^*) est la régression affine de X_t (respectivement X_{t+h}) sur $X_{t-1}, \dots, X_{t-h+1}$.

Proposition 7 On peut démontrer la relation suivante :

$$\begin{bmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \\ \vdots \\ \rho(h) \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(h-1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(h-1) \\ \rho(2) & \rho(1) & 1 & \dots & \rho(h-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho(h-1) & \rho(h-2) & \rho(h-3) & \dots & 1 \end{bmatrix} \begin{bmatrix} \tau(1) \\ \tau(2) \\ \tau(3) \\ \vdots \\ \tau(h) \end{bmatrix} \quad (1.15)$$

On a, en particulier $\rho(1) = \tau(1)$. Les coefficients d'autocorrélation partielle sont donc des combinaisons linéaires des coefficients d'autocorrélation et vice-versa.

Proposition 8 (caractérisation des processus AR et MA) Les fonctions d'autocorrélation et d'autocorrélation partielle permettent de caractériser les processus AR et MA.

1. Si $(X_t, t \in \mathbb{Z})$ admet une représentation $\text{AR}(p)$, alors $\forall h > p, \tau(h) = 0$.
2. si $(X_t, t \in \mathbb{Z})$ admet une représentation $\text{MA}(q)$, alors $\forall h > q, \rho(h) = 0$.

Si on note pour $i \geq 0$ et $j \geq 1$, $\Delta(i, j)$ le déterminant

$$\Delta(i, j) = \begin{vmatrix} \rho(i) & \rho(i-1) & \dots & \rho(i-j+1) \\ \rho(i+1) & \rho(i) & \dots & \rho(i-j+2) \\ \vdots & \vdots & \dots & \vdots \\ \rho(i+j-1) & \rho(i+j-2) & \dots & \rho(i) \end{vmatrix}$$

on obtient le résultat suivant :

Proposition 9 (caractérisation des processus ARMA) Le processus $(X_t, t \in \mathbb{Z})$ admet une représentation $\text{ARMA}(p, q)$ minimale si :

$$\begin{aligned} \forall i \geq q+1, j \geq p+1, \Delta(i, j) &= 0 \\ \forall i \geq q, \Delta(i, p) &\neq 0 \\ \forall j \geq p, \Delta(q, j) &\neq 0 \end{aligned}$$

1.6.7 Estimation des paramètres des modèles ARMA

Dans ce paragraphe, on dispose de T observations successives (x_1, \dots, x_T) d'un processus $\text{ARMA}(p, q)$, où p et q sont connus.

Méthodes de pseudo-maximum de vraisemblance

Supposons que la série X_t soit un MA(1). Il s'agit d'estimer le paramètre θ_1 à partir des observations. Du modèle $X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$, on déduit $x_t = e_t + \theta_1 e_{t-1}$ où e_t est la réalisation du bruit ε_t à la date t . Si on fixe $\varepsilon_0 = 0$, on peut déterminer tous les e_t par récurrence en fonction des valeurs observées x_t . Il vient :

$$\forall t \geq 1, e_t = \sum_{i=0}^{t-1} \theta_1^i x_{t-i} \quad (1.16)$$

Comme pour les modèles linéaires, la représentation MA(1) est d'autant plus efficace que la variance du bruit est petite. On estimera alors θ_1 par :

$$\theta_1 = \text{Arg min}_{\theta} S_C(\theta)$$

où

$$S_C(\theta) = \sum_{t=1}^T e_t^2 = \sum_{t=1}^T \left(\sum_{i=0}^{t-1} \theta_1^i x_{t-i} \right)^2$$

$S_C(\theta)$ est la somme des carrés conditionnelle à $\varepsilon_0 = 0$.

Dans le cas général d'un ARMA(p, q) $X_t - \sum_{i=1}^p \phi_i X_{t-i} = \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}$, on fixe les conditions initiales $\varepsilon_p = e_p = 0, \varepsilon_{p-1} = e_{p-1} = 0, \dots, \varepsilon_{p-q+1} = e_{p-q+1} = 0$ et $X_1 = x_1, \dots, X_p = x_p$ puis on détermine par récurrence

$$\forall t \geq p+1, e_t = x_t - \sum_{i=1}^p \phi_i x_{t-i} + \sum_{j=1}^q \theta_j e_{t-j}$$

e_t dépend donc des paramètres $\beta = ((\phi_i)_i, (\theta_j)_j)$, on calcule la somme des carrés conditionnelle

$$S_C(\beta) = \sum_{t=p+1}^T e_t^2 \text{ et on estime}$$

$$\hat{\beta} = \text{Arg min}_{\beta} S_C(\beta)$$

Cette méthode a été améliorée par Box et Jenkins (1976). Si on reprend le modèle MA(1), on peut écrire : $X_t = \varepsilon'_t - \theta_1 \varepsilon'_{t+1}$ où les ε'_t sont les innovations en temps inversé. En posant alors $e'_T = x_T$ et $e'_t = x_{t+1} + \theta_1 e'_{t+1}$ pour $t \leq T-1$. Ceci permet de trouver e'_1 puis la prévision rétrospective (ou arrière) $\hat{X}_0 = -\theta_1 e'_1$. Il suffit alors de poser $e_0 = \hat{X}_0$ et de recalculer $S_C(\theta_1)$, en utilisant l'équation (1.16). Ceci conduit à une nouvelle estimation de $\hat{\theta}_1$ en minimisant $S_C(\theta_1)$ et on peut ré-itérer cette opération.

Si on suppose connues les valeurs initiales (x^*, e^*) nécessaires pour amorcer les récurrences et si on suppose que $\varepsilon_t \rightsquigarrow NIID(0, \sigma^2)$ alors la vraisemblance conditionnelle de (x_1, \dots, x_T) est donnée par :

$$L_C(\beta, \sigma^2) = \frac{1}{(2\pi)^{T/2} \sigma^T} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^T e_t^2(\beta) \right\} \quad (1.17)$$

Le maximum de cette vraisemblance conditionnelle s'obtient aisément :

$$\hat{\beta} = \min_{\beta} S_C(\beta) \text{ et } \hat{\sigma}^2 = \frac{1}{T} S_C(\hat{\beta})$$

La vraisemblance non conditionnelle

On fait toujours l'hypothèse que le bruit blanc ε est gaussien. Dans ces conditions, (X_1, \dots, X_T) est un vecteur gaussien dont la matrice de covariance est donnée à l'aide de la matrice :

$$\Gamma = \begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(T-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(T-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(T-1) & \gamma(T-2) & \dots & \gamma(0) \end{pmatrix}$$

où $\gamma(h)$ est la fonction d'autocovariance du processus $(Y_t = \frac{X_t}{\sigma}, t \in \mathbb{Z})$ (où σ^2 est la variance de ε_t). Pour les modèles ARMA, chaque $\gamma(h)$ dépend du paramètre β et donc $\Gamma = \Gamma(\beta)$. Avec ces notations, on montre que la fonction de vraisemblance de (x_1, \dots, x_T) est :

$$L_V(\beta, \sigma^2) = \frac{1}{(2\pi)^{T/2} \sigma^T (\det \Gamma(\beta))^{1/2}} \exp \left\{ -\frac{1}{2\sigma^2} {}^t x \Gamma(\beta)^{-1} x \right\} \quad (1.18)$$

où $x = {}^t(x_1, \dots, x_p)$.

Comme pour les modèles linéaires généralisés, le maximum de vraisemblance est solution de :

$$\hat{\beta} = \text{Arg min}_{\beta} S_V(\beta) \text{ et } \hat{\sigma}^2 = \frac{1}{T} {}^t x \Gamma^{-1}(\hat{\beta}) x$$

où $S_C(\beta) = (\det \Gamma(\beta))^{1/T} ({}^t x \Gamma^{-1}(\beta) x)$. Le calcul du maximum de vraisemblance est ardu puisque à chaque étape de l'algorithme de maximisation, il demande d'inverser et de calculer le déterminant d'une matrice carrée d'ordre T . Il existe toutefois des algorithmes complexes qui permettent d'obtenir assez rapidement ces estimateurs.

Propriétés asymptotiques des estimateurs

Lorsque $T \rightarrow +\infty$, les différentes méthodes d'estimations sont équivalentes. Sous certaines conditions générales, ces estimateurs sont asymptotiquement sans biais et asymptotiquement normalement distribués. En particulier, les estimateurs du pseudo-maximum de vraisemblance du processus gaussien (1.17) ont les "bonnes" propriétés même si le bruit ε_t n'est pas normalement distribué. Pour obtenir la matrice de covariance asymptotique des estimateurs, on utilise la matrice d'information de Fisher et

$$\begin{aligned} V_{as}(\hat{\beta}) &= - \left\{ E \left[\frac{\partial^2 \ln L_C(\beta)}{\partial \beta \partial \beta'} \right] \right\}^{-1} \\ &= 2\sigma^2 \left\{ E \left[\frac{\partial^2 S_C(\beta)}{\partial \beta \partial \beta'} \right] \right\}^{-1} \end{aligned}$$

Comme $E \left[e_t(\beta) \frac{\partial^2 e_t(\beta)}{\partial \beta \partial \beta'} \right] = 0$, car le terme ne fait intervenir que des e_{t-i} pour $i > 0$ et que les e_t sont 2 à 2 non corrélés, on trouve

$$E \left[\frac{\partial^2 S_C(\beta)}{\partial \beta \partial \beta'} \right] = 4 \sum_{t=1}^T E \left\{ \left(\frac{\partial e_t(\beta)}{\partial \beta} \right) {}^t \left(\frac{\partial e_t(\beta)}{\partial \beta'} \right) \right\} \quad (1.19)$$

Par exemple, pour un modèle MA(1), il vient $V_{as}(\theta_1) = \frac{1}{T} (1 - \theta_1^2)$.

Dans la pratique, on remplace β par sa valeur estimée $\hat{\beta}_T$ et si on ne sait pas calculer l'espérance (1.19), cette dernière est remplacée par la moyenne empirique.

Grâce à ces résultats asymptotiques, il est possible de fabriquer des intervalles de confiance sur les paramètres des modèles, ou ce qui est équivalent, de construire des tests d'hypothèse sur ces paramètres.

1.6.8 Identification et vérification

C'est la pratique de recherche d'un modèle ARIMA compatible avec les données.

La démarche est la suivante :

1. Identification *a priori* du modèle : on recherche des valeurs plausibles du triplet (p, d, q) .
2. Estimation : pour chaque triplet, on estime le modèle.
3. Vérification : on soumet les modèles à divers tests statistiques de validation : significativité des coefficients, non corrélation des résidus ...
4. Identification *a posteriori* : si plus d'un modèle est compatible, on peut peut-être les distinguer par leur pouvoir prédictif ou leur quantité d'information.

Identification *a priori*

On commence à calculer la fonction d'autocorrélation empirique $\hat{\rho}(h)$ définie par :

$$\forall h \geq 0, \hat{\rho}(h) = \frac{\frac{1}{T-h} \sum_{t=1}^{T-h} (x_t - \bar{x}_T)(x_{t+h} - \bar{x}_T)}{\sum_{t=1}^T (x_t - \bar{x}_T)^2} \quad (1.20)$$

où \bar{x}_T est la moyenne empirique des observations : $\bar{x}_T = \frac{1}{T} \sum_{t=1}^T x_t$. Si les premiers $\hat{\rho}(h)$ sont assez proches de 1 (ou même si les premiers $\hat{\rho}(h)$ sont proches entre eux), il faut différencier la série. On recommence cette première étape tant que le même phénomène apparaît sur le corrélogramme des séries différenciées.

Pour choisir la valeur de q dans un modèle MA(q), on utilise la propriété $\rho(h) = 0$ pour $h > q$. Comme on ne dispose pas de la fonction d'autocorrélation exacte mais de sa version empirique, il faut tenir compte de la variance de son estimation. On démontre que l'estimateur (1.20) est asymptotiquement sans biais et gaussien. Sa variance asymptotique est donnée par :

$$V_{as}(\sqrt{T}\hat{\rho}) \approx 1 + \sum_{i=1}^q \rho^2(i)$$

On peut donc estimer l'écart-type de $\hat{\rho}(h)$ par

$$\hat{\sigma}(\hat{\rho}(h)) = \frac{1}{\sqrt{T}} \left(1 + \sum_{i=1}^q \hat{\rho}(i)^2 \right)^{1/2}$$

On dira alors que le $\rho(h)$ est significativement non nul au niveau α , si $\hat{\rho}(h)$ est en dehors de l'intervalle de confiance $I_\alpha = [-u_{\alpha/2} \hat{\sigma}(\hat{\rho}(h)), -u_{\alpha/2} \hat{\sigma}(\hat{\rho}(h))]$ où $-u_\alpha$ est le premier α -quantile de la loi normale $\mathcal{N}(0, 1)$ ($u_{2,5\%} = 1,96$).

Pour choisir la valeur de p dans un modèle AR(p), on agit de la même manière en remplaçant la fonction d'autocorrélation empirique par la fonction d'autocorrélation partielle empirique. Celle-ci est calculée en remplaçant dans la formule (1.15) les coefficients d'autocorrélation par leurs versions empiriques données par (1.20). On montre encore que ces coefficients d'autocorrélation partielle empirique sont asymptotiquement sans biais et normalement distribués. Leur variance asymptotique vaut :

$$V_{as}(\sqrt{T} \hat{\tau}(h)) \approx \frac{1}{T}$$

Pour choisir le couple (p, q) d'un modèle ARMA(p, q), on peut utiliser la méthode du coin présentée dans la proposition 9 en remplaçant encore les coefficients d'autocorrélation par leur version empirique. L'idéal serait de connaître la variance asymptotique de $\hat{\Delta}(i, j)$ pour déterminer s'il est ou non significativement non nul. Cela est possible mais assez délicat.

Vérification

Une fois estimés les paramètres du modèle, on peut utiliser divers tests statistiques pour valider ou non cette modélisation.

En utilisant les propriétés asymptotiques des estimateurs décrites dans le paragraphe 1.6.7, il est facile de construire des tests d'hypothèses simples

1. Test d'un modèle ARMA(p-1, q) contre ARMA(p, q) : on calcule la statistique de test $t = \frac{|\hat{\phi}_p|}{\sqrt{V_{as}(\hat{\phi}_p)}}$ et on rejette l'hypothèse nulle si cette statistique est plus grande que la valeur critique $u_{\alpha/2}$ lue dans la table de la loi normale ($u_{2,5\%} = 1,96$).
2. Test d'un modèle ARMA(p, q-1) contre un ARMA(p, q) : on calcule la statistique de test $t = \frac{|\hat{\theta}_q|}{\sqrt{V_{as}(\hat{\theta}_q)}}$ et on rejette l'hypothèse nulle si cette statistique est plus grande que la valeur critique $u_{\alpha/2}$.

Un autre type de vérification concerne les résidus du modèle : si $(\varepsilon_t, t \in \mathbb{Z})$ est un bruit blanc, alors les coefficients d'autocorrélation sont nul pour $h \geq 1$. On peut alors examiner la fonction d'autocorrélation empirique des résidus et s'assurer que les coefficients ne sont pas significativement non nuls.

On peut aussi effectuer le test du "portmanteau". L'hypothèse nulle de ce test est $H_0 : \rho(1) = \dots \rho(K) = 0$, l'hypothèse alternative est $H_1 : \text{non}H_0$. La statistique du test (statistique de Box et Pierce) est

$$Q_{BP} = T \sum_{h=1}^K \hat{\rho}(h)^2(\hat{\varepsilon})$$

Sous l'hypothèse nulle, cette statistique converge en loi (quand $T \rightarrow +\infty$) vers χ_{K-p-q}^2 . L'hypothèse H_0 est rejetée au seuil de α si la valeur calculée de Q_{BP} est supérieure au quantile $1 - \alpha$ de la loi du χ^2 à $K - p - q$ degrés de liberté.

Pour des raisons de convergence des estimateurs, cette statistique peut être remplacée par celle de Ljung et Box définie par

$$Q_{LB} = T(T+2) \sum_{h=1}^K \frac{\hat{\rho}(h)^2(\hat{\varepsilon})}{T-h}$$

qui converge aussi en loi vers une loi du Khi-2 à $K - p - q$ degrés de liberté.

Identification *a posteriori*

Pour choisir entre plusieurs modèles qui ont passé la phase de vérification, on peut considérer les deux critères suivants :

1. Critère du pouvoir prédictif : on choisit le modèle qui minimise la variance du bruit. On peut alors utiliser les différents indicateurs qui suivent et on choisira le modèle qui a le plus petit (ou le plus grand suivant le cas) :
 - La variance du bruit elle-même (le plus petit),
 - Le coefficient de détermination, comme pour les modèles linéaires : $R^2 = 1 - \frac{\hat{\sigma}^2(\hat{\varepsilon}_t)}{V(X_t)}$ avec $V(X_t)$ variance empirique de la série initiale (variance totale). On prendra, parmi les modèles, celui qui à le R^2 le plus grand.
 - Le coefficient de détermination modifié (pour tenir compte du nombre de paramètres) $\bar{R}^2 = 1 - \frac{T-p-q}{T-1} \frac{\hat{\sigma}^2(\hat{\varepsilon}_t)}{V(X_t)}$. Là encore, on choisira, parmi les modèles, celui qui à le \bar{R}^2 le plus grand.
 - On peut même calculer la statistique de Fisher comme pour les modèles linéaires : $F = \frac{T-p-q}{p+q} \frac{V(X_t) - \hat{\sigma}^2(\hat{\varepsilon}_t)}{\hat{\sigma}^2(\hat{\varepsilon}_t)}$ On choisira encore, parmi les modèles possibles, celui qui a le F le plus grand.
2. Critère de la quantité d'information : la modélisation ARMA(p,q) est une approximation de la réalité et la vraie loi inconnue des observations ne satisfait pas nécessairement un tel modèle. On va donc mesurer l'écart entre la vraie loi inconnue et celle du modèle proposé. Si f_0 est la densité inconnue des observations et $\{f, f \in \mathcal{F}_{p,q}\}$ la famille des densités correspondant au modèle ARMA(p, q), on calcule *l'information de Kullback* :

$$I(f_0, \mathcal{F}_{p,q}) = \min_{f \in \mathcal{F}_{p,q}} \int \ln \left(\frac{f_0(x)}{f(x)} \right) f(x) dx$$

Cette quantité est toujours positive et s'annule si $f_0 \in \mathcal{F}_{p,q}$. Le problème est qu'on ne connaît pas f_0 mais si on est capable d'en connaître un "bon" estimateur alors l'information de Kullback pourra être utilisée comme critère de choix. Les deux approximations les plus classiques sont :

- Le critère d'Akaike : $AIC(p, q) = \ln \hat{\sigma}^2(\hat{\varepsilon}_t) + 2 \frac{p+q}{T}$
- Le critère d'information bayésien : $BIC(p, q) = \ln \hat{\sigma}^2(\hat{\varepsilon}_t) + (p+q) \frac{\ln T}{T}$

Parmi tous les modèles ARMA possibles, on choisira alors le modèle dont le critère AIC ou le critère BIC est le plus petit.

Tous les calculs, les statistiques et les indicateurs présentés dans ce paragraphe sont effectués dans la plupart des logiciels statistiques.

1.6.9 Exemple simulé

On considère un échantillon de $T = 500$ observations d'une simulation d'un modèle défini *a priori*, non précisé. Peut-on identifier ce modèle?

Le graphique (FIG. 1.12) ne présente pas de saisonnalité évidente. On peut envisager de traiter la série directement.

Les graphes (FIG. 1.13) et (FIG. 1.14) montrent les fonctions d'autocorrélations et d'autocorrélation partielles empiriques obtenues par le programme S-PLUS suivant :

```
> acY <- acf(Y, type = "correlation", plot = T)
> pacY <- acf(Y, type = "partial", plot = T)
```

Les coefficients $\hat{\rho}(4)$, $\hat{\rho}(7)$, $\hat{\rho}(27)$ sont significativement non nuls à 5%. Il faudrait estimer au moins un MA(7).

De même, les coefficients $\hat{\tau}(4)$, $\hat{\tau}(7)$, sont significativement non nuls à 5%. Il faudrait donc envisager un modèle AR(7).

La méthode des coins, programmée dans S-PLUS donne le résultat suivant :

```
> MatDelta <- FaireMatDelta(Y, 8, 8)
> MatDelta
```

	[,1]	[,2]	[,3]	[,4]
[1,]	-0.05293422192	0.0028020318506	-0.00231237515660	1.583423126e-002
[2,]	-0.04348003119	-0.0004217282966	-0.00007043840290	2.367417688e-007
[3,]	-0.04368140921	-0.0022780144152	-0.00004092360077	6.931340434e-004
[4,]	-0.09627591819	0.0081984035940	-0.00075803141860	5.478188724e-005
[5,]	-0.02451040037	0.0006007597263	-0.00001502753232	4.699943314e-006
[6,]	0.03672213852	-0.0003464976311	-0.00027419073978	-1.923707531e-005
[7,]	-0.06915485114	0.0039796827918	-0.00022811221613	2.807497136e-006
[8,]	0.02185903862	-0.0001812675405	-0.00005893318464	2.449149691e-006

	[,5]	[,6]	[,7]	[,8]
[1,]	-6.488107672e-003	-2.605097133e-003	-2.342698122e-003	-4.442018877e-005
[2,]	1.339548733e-005	-6.942324981e-006	7.450528803e-009	-6.896853100e-009
[3,]	-5.476632717e-004	2.561279698e-004	-6.316639285e-005	8.487117635e-006
[4,]	5.684236517e-005	-8.531094699e-006	2.413706981e-008	6.205211575e-007
[5,]	-1.115557423e-007	3.545675318e-008	-7.731289886e-008	1.012466364e-008
[6,]	-1.099635280e-006	3.465083998e-008	5.010493326e-008	-7.618944005e-009
[7,]	-2.197026902e-007	4.006754969e-009	-2.056674284e-011	-3.303664085e-011
[8,]	-4.965063815e-008	-4.782383647e-009	2.608961082e-010	2.002498584e-011

Il semble qu'il y ait une rupture entre après la première ligne et la première colonne (tous les coefficients à partir du (2,2) sont au moins 6 ou 7 fois plus petits que $\hat{\Delta}(1,1)$). On est donc conduit à estimer un modèle ARMA(1, 1).

```
> p <- 1
```

```

> q <- 1
> ModeleARMA <- list(order = c(p, 0, q))
> ResARMA <- arima.mle(Y, model = ModeleARMA, max.iter = 100, max.fcal = 1000)
> for(i in 1:p) {
  Aux <- c(ResARMA$model$ar[i], ResARMA$model$ar[i]/sqrt(ResARMA$var.coef[i, i]))
  print(Aux)
}
[1] 0.8193256611 13.4975089853
> for(i in 1:q) {
  Aux <- c(ResARMA$model$ma[i], ResARMA$model$ma[i]/sqrt(ResARMA$var.coef[i + p, i + p]))
  print(Aux)
}
[1] 0.8982321552 19.3023050293
> ResARMA$converged
[1] T
> ResARMA$loglik
[1] 1477.640685
> ResARMA$aic
[1] 1481.640685
> ResARMA$sigma2
[1] 0.2565573519

```

Les deux paramètres estimés sont significativement non nuls (dans les deux cas, la statistique de test est très grande). Les deux polynômes ont des racines en dehors du cercle unité. Sur le plan

La phase de validation est effectuée à l'aide du programme suivant :

```

> Aux <- c("R2", 1 - ResARMA$sigma2/Sigma2Y)
> print(Aux)
[1] "R2" "0.0188804476776946"
> Aux <- c("R2barre", 1 - (ResARMA$sigma2/Sigma2Y * (length(Y) - 1))/(length(Y) -
  p - q))
> print(Aux)
[1] "R2barre" "0.0178973619539248"
> FTest <- ((Sigma2Y - ResARMA$sigma2)/(p + q))/(ResARMA$sigma2/(length(Y) - p -
  q))
> Aux <- c("F", FTest, 1 - pf(FTest, p + q, length(Y) - p - q))
> print(Aux)
[1] "F" "9.6026456397381" "7.4001991227646e-005"
> ResResid <- arima.diag(ResARMA, resid = T)
> ResResid$gof
$statistic:
[1] 0.2339935596 3.8552610813 3.8564672050 7.0140032858 9.4279028381
[6] 10.3081689831 10.3113139266 10.3578456037 12.5061145300 12.5199901922
[11] 12.5234768784 13.2029368742 13.2108224716

$df:
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13

```

\$p.value:

```
[1] 0.62857834555 0.14549252902 0.27738601468 0.13515006374 0.09316636713
[6] 0.11225995969 0.17160725186 0.24080249730 0.18625649225 0.25176078200
[11] 0.32560707157 0.35446460738 0.43166226630
```

\$lag:

```
[1] 3 4 5 6 7 8 9 10 11 12 13 14 15
```

Le test du “portmanteau” ne montre pas une autocorrélation des résidus. Nous avons donc choisi un modèle ARMA correct pour représenter la série initiale.

Par contre, la variance du bruit est encore très grande: le R^2 du modèle n’est que de l’ordre de 2%, même si, par la statistique de Fisher, on peut le considérer comme non nul. Il doit être difficile de faire des prévisions correctes.

Le modèle obtenu est

$$(Id - 0,9B)Y_t = (1 - 0,8B)\varepsilon_t$$

Il s’écrit sous forme de moyenne mobile infini de la manière suivante :

$$Y_t = \varepsilon_t + 0,1 \sum_{i=1}^{+\infty} (0,9)^i \varepsilon_{t-i}$$

La variance de Y_t vaut alors :

$$\begin{aligned} \text{var}(Y_t) &= \text{var}(\varepsilon_t) \left[1 + 0,1^2 \sum_{i=0}^{+\infty} 0,9^{2i} \right] \\ &= \frac{20}{19} \text{var}(\varepsilon_t) \end{aligned}$$

Il est donc normal de trouver un R^2 petit.

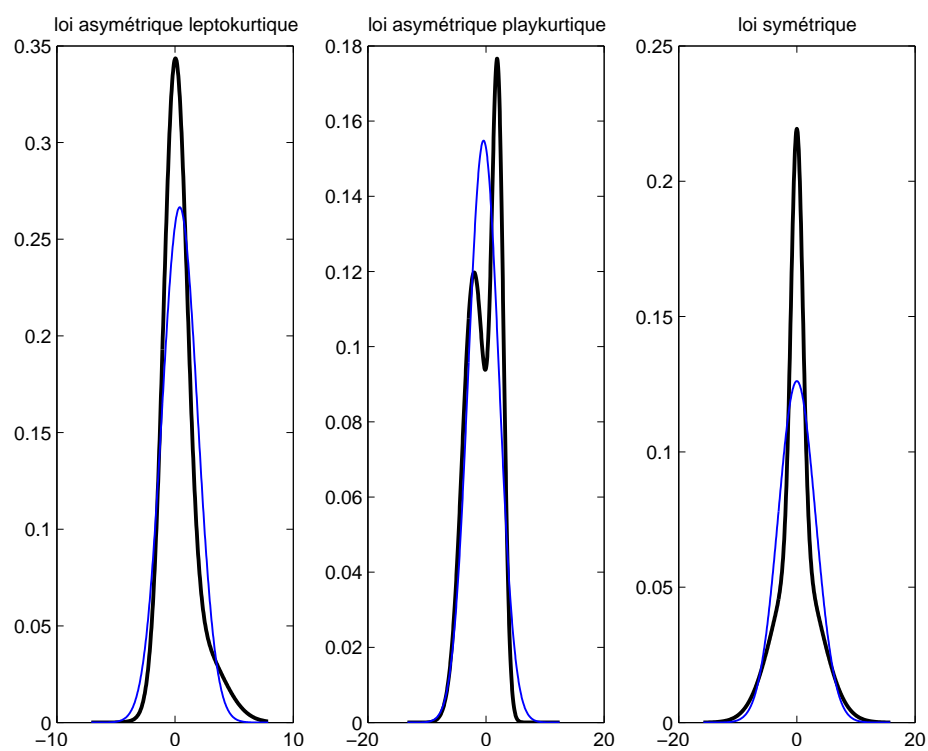


FIG. 1.6 – 3 densités de loi contaminée : symétrique, asymétrique leptokurtique et play-kurtique

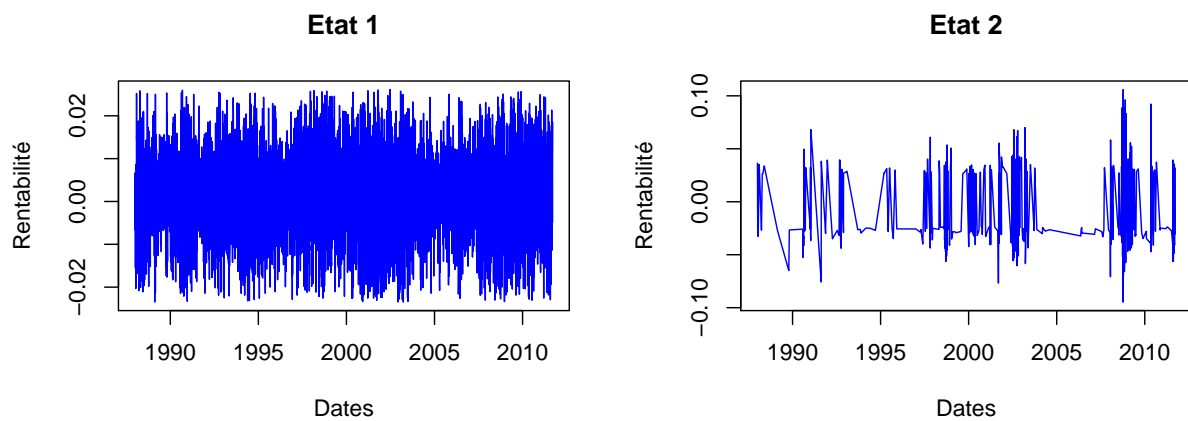


FIG. 1.7 – *Les deux régimes pour les rentabilités journalières de l'indice CAC 40*

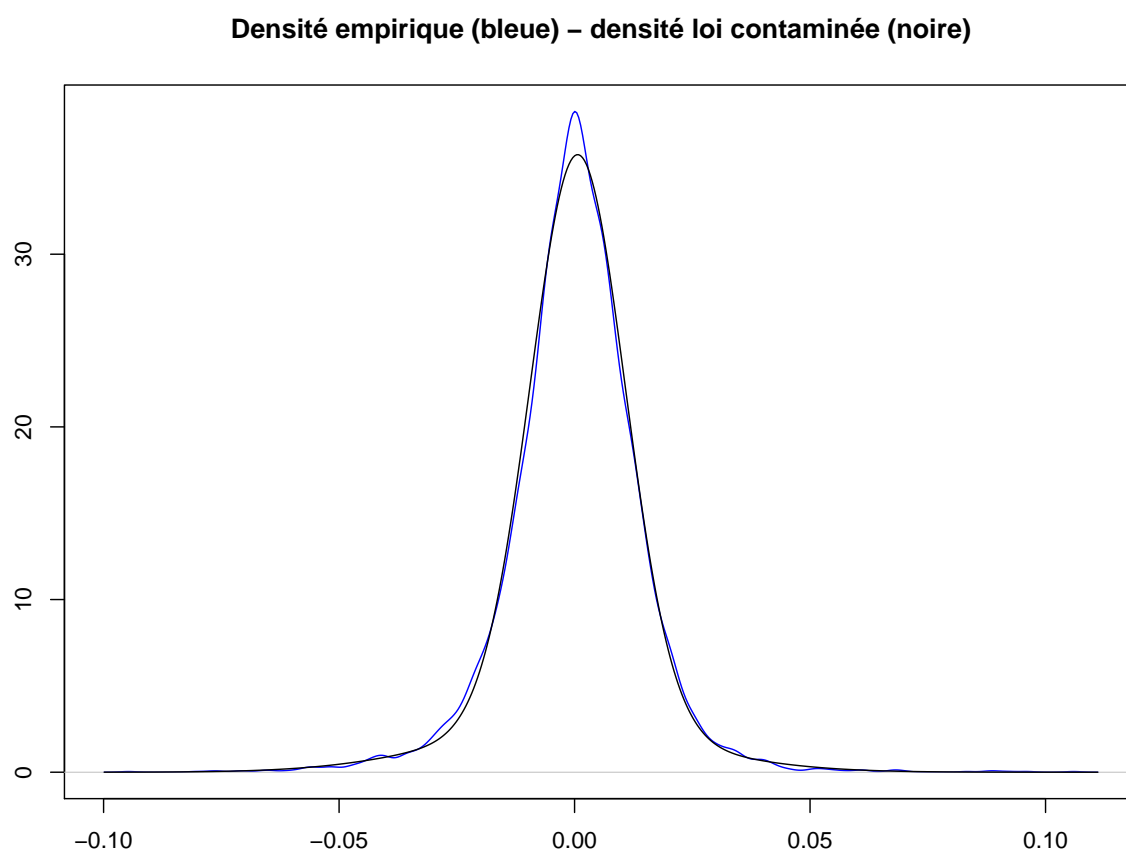


FIG. 1.8 – *Densité noyau gaussien : trait gras - modèle mélange : trait fin*

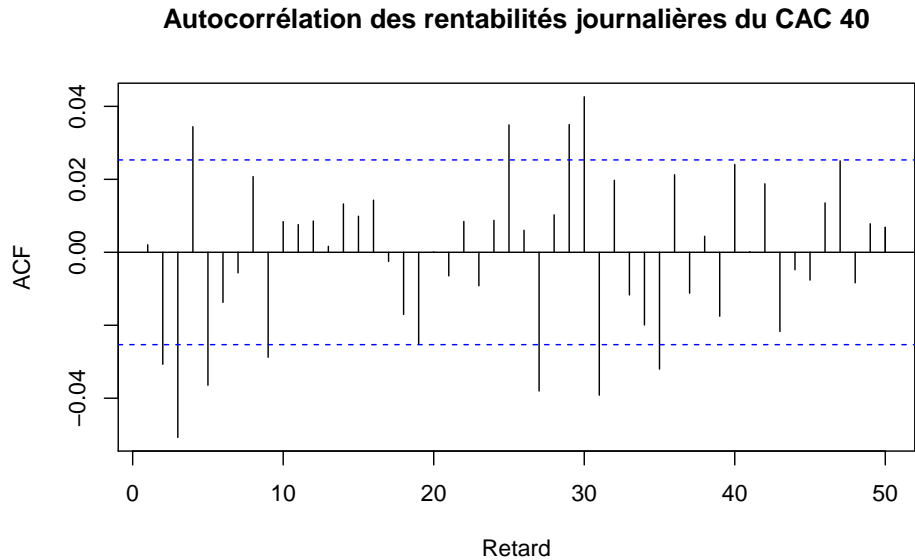


FIG. 1.9 – *Fonction d'autocorrélation des rentabilités journalières du Cac40*

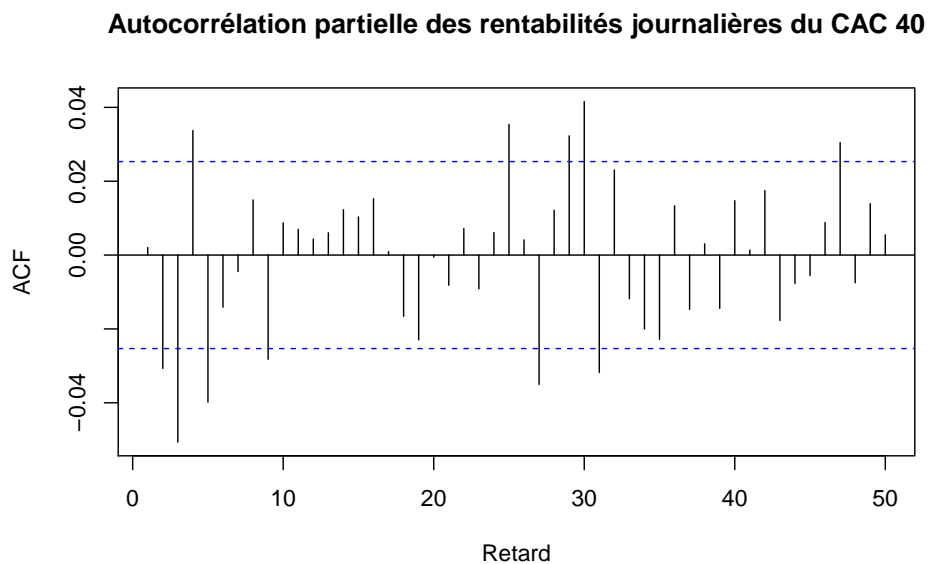


FIG. 1.10 – *Fonction d'autocorrélation partielle des rentabilités journalières du Cac40*

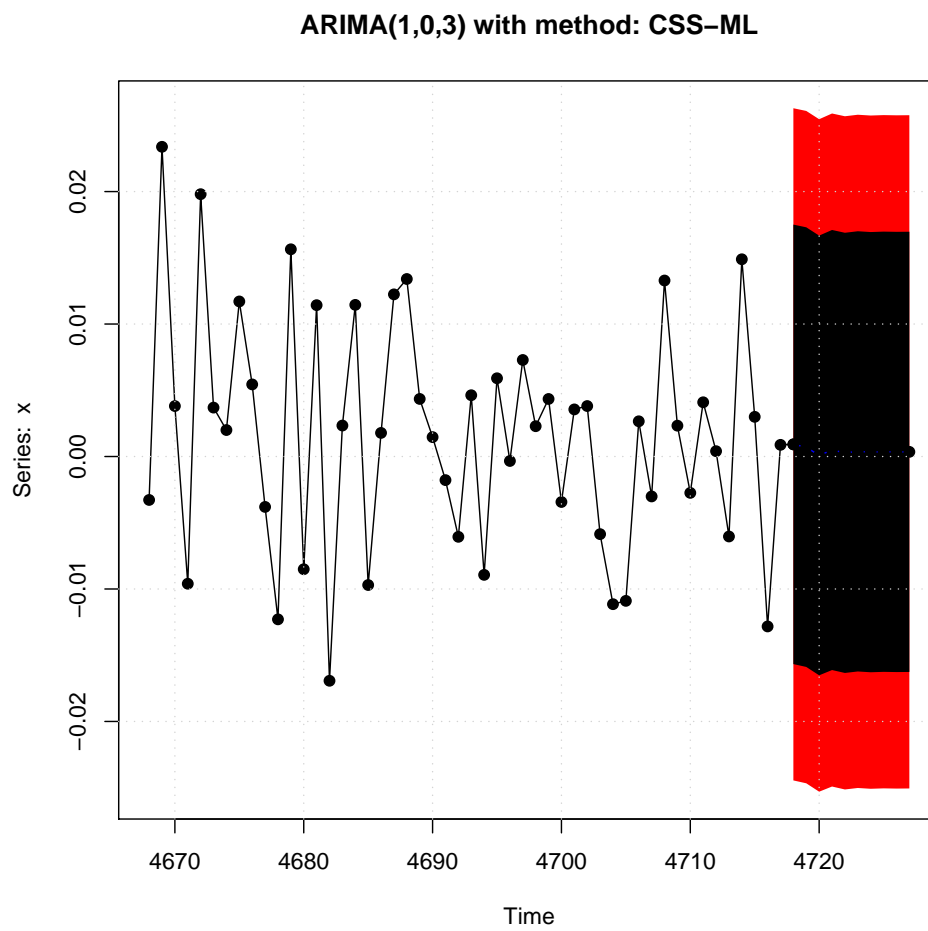


FIG. 1.11 – *Prévision des rentabilités journalières du CAC modèle ARMA(1,3)*

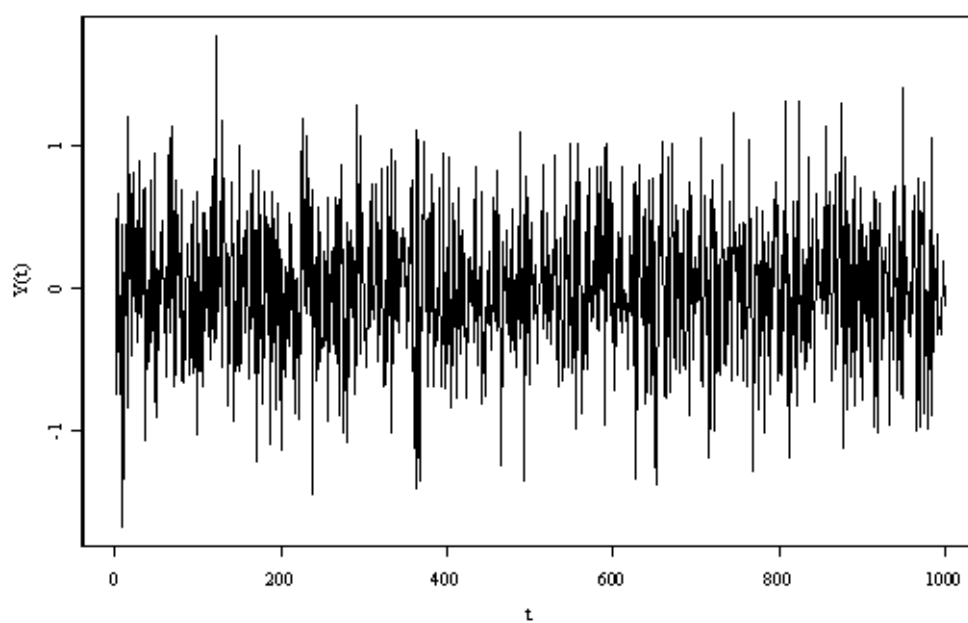


FIG. 1.12 – *Trajectoire d'un processus ARMA simulé*

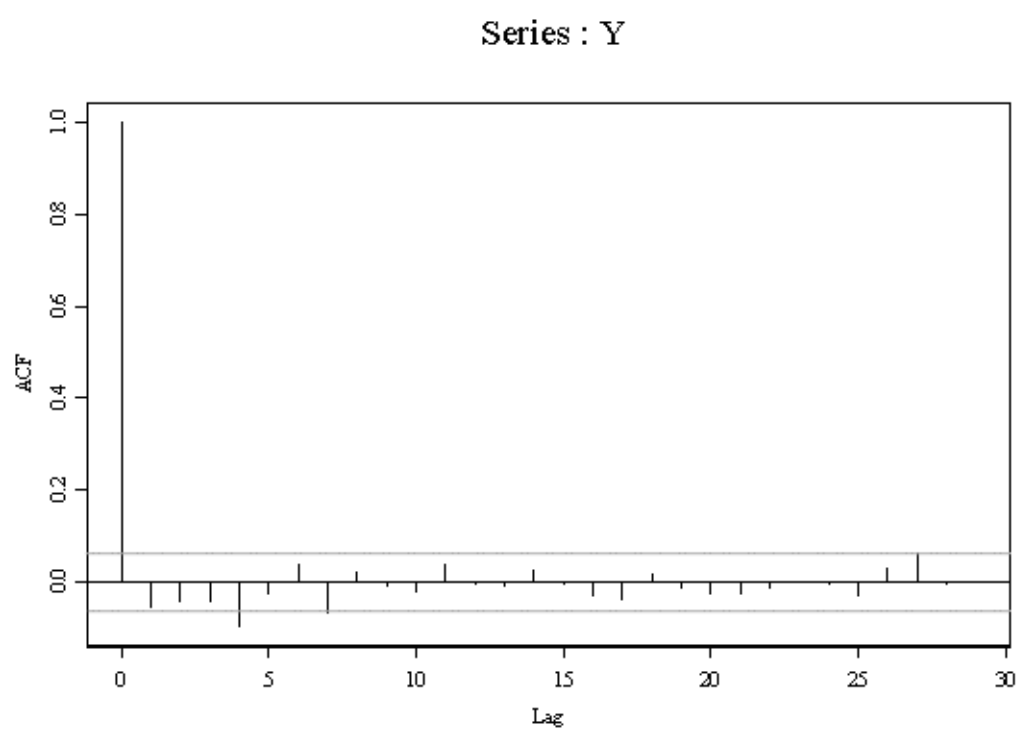


FIG. 1.13 – Autocorrélations d'un processus ARMA simulé

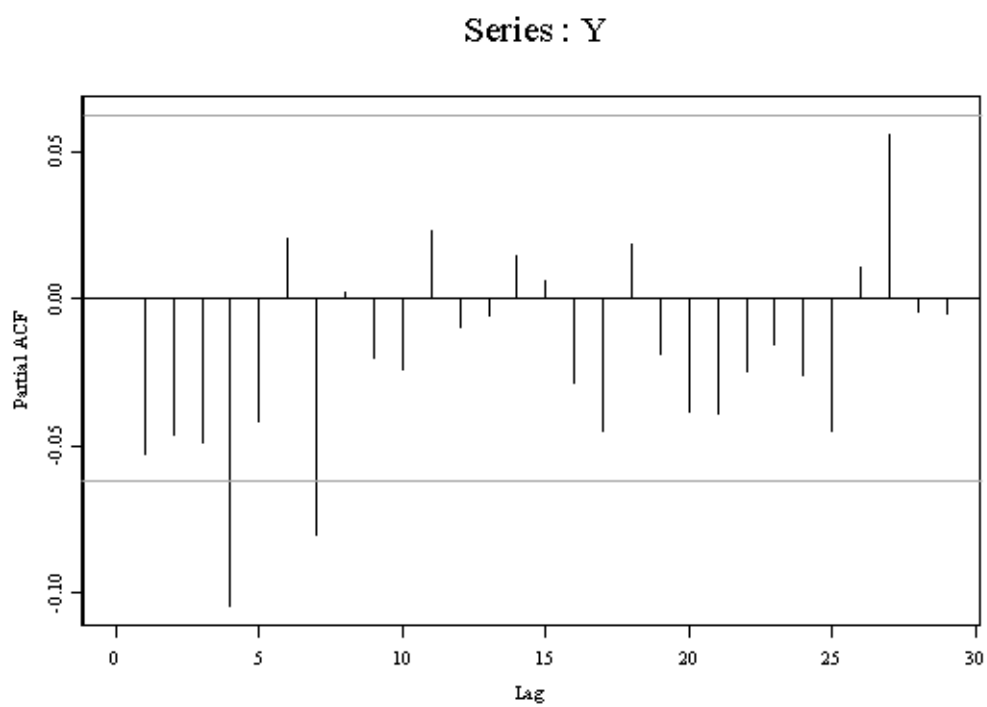


FIG. 1.14 – Autocorrélations d'un processus ARMA simulé

Chapitre 2

Modèles ARCH univariés

2.1 Le modèle ARCH(1)

Les modèles ARCH (AutoRegressive Conditional Heteroskedasticity ou modèle à variance conditionnelle autorégressive) ont été introduits par Engle (1982) pour prendre en compte des variances conditionnelles instables dans le temps.

2.1.1 Définition

Soit $Y = (Y_t, t = 0, 1, \dots)$ un processus aléatoire en temps discret. On note $\underline{Y}_{t-1} = \{Y_{t-1}, Y_{t-2}, \dots\}$. On impose au processus (Y) les conditions suivantes :

$$\begin{aligned} \forall t \in \{1, \dots\}, E(Y_t | \underline{Y}_{t-1}) &= 0 \\ \forall t \in \{1, \dots\}, Y_t^2 &= c + a Y_{t-1}^2 + u_t \quad (2.1) \end{aligned}$$

où $u = (u_t, t = 1, \dots)$ est un bruit blanc, $a, c \in \mathbb{R}$. Pour assurer l'existence (condition de positivité) de ce processus, on impose les conditions (suffisantes) suivantes :

$$\begin{aligned} a &\geq 0 \\ \forall t, c + u_t &> 0 \end{aligned}$$

Ce qui impose que le support de la loi de u_t soit inclus dans $] -c, +\infty[$.

Le graphique (2.1) montre une trajectoire de 2000 observations d'un processus ARCH(1), avec $c = 0,1$ et $a = 0,9$.

Remarques

1. Le processus est markovien d'ordre 1 : $P(Y_t < y_t | \underline{Y}_{t-1} = \underline{y}_{t-1}) = P(Y_t < y_t | Y_{t-1} = y_{t-1})$
2. La variance conditionnelle du processus (Y) dépend du temps :

$$V(Y_t | \underline{Y}_{t-1}) = c + a Y_{t-1}^2$$

3. L'espérance (non conditionnelle) du processus (Y) est nulle

$$E(Y_t) = E(E(Y_t | \underline{Y}_{t-1})) = E(0) = 0$$

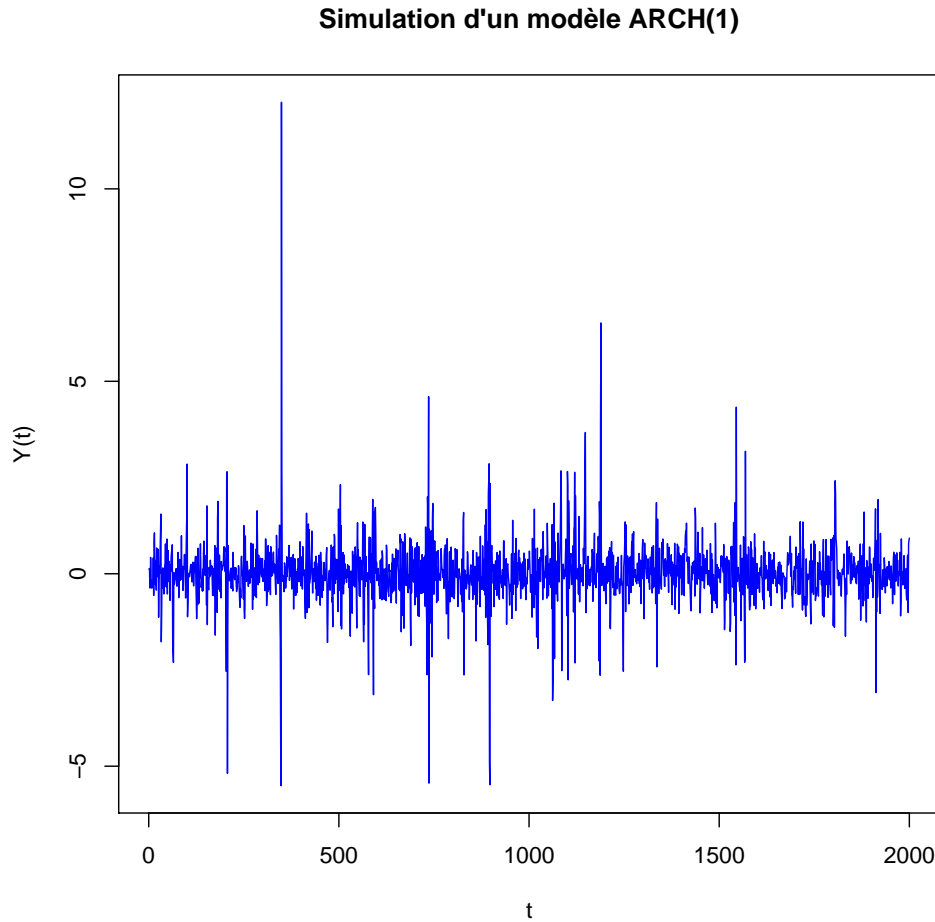


FIG. 2.1 – Simulation d'un modèle ARCH(1)

4. La variance (non conditionnelle) du processus (Y) est solution de l'équation :

$$\forall t > 0, h_t = V(Y_t) = E(Y_t^2) = c + a h_{t-1}$$

Pour que cette variance soit invariante dans le temps, il faut que $h_t = h_0 = \frac{c}{1-a}$ et donc que $0 \leq a < 1$.

2.1.2 Quelques propriétés du processus Y

Propriété 1 : On peut remarquer que le processus (Y) est orthogonal à tout son passé *i.e.* :

$$\forall t > 1, \forall h > 0, E(Y_t | \underline{Y}_{t-h}) = 0$$

Démonstration

Ceci est une conséquence du théorème des “projections itérées”, puisque $\underline{Y}_{t-h} \subset \underline{Y}_{t-1}$, il vient :

$$E(Y_t | \underline{Y}_{t-h}) = E(E(Y_t | \underline{Y}_{t-h}, \underline{Y}_{t-1})) = E(E(Y_t | \underline{Y}_{t-1}) | \underline{Y}_{t-h}) = E(0 | \underline{Y}_{t-h}) = 0$$

Propriété 2 : La covariance conditionnelle entre 2 instants distincts du processus Y est nulle :

$$\forall t \geq 0, \forall k > 0, \forall h > 0, \text{Cov} \left(Y_t, Y_{t+k} \mid \underline{Y_{t-h}} \right) = 0$$

Démonstration

Ceci est le résultat de la propriété précédente :

$$\begin{aligned} \text{Cov} \left(Y_t, Y_{t+k} \mid \underline{Y_{t-h}} \right) &= E \left[Y_t Y_{t+k} \mid \underline{Y_{t-h}} \right] - E \left(Y_t \mid \underline{Y_{t-h}} \right) E \left(Y_{t+k} \mid \underline{Y_{t-h}} \right) \\ &= E \left[Y_t Y_{t+k} \mid \underline{Y_{t-h}} \right] \\ &= E \left[E \left[Y_t Y_{t+k} \mid \underline{Y_{t-k+1}} \right] \mid \underline{Y_{t-h}} \right] \\ &\quad (\text{Th. des projections it.}) \\ &= E \left[Y_t E \left[Y_{t+k} \mid \underline{Y_{t-k+1}} \right] \mid \underline{Y_{t-h}} \right] \\ &\quad (\text{car } Y_t \in \underline{Y_{t-k+1}}) \\ &= 0 \end{aligned}$$

Par conséquent, la covariance (non conditionnelle) entre 2 instants distincts est aussi nulle (puisque $\text{Cov} (Y_t, Y_{t+k}) = E (Y_t Y_{t+k}) = E \left[E (Y_t Y_{t+k} \mid \underline{Y_{t-1}}) \right] = E \left[\text{Cov} (Y_t, Y_{t+k} \mid \underline{Y_{t-1}}) \right] = 0$).

On vient donc de définir un *bruit blanc "faible"* : les variables Y_t sont centrées, 2 à 2 non corrélées et de variance (inconditionnelle) constante.

Propriété 3 : La variance de Y_t sachant $\underline{Y_{t-h}}$ est donnée par :

$$V \left(Y_t \mid \underline{Y_{t-h}} \right) = c \frac{1 - a^h}{1 - a} + a^h Y_{t-h}^2$$

Démonstration

Il suffit d'itérer l'équation (2.1) :

$$Y_t^2 = c \sum_{k=0}^{h-1} a^k + a^h Y_{t-h}^2 + \sum_{k=1}^{h-1} a^k u_{t-k}$$

Et en prenant l'espérance conditionnelle de chacun des membres de cette équation conditionnellement à $\underline{Y_{t-h}}$, on obtient le résultat.

Propriété 4 : L'écart entre la variance conditionnelle et la variance inconditionnelle est donnée par :

$$V \left(Y_t \mid \underline{Y_{t-h}} \right) - V (Y_t) = a^h [Y_{t-h}^2 - E (Y_{t-h}^2)]$$

Comme la limite (quand $h \rightarrow \infty$) de la variance conditionnelle, est égale à la variance inconditionnelle, il y a un "retour vers la moyenne" de la variance (voir graphique (2.2)).

Pour illustrer ces diverses propriétés, la série simulée est traitée dans R. Dans un premier temps, on estime un modèle ARMA sur la série brute (Y_t). Le graphique (2.3) des autocorrélations

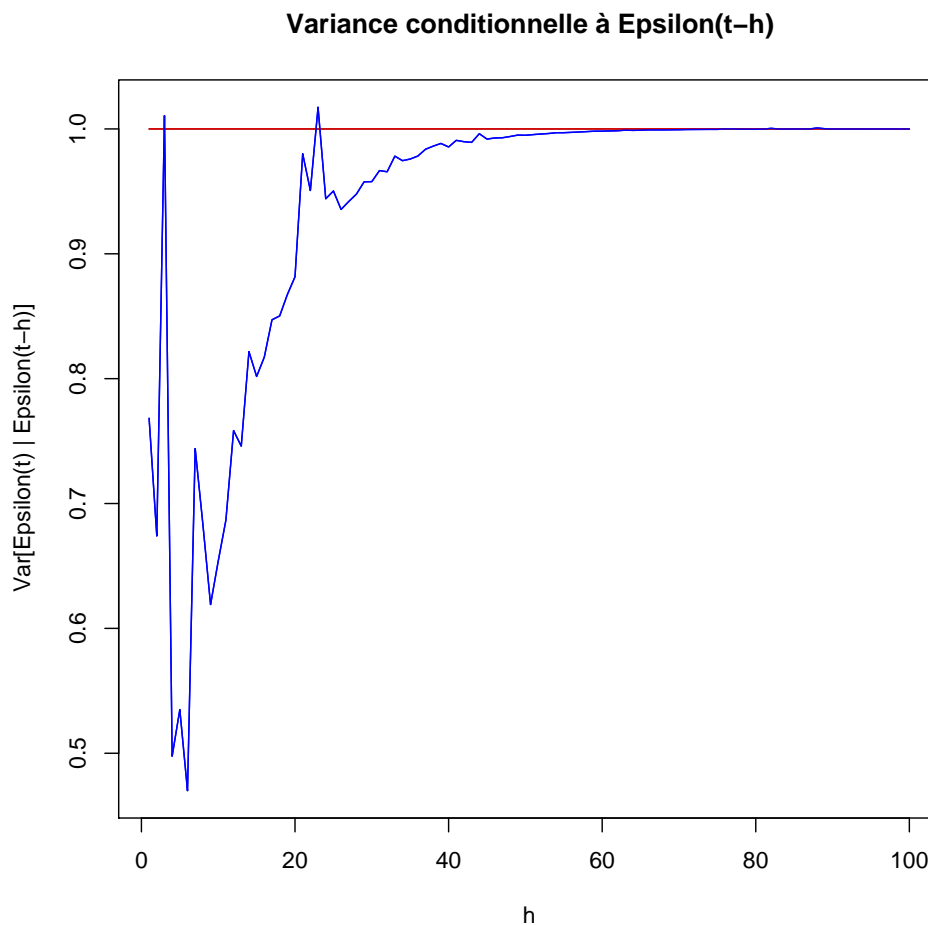


FIG. 2.2 – $V\left(Y_t \mid \underline{Y_{t-h}}\right)$ en fonction de h

partielles de la série montre que les coefficients 2, 4 et 7 sont significativement (au niveau 5%) non nuls. On va donc estimer un modèle AR(7). Les coefficients 2, 4 et 7 du modèle AR sont significativement non nuls (à 5%). La variance estimée du résidu de ce modèle est égale à 0,6524 alors que la variance totale (la variance empirique de Y_t) est égale à 0,6637204. le modèle AR(7) n'explique donc que $R^2 = 1 - \frac{0,6524}{0,6637204} = 1,70\%$ de la variance totale !

```
> summary(Yt) #Yt contient les données
      Min.   1st Qu.     Median       Mean   3rd Qu.      Max.
-5.506000 -0.318100  0.002206  0.013930  0.355800 12.250000
> mean(Yt)
[1] 0.0139306
> var(Yt)
[1] 0.6637204
> sd(Yt)
```



```
[1] 0.8146904
> acf.par<-pacf(Yt, plot = T)
```

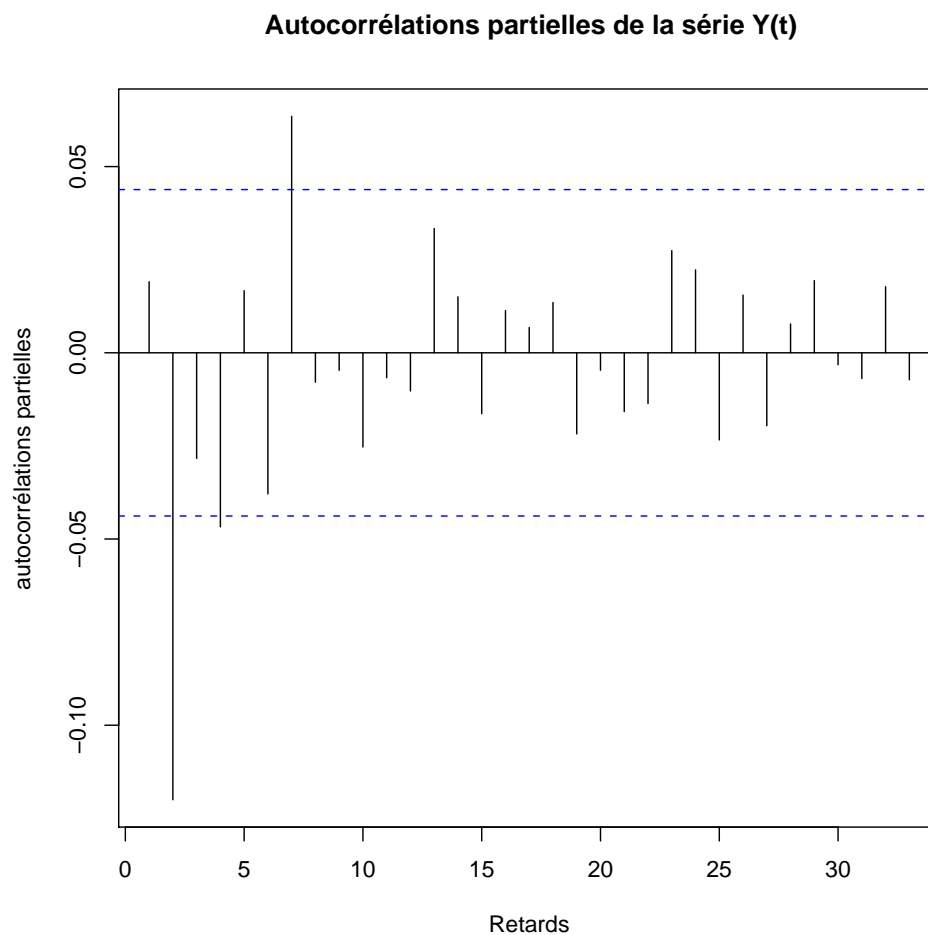


FIG. 2.3 – Autocorrélations partielles de Y_t

```
> library(tseries)
> arma.fit <- arma(Yt, lag=list(ar=c(2,4,7)))
Warning message:
order is ignored in: arma(Yt, lag = list(ar = c(2, 4, 7)))
> summary(arma.fit)
```

```
Call:
arma(x = Yt, lag = list(ar = c(2, 4, 7)))
```

```
Model:
```

ARMA(7,0)

Residuals:

Min	1Q	Median	3Q	Max
-5.67388	-0.34991	-0.01127	0.35814	11.58743

Coefficient(s):

	Estimate	Std. Error	t value	Pr(> t)	
ar2	-0.12647	0.02231	-5.669	1.44e-08	***
ar4	-0.04620	0.02233	-2.069	0.0385	*
ar7	0.06013	0.02218	2.711	0.0067	**
intercept	0.01566	0.01806	0.867	0.3861	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Fit:

sigma^2 estimated as 0.6524, Conditional Sum-of-Squares = 1299.52, AIC = 4829.47

Dans un deuxième temps, nous traitons la série Y_t^2 . Le graphique (2.4) donne les autocorrélations partielles de cette série. On peut déjà remarquer que leurs amplitudes sont beaucoup plus grandes que celles de la série initiale. Les deux premiers coefficients d'autocorrélation partielle de Y_t^2 étant significativement non nuls (au niveau 5%), il est naturel d'estimer un modèle AR(2). Les résultats montrent que les coefficients estimés sont significativement non nuls (à 5%). Cependant, nous verrons par la suite que cette façon d'estimer les paramètres d'un modèle ARCH n'est pas bonne. Le pourcentage de variance expliquée par ce modèle est égale à $R^2 = 10,2\%$, à comparer au 1,70% du 1^{er} modèle.

```
> YtCarre<-Yt^2
> summary(YtCarre)
      Min.    1st Qu.    Median      Mean   3rd Qu.      Max.
4.036e-07 2.248e-02 1.125e-01 6.636e-01 3.961e-01 1.501e+02
> mean(YtCarre)
[1] 0.6635826
> var(YtCarre)
[1] 15.86441
> sd(YtCarre)
[1] 3.983015
> acf.YtCarre<-pacf(YtCarre, plot = T,
  main='Autocorrélations partielles de la série Y(t)^2',
  xlab='Retards', ylab='Autocorrélations partielles')

> arma.fitYt2 <- arma(YtCarre, order=c(2,0))
```

```

> summary(arma.fitYt2)

Call:
arma(x = YtCarre, order = c(2, 0))

Model:
ARMA(2,0)

Residuals:
      Min       1Q   Median       3Q      Max
-38.9602  -0.4436  -0.3765  -0.1553  139.4273

Coefficient(s):
              Estimate Std. Error t value Pr(>|t|)
ar1           0.29741    0.02232   13.324 < 2e-16 ***
ar2           0.05922    0.02232    2.653  0.00797 **
intercept     0.42779    0.08613    4.967  6.8e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Fit:
sigma^2 estimated as 14.24, Conditional Sum-of-Squares = 28442.61, AIC = 10994.24

Les vrai modèle étant un ARCH(1), ce modèle a été estimé dans R. Les résultats sont les suivants :
> arma.fitYt2Vrai <- arma(YtCarre, order=c(1,0))
> summary(arma.fitYt2Vrai)

Call:
arma(x = YtCarre, order = c(1, 0))

Model:
ARMA(1,0)

Residuals:
      Min       1Q   Median       3Q      Max
-40.0107  -0.4587  -0.3907  -0.1675  140.0212

Coefficient(s):
              Estimate Std. Error t value Pr(>|t|)
ar1           0.31620    0.02121   14.905 < 2e-16 ***
intercept     0.45407    0.08566    5.301 1.15e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Fit:
sigma^2 estimated as 14.29, Conditional Sum-of-Squares = 28542.94, AIC = 10998.28

```

On peut remarquer que la part de la variance expliquée passe de 10,2% pour le modèle ARCH(2) à 9.90% pour le modèle ARCH(1). Ce qui signifie que ces deux modèles sont très proches. Le modèle ARCH(1) estimé (par cette méthode) est donc $\varepsilon_t^2 = 0.45407 + 0.31620 \varepsilon_{t-1}^2 + u_t$

2.1.3 Distribution du processus Y

Pour pouvoir étudier ce processus, il faut préciser la distribution conditionnelle de $Y_t | Y_{t-1}$. Engle (1982) a initialement introduit un processus (Y) avec une distribution conditionnellement normale. On peut donc écrire :

$$Y_t | Y_{t-1} \rightsquigarrow \mathcal{N}(0, c + a Y_{t-1}^2)$$

Il faut remarquer que ce processus, qui est conditionnellement gaussien, n'est pas marginalement gaussien (ou encore qu'un vecteur de n Y_t successifs n'est pas normalement distribué).

Trouver la loi de Y_t n'étant pas possible (et de toute façon, inutile), il peut être intéressant de chercher quelques caractéristiques de cette distributions. Il est facile de montrer que les moments d'ordre impair de Y_t sont nuls et qu'il faut des conditions sur le paramètre a pour que le moment d'ordre $2k$ existe. Par exemple :

$$\begin{aligned} E(Y_t^2) < \infty &\iff a < 1 \\ E(Y_t^4) < \infty &\iff 3a^2 < 1 \end{aligned}$$

On peut calculer le coefficient d'aplatissement de la série : si $0 \leq a < \frac{1}{\sqrt{3}}$ alors $\gamma_4(Y_t) = 3 \frac{1-a^2}{1-3a^2} > 3$ et conclure que la distribution marginale est toujours leptokurtique (le coefficient d'aplatissement de Y_t est soit strictement supérieur à 3 soit infini).

Le modèle ARCH(1) gaussien présente donc beaucoup d'avantage pour la représentation des séries de rentabilités boursières puisque

1. Sa distribution est symétrique, leptokurtique.
2. La corrélation entre deux instants distincts est nulle.
3. Par contre, il y a autocorrélation entre les carrés du processus.

2.2 Propriétés générales des processus de type ARCH

2.2.1 Quelques autres modèles

La classe des processus ARCH a connu un succès extraordinaire dans les études économiques et financières. Le premier modèle (modèle ARCH) a été proposé par Engle (1982). Depuis, Bollerslev, Chou, et Kroner (1992) recense plus de 400 articles sur ce domaine. On ne présente ici que les modèles unidimensionnels.

Nous nous intéressons à un processus $(Y_t, t \in \mathbb{N})$ à valeurs réelles, dont on va modéliser l'espérance et la variance conditionnelle à son propre passé.

Définition 15 On fixe une mesure de probabilité λ sur \mathbb{R} , d'espérance 0 et de variance 1. Une variable aléatoire est de **type** λ si $Y = 0$ λ p.p. ou bien s'il existe $\mu \in \mathbb{R}$ et $\sigma \in \mathbb{R}_+$ tels que la loi de $\frac{Y-\mu}{\sigma}$ soit λ .

Définition 16 *Le processus $(Y_t, t \in \mathbb{N})$ est un processus autorégressif conditionnellement hétéroscédastique si la loi de Y_t conditionnellement à la tribu*

$$\mathcal{F}_{t-1} = \sigma(Y_s, s \leq t-1)$$

est de type λ , d'espérance μ_t et de variance σ_t^2 .

Dans la plupart des modèles, la loi λ est la loi $N(0,1)$, et on donne une représentation paramétrique de l'espérance conditionnelle μ_t et de la variance conditionnelle σ_t^2 .

Les modèles "ARCH purs" sont tels que $\mu_t = 0$. Des modèles plus conformes à la théorie définissent l'espérance conditionnelle μ_t par une fonction croissante de la volatilité conditionnelle σ_t^2 : ce sont les modèles de type ARCH *in mean* (où classiquement $\mu_t = \alpha + \beta \sigma_t^2$ ou bien $\mu_t = \alpha + \beta \sigma_t$ avec $\alpha, \beta \in \mathbb{R}$)

Pour que les modèles soient opérationnels, on définit de manière paramétrique l'espérance et la variance conditionnelle :

$$\begin{aligned}\mu_t &\equiv \mu_t(\theta) \\ \sigma_t^2 &\equiv \sigma_t^2(\theta)\end{aligned}$$

où $\theta \in \mathbb{R}^p$ est un paramètre de dimension $p > 0$.

Par exemple, si on considère le modèle ARCH *in mean*(1), $\theta =^t (\alpha, \beta, c, a)$ et

$$\begin{aligned}\sigma_t^2(\theta) &= c + a [Y_{t-1} - \mu_{t-1}(\theta)]^2 \\ \mu_t(\theta) &= \alpha + \beta \sigma_t^2(\theta)\end{aligned}$$

Le succès de ce type de modèles n'est pas seulement dû à ses propriétés mathématiques, mais aussi sur le fait qu'il propose une explication acceptable (en terme économique) à l'instabilité de la volatilité des titres dans le temps. En effet, il est raisonnable de penser qu'un investisseur qui révisé ses anticipations, tient compte de la valeur de la rentabilité des périodes précédentes. Ces anticipations modifient à la fois l'espérance de gains et le risque liés aux investissements, ce qui conduit assez naturellement à une modélisation ARCH.

Les modèles les plus utilisés pour des problèmes financiers sont les suivants :

Modèle ARCH(p) Engle (1982)

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i Y_{t-i}^2$$

avec $\alpha_0 > 0$ et $\forall i \in \{1 \dots p\}, \alpha_i \geq 0$

Modèle GARCH(p, q) Bollerslev (1986)

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^q \alpha_j Y_{t-j}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2$$

avec $\alpha_0 > 0, \forall j \in \{1 \dots q\}, \alpha_j \geq 0$ et $\forall i \in \{1 \dots p\}, \beta_i \geq 0$

Modèle GARCH-M Engle, Lilien, et Robbins (1987)

$$\begin{aligned}
Y_t &= X_t \beta + \delta \sigma_t + \varepsilon_t \\
\varepsilon &\sim \text{GARCH}(p, q) \\
V(\varepsilon_t | \varepsilon_{t-1}) &= \sigma_t^2
\end{aligned}$$

δ, β sont des paramètres réels et X_t un ensemble de variable(s) explicative(s).

Modèle EGARCH(p, q) Nelson (1990)

$$\ln \sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i [\phi Z_{t-i} + \gamma (|Z_{t-i}| - E\{|Z_{t-i}|\})] + \sum_{j=1}^q \beta_j \ln \sigma_{t-j}^2$$

Avec et $\forall t, Z(t) = \frac{Y_t}{\sigma_t}$, et pas de restrictions sur les paramètres.

Contrairement aux deux autres modèles, celui de Nelson permet une asymétrie de la variance conditionnelle. Cette asymétrie est constatée sur les marchés financiers : le risque n'étant probablement pas perçu de la même façon suivant que le marché est haussier ou baissier.

Modèle APARCH(p, q) Ding, Granger, et Engle (1993)

$$\sigma_t^\delta = \alpha_0 + \sum_{i=0}^p \alpha_i (|Y_{t-i}| - \gamma_i Y_{t-i})^\delta + \sum_{j=0}^q \beta_j \sigma_{t-1}^\delta$$

avec $\alpha_0 > 0, \forall i \in \{1 \dots p\}, \alpha_i \geq 0, -1 < \gamma_i < 1, \forall j \in \{1 \dots q\}, \beta_j \geq 0$, et $\delta > 0$ Comme le modèle EGARCH, ce modèle permet une asymétrie sur la variance conditionnelle. Ce modèle permet aussi d'expliquer le phénomène de "mémoire longue" cf. paragraphe 2.7).

2.2.2 Stationnarité du second ordre des modèles ARCH

Pour chaque modèle ARCH, il faut déterminer les conditions de stationnarité du second ordre à partir des équations qui déterminent la variance conditionnelle. On montre par exemple :

Proposition 10 *Un modèle ARCH(p) avec des coefficients $\alpha_i \geq 0$ est asymptotiquement stationnaire au second ordre si*

$$\sum_{i=1}^p \alpha_i < 1$$

On obtient de même la condition de stationnarité du second ordre pour le processus GARCH(p, q) :

Proposition 11 *Un modèle GARCH(p, q) avec des coefficients $\alpha_j \geq 0, \beta_i \geq 0$ est asymptotiquement stationnaire au second ordre si*

$$\sum_{j=1}^q \alpha_j + \sum_{i=1}^p \beta_i < 1$$

Démonstration

Introduisons l'innovation correspondant au carré du processus : $u_t = Y_t^2 - \sigma_t^2$. En remplaçant σ_t^2 par $Y_t^2 - u_t$ dans la formalisation GARCH, il vient :

$$\begin{aligned} Y_t^2 - u_t &= \alpha_0 + \sum_{j=1}^q \alpha_j Y_{t-j}^2 + \sum_{i=1}^p \beta_i (Y_{t-i}^2 - u_{t-i}) \\ \Leftrightarrow Y_t^2 &= \alpha_0 + \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) Y_{t-j}^2 + u_t - \sum_{i=1}^p \beta_i u_{t-i} \end{aligned}$$

avec la convention $\alpha_j = 0$ pour $j > q$ et $\beta_i = 0$ pour $i > q$.

C'est une sorte de représentation ARMA($\max(p, q), p$) du processus Y_t^2 sauf que le terme d'erreur $u = (u_t, t = 0, \dots)$ n'est pas nécessairement de variance constante.

En prenant l'espérance de chacun des membres de la dernière équation, il vient :

$$E(Y_t^2) = \alpha_0 + \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) E(Y_{t-i}^2)$$

Donc, si les racines du polynôme caractéristique

$$P(B) = 1 - \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) B^i$$

sont strictement à l'extérieur du disque unité, la suite des $E(Y_t^2)$ converge et le processus est asymptotiquement stationnaire au second ordre.

Il ne reste plus qu'à montrer que cette condition est équivalente à la condition $\sum_{i=1}^{\max(p,q)} \alpha_i + \beta_i < 1$.

\Rightarrow Comme on a $P(0) = 1$ et $P(1) = 1 - \sum_{i=1}^{\max(p,q)} \alpha_i + \beta_i$, il est nécessaire que $P(1) > 0$ sinon P aurait une racine réelle comprise entre 0 et 1.

\Leftarrow Supposons que P ait une racine z de module strictement inférieur à 1, alors

$$\begin{aligned} 1 &= \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) z^i \\ &= \left| \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) z^i \right| \\ &\leq \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) |z|^i \\ &\leq \sum_{i=1}^{\max(p,q)} \alpha_i + \beta_i \\ &< 1 \end{aligned}$$

D'où la contradiction. ■

2.2.3 Coefficient d'aplatissement d'un processus de type ARCH conditionnellement gaussien

Soit $Y = (Y_t, t = 0, \dots)$ un processus de type ARCH conditionnellement gaussien, centré et stationnaire. On peut écrire $Y_t | \underline{Y}_{t-1} \rightsquigarrow \mathcal{N}(0, \sigma_t^2)$. Par conséquent, les moments centrés conditionnels d'ordre 2 et 4 de ce processus sont liés par :

$$E(Y_t^4 | \underline{Y}_{t-1}) = 3 \left[E(Y_t^2 | \underline{Y}_{t-1}) \right]^2$$

Le coefficient d'aplatissement de Y_t est défini par :

$$\gamma_4(Y_t) = \frac{E[Y_t^4]}{E[Y_t^2]^2}$$

On peut écrire :

$$E(Y_t^4) = E \left[E(Y_t^4 | \underline{Y}_{t-1}) \right] = 3 E \left[E(Y_t^2 | \underline{Y}_{t-1}) \right]^2$$

et donc

$$\begin{aligned} \gamma_4(Y_t) &= 3 \frac{E \left[E(Y_t^2 | \underline{Y}_{t-1}) \right]^2}{E[Y_t^2]^2} \\ &= 3 \frac{E \left[E(Y_t^2 | \underline{Y}_{t-1}) \right]^2 - E[Y_t^2]^2 + E[Y_t^2]^2}{E[Y_t^2]^2} \\ &= 3 + 3 \frac{E \left[E(Y_t^2 | \underline{Y}_{t-1}) \right]^2 - E[Y_t^2]^2}{E[Y_t^2]^2} \\ &= 3 + 3 \frac{E \left[E(Y_t^2 | \underline{Y}_{t-1}) \right]^2 - E \left[E(Y_t^2 | \underline{Y}_{t-1}) \right]^2}{E[Y_t^2]^2} \\ &= 3 + 3 \frac{V \left[E(Y_t^2 | \underline{Y}_{t-1}) \right]}{E[Y_t^2]^2} \\ &= 3 + 3 \frac{V(\sigma_t^2)}{E[Y_t^2]^2} \end{aligned}$$

Le deuxième terme de cette somme est strictement positif dès que σ_t n'est pas une constante *i.e.* dès que le processus est un "vrai" modèle ARCH. Le coefficient d'aplatissement est alors strictement plus grand que 3 et la distribution est leptokurtique.

2.3 Estimation des paramètres des modèles ARCH

2.3.1 Le pseudo maximum de vraisemblance

On note θ le vecteur des paramètres du modèle considéré et $l_t(y; \theta)$ la vraisemblance de Y_t conditionnellement à son passé. La fonction de vraisemblance du vecteur $(Y_1 \dots Y_T)$ conditionnelle à

Y_0 est alors donnée par :

$$L(y; \theta) = \prod_{t=1}^T l_t(y; \theta)$$

L'estimateur est alors donné par

$$\hat{\theta}_T = \underset{\theta}{\text{ArgMax}} L(y; \theta)$$

Les propriétés de $\hat{\theta}_T$ appelé “Pseudo maximum de vraisemblance” dépendent à la fois de la vraie distribution sous-jacente et de celle utilisée pour le calcul (ici, on utilise la loi normale). Sous diverses conditions de régularités, cet estimateur est convergent, même si la loi conditionnelle de Y_t n'est pas normale. De plus cet estimateur est asymptotiquement sans biais, normal et sa matrice de covariance asymptotique est donnée par :

$$V_{as}(\hat{\theta}_T - \theta) = \frac{1}{T} J^{-1} I J^{-1}$$

avec

$$I = E_0 \left[\frac{\partial \ln l_t(y; \theta)}{\partial \theta} \frac{\partial \ln l_t(y; \theta)}{\partial \theta'} \right]$$

et

$$J = E_0 \left[- \frac{\partial^2 \ln l_t(y; \theta)}{\partial \theta^2} \right]$$

et E_0 désigne l'espérance prise sous la vraie loi.

Il est, en pratique, impossible de calculer les deux matrices I et J . Il faut donc, à partir des données, en calculer des estimateurs et en déduire un estimateur de la matrice de covariance de $\hat{\theta}_T$. Le plus simple consiste à remplacer l'espérance E_0 par la moyenne empirique, et la vraie valeur de θ par sa valeur estimée $\hat{\theta}_T$. On obtient alors :

$$\hat{I}_T = \frac{1}{T} \sum_{t=1}^t \frac{\partial \ln l_t(y; \hat{\theta}_T)}{\partial \theta} \frac{\partial \ln l_t(y; \hat{\theta}_T)}{\partial \theta'}$$

$$\hat{J}_T = - \frac{1}{T} \sum_{t=1}^t \frac{\partial^2 \ln l_t(y; \hat{\theta}_T)}{\partial \theta^2}$$

exemple : échantillonnage

On considère $Y_t, t = 1 \dots T$, une suite de variables aléatoires i.i.d. de moyenne m et de variance σ^2 . Dans ces conditions, on a immédiatement :

$$l_t(y; \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2} \left(\frac{Y_t - m}{\sigma} \right)^2 \right\}$$

On obtient très facilement :

$$\hat{\theta}_T = \left(\begin{array}{l} \hat{m}_T = \frac{1}{T} \sum_{t=1}^T Y_t \\ \hat{\sigma}_T^2 = \frac{1}{T} \sum_{t=1}^T (Y_t - \hat{m}_T)^2 \end{array} \right)$$

et

$$\frac{\partial \ln l_t(y; \theta)}{\partial \theta} = \begin{pmatrix} \frac{Y_t - m}{\sigma^2} \\ \frac{(Y_t - m)^2}{2\sigma^4} - \frac{1}{2\sigma^2} \end{pmatrix}$$

$$\frac{\partial^2 \ln l_t(y; \theta)}{\partial \theta^2} = \begin{pmatrix} -\frac{1}{\sigma^2} & -\frac{Y_t - m}{\sigma^4} \\ -\frac{Y_t - m}{\sigma^4} & -\frac{(Y_t - m)^2}{\sigma^6} + \frac{1}{2\sigma^4} \end{pmatrix}$$

Il vient donc

$$I = \begin{pmatrix} \frac{1}{\sigma^2} & \frac{1}{2\sigma^3} E_0(u_t^3) \\ \frac{1}{2\sigma^3} E_0(u_t^3) & \frac{1}{4\sigma^4} [E_0(u_t^4) - 1] \end{pmatrix}$$

où $u_t = \frac{Y_t - m}{\sigma}$ est la variable centrée réduite.

et

$$J = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

Par conséquent :

$$V_{as}(\hat{\theta}_T - \theta) = \frac{1}{T} \begin{pmatrix} \sigma^2 & \sigma^3 E[u_t^3] \\ \sigma^3 E[u_t^3] & \sigma^4 \{E[u_t^4] - 1\} \end{pmatrix}$$

Si la vraie loi de Y_t est la loi normale, on a $E[u_t^3] = 0$ et les deux estimateurs \hat{m}_T et $\hat{\sigma}_T^2$ sont non corrélés. Au contraire, plus la vraie loi de Y_t présente des asymétries, plus les deux estimateurs sont corrélés.

2.3.2 modèle de régression avec résidus ARCH

On suppose ici que le processus $(Y_t, t \in \mathbb{N})$ est décrit par une équation du type :

$$\forall t \in \mathbb{N}, Y_t = m(\theta, \underline{Y}_{t-1}, \underline{X}_t) + h^{1/2}(\theta, \underline{Y}_{t-1}, \underline{X}_t) u_t$$

où u_t est un bruit blanc de variance égale à 1, et $(X_t, t \in \mathbb{N})$ est un processus donné.

$m(\theta, \underline{Y}_{t-1}, \underline{X}_t)$ est alors l'espérance conditionnelle de Y_t sachant le passé et $h(\theta, \underline{Y}_{t-1}, \underline{X}_t)$ est sa variance conditionnelle. Pour simplifier les écritures, on notera $m_t(\theta) = m(\theta, \underline{Y}_{t-1}, \underline{X}_t)$ et $h_t(\theta) = h(\theta, \underline{Y}_{t-1}, \underline{X}_t)$.

Le logarithme de la pseudo vraisemblance est alors donné par :

$$\ln L_{y,x}(\theta) = -\frac{1}{2} \left\{ T \ln \sqrt{2\pi} + \sum_{t=1}^T \ln h_t(\theta) + \sum_{t=1}^T \frac{(y_t - m_t(\theta))^2}{h_t(\theta)} \right\}$$

Les conditions du premier ordre sont donc données par :

$$\frac{\partial \ln L_{y,x}(\hat{\theta}_T)}{\partial \theta} = \sum_{t=1}^T \frac{1}{h_t^{1/2}(\hat{\theta}_T)} \frac{\partial m_t(\hat{\theta}_T)}{\partial \theta} u_t(\hat{\theta}_T) + \frac{1}{2} \sum_{t=1}^T \frac{1}{h_t(\hat{\theta}_T)} \frac{\partial h_t(\hat{\theta}_T)}{\partial \theta} [u_t^2(\hat{\theta}_T) - 1] = 0$$

où $u_t(\hat{\theta}_T) = \frac{y_t - m_t(\hat{\theta}_T)}{h_t^{1/2}(\hat{\theta}_T)}$ est le résidu centré réduit.

On obtient ensuite :

$$J = E_0 \left[\frac{1}{h_t(\theta)} \frac{\partial m_t(\theta)}{\partial \theta} \frac{\partial m_t(\theta)}{\partial \theta'} + \frac{1}{2h_t^2(\theta)} \frac{\partial h_t(\theta)}{\partial \theta} \frac{\partial h_t(\theta)}{\partial \theta'} \right]$$

et

$$I = E_0 \left[\frac{1}{4h_t^2(\theta)} \frac{\partial h_t(\theta)}{\partial \theta} \frac{\partial h_t(\theta)}{\partial \theta'} (K_t(\theta) - 1) + \frac{1}{h_t(\theta)} \frac{\partial m_t(\theta)}{\partial \theta} \frac{\partial m_t(\theta)}{\partial \theta'} \right. \\ \left. + \frac{1}{2h_t(\theta)^{3/2}} \left(\frac{\partial h_t(\theta)}{\partial \theta} \frac{\partial m_t(\theta)}{\partial \theta'} + \frac{\partial m_t(\theta)}{\partial \theta} \frac{\partial h_t(\theta)}{\partial \theta'} \right) A_t(\theta) \right]$$

où $A_t(\theta) = E \left[u_t^3(\theta) \mid Y_{t-1}, X_t \right]$ est le coefficient d'asymétrie conditionnel et $K_t(\theta) = E \left[u_t^4(\theta) \mid Y_{t-1}, X_t \right]$ est le coefficient d'aplatissement conditionnel.

On estime alors les matrices I et J par

$$\bar{J}_T = \frac{1}{T} \sum_{t=1}^T \left\{ \frac{1}{h_t(\hat{\theta}_T)} \frac{\partial m_t}{\partial \theta}(\hat{\theta}_T) \frac{\partial m_t}{\partial \theta'}(\hat{\theta}_T) + \frac{1}{2h_t^2(\hat{\theta}_T)} \frac{\partial h_t}{\partial \theta}(\hat{\theta}_T) \frac{\partial h_t}{\partial \theta'}(\hat{\theta}_T) \right\} \\ \bar{I}_T = \frac{1}{T} \sum_{t=1}^T \left\{ \frac{1}{4h_t^2(\hat{\theta}_T)} \frac{\partial h_t}{\partial \theta}(\hat{\theta}_T) \frac{\partial h_t}{\partial \theta'}(\hat{\theta}_T) \left(u_t^4(\hat{\theta}_T) - 1 \right) + \frac{1}{h_t(\hat{\theta}_T)} \frac{\partial m_t}{\partial \theta}(\hat{\theta}_T) \frac{\partial m_t}{\partial \theta'}(\hat{\theta}_T) \right. \\ \left. + \frac{1}{2h_t(\hat{\theta}_T)^{3/2}} \left\{ \frac{\partial h_t}{\partial \theta}(\hat{\theta}_T) \frac{\partial m_t}{\partial \theta'}(\hat{\theta}_T) + \frac{\partial m_t}{\partial \theta}(\hat{\theta}_T) \frac{\partial h_t}{\partial \theta'}(\hat{\theta}_T) \right\} u_t^3(\hat{\theta}_T) \right\}$$

Remarquons que si la loi conditionnelle (*i.e.* la loi de $u_t(\theta)$) est normale, alors $A_t(\theta) = 0$ et $K_t(\theta) = 3$ et $I = J$.

exemple : modèle AR(1)-ARCH(1)

On a alors $m_t(\theta) = m_t(\beta) = \beta_0 + \beta_1 Y_{t-1}$ et $h_t(\theta) = \alpha_0 + \alpha_1 (Y_{t-1} - m_{t-1}(\beta))^2$

Remarquons que dans ce modèle, si on centre le processus Y_t en posant $Z_t = Y_t - m_t(\theta)$, Z_t est un ARCH(1).

En posant $\theta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \alpha_0 \\ \alpha_1 \end{pmatrix}$, il vient :

$$\frac{\partial m_t(\theta)}{\partial \theta} = \begin{pmatrix} 1 \\ y_{t-1} \\ 0 \\ 0 \end{pmatrix}$$

et

$$\frac{\partial h_t(\theta)}{\partial \theta} = \begin{pmatrix} -2\alpha_1(y_{t-1} - m_{t-1}(\theta)) \frac{\partial m_{t-1}(\theta)}{\partial \beta_0} \\ -2\alpha_1(y_{t-1} - m_{t-1}(\theta)) \frac{\partial m_{t-1}(\theta)}{\partial \beta_1} \\ 1 \\ (y_{t-1} - m_{t-1}(\theta))^2 \end{pmatrix} = \begin{pmatrix} -2\alpha_1(y_{t-1} - m_{t-1}(\theta)) \\ -2\alpha_1(y_{t-1} - m_{t-1}(\theta))y_{t-2} \\ 1 \\ (y_{t-1} - m_{t-1}(\theta))^2 \end{pmatrix}$$

L'estimateur du pseudo-maximum de vraisemblance est alors solution du système d'équations :

$$\begin{aligned}
\sum_{t=1}^T \frac{z_t(\hat{\theta}_T) - \hat{\alpha}_1 z_{t-1}(\hat{\theta}_T)}{h_t(\hat{\theta}_T)} &= 0 \\
\sum_{t=1}^T \frac{y_{t-1} z_t(\hat{\theta}_T) - \hat{\alpha}_1 y_{t-2} z_{t-1}(\hat{\theta}_T)}{h_t(\hat{\theta}_T)} &= 0 \\
\sum_{t=1}^T \frac{u_t(\hat{\theta}_T)^2 - 1}{2 h_t(\hat{\theta}_T)} &= 0 \\
\sum_{t=1}^T \frac{z_{t-1}(\hat{\theta}_T) (u_t(\hat{\theta}_T)^2 - 1)}{2 h_t(\hat{\theta}_T)} &= 0
\end{aligned}$$

où $z_t(\hat{\theta}_T) = y_t - m_t(\hat{\theta}_T)$

2.4 Tests dans les modèles ARCH

Comme on connaît la matrice de covariance asymptotique de l'estimateur du pseudo maximum de vraisemblance et que celui-ci est gaussien, on peut facilement construire un test d'hypothèse du type $H_0 : \theta_i = \theta_i^{(0)}$ contre $H_1 : \theta_i \neq \theta_i^{(0)}$, où θ_i est la i^e coordonnée de θ .

En effet, la statistique de test $\mathcal{T} = \frac{\hat{\theta}_i - \theta_i^{(0)}}{\sqrt{\hat{V}_{as}(\hat{\theta}_i)}}$ suit asymptotiquement une loi de Student à $T - r$ d.d.l. (si T est le nombre d'observations et r le nombre de paramètres estimés). Comme T est grand (puisqu'on utilise la matrice de covariance asymptotique), la valeur critique du test peut être lue dans la table de la loi normale, plutôt que dans celle de la loi de Student.

2.5 Exemple sous R

Voici comment on peut programmer le modèle AR(1)-ARCH(1) dans R :

2.5.1 fichier AR-ARCH.R

```

getalpha <- function(teta)
{
  teta[3:4]
}
getbeta <- function(teta)
{
  teta[1:2]
}
mt <- function(Y, teta, tps)
{
  teta[1] + teta[2] * Y[tps-1]
}
dmtsurdteta <- function(Y, teta, tps)
{
  c(1, Y[tps-1], 0, 0)
}
ht <- function(Y, teta, tps)
{
  mtm1 <- mt(Y, teta, tps-1)
  teta[3] + teta[4] * (Y[tps-1] - mtm1)^2
}

```

```

zt <- function(Y, teta, tps)
{
  Y[tps] - mt(Y, teta, tps)
}
ut <- function(Y, teta, tps)
{
  zt(Y, teta, tps)/sqrt(ht(Y, teta, tps))
}
dhtsurdteta <- function(Y, teta, tps)
{
  ztm1 <- zt(Y, teta, tps-1)
  -2*teta[4]*ztm1*c(1,Y[tps-2],0,0) +c(0, 0, 1, ztm1^2)
}
logpmv <- function(teta, Y)
{
  TT <- length(Y)
  f <- 0
  nbparam <- length(teta)
  for (tt in 3:TT)
  {
    hht <- ht(Y, teta, tt)
    mmt <- mt(Y, teta, tt)
    ZZt <- Y[tt] - mmt
    ut <- ZZt/sqrt(hht)
    f <- f - 0.5*(log(hht) + ut^2)
  }
  -f
}
gradlogpmv <- function(teta, Y)
{
  TT <- length(Y)
  nbparam <- length(teta)
  g <- 0*c(1:nbparam)
  for (tt in 3:TT)
  {
    hht <- ht(Y, teta, tt)
    mmt <- mt(Y, teta, tt)
    ZZt <- Y[tt] - mmt
    ut <- ZZt/sqrt(hht)
    dmt <- dmtsurdteta(Y, teta, tt)
    dht <- dhtsurdteta(Y, teta, tt)
    g <- g + ut/sqrt(hht)*dmt + 0.5*(ut^2-1)*dht
  }
  -g
}
IChapeau <-function(Y, teta)
{
  TT <- length(Y)
  NbParam <- length(teta)
  Ic <- matrix(data=0, nrow=NbParam, ncol=NbParam)
  for (tt in 3:TT)
  {
    hht <- ht(Y, teta, tt)
    mmt <- mt(Y, teta, tt)
    dmt <- dmtsurdteta(Y, teta, tt)
  }
}

```

```

    dht <- dhtsurdteta(Y, teta, tt)
    ut <- (Y[tt]-mmt)/sqrt(hht)
    Kt <- ut^4
    At <- ut^3
    Ic <- Ic+(Kt-1)/(4*hht^2)*dht%%t(dht)+dmt%%t(dmt)/hht
        +At/(2*hht^1.5)*(dht%%t(dmt)+dmt%%t(dht))
  }
  Ic <- Ic/(TT-2)
  Ic
}
JChapeau <- function(Y, teta)
{
  TT <- length(Y)
  NbParam <- length(teta)
  Jc <- matrix(data=0, nrow=NbParam, ncol=NbParam)
  for (tt in 3:TT)
  {
    hht <- ht(Y, teta, tt)
    mmt <- mt(Y, teta, tt)
    dmt <- dmtsurdtdeta(Y, teta, tt)
    dht <- dhtsurdteta(Y, teta, tt)
    Jc <- Jc+dmt%%t(dmt)/hht+dht%%t(dht)/(2*hht^2)
  }
  Jc <- Jc/(TT-2)
  Jc
}
Vas <- function(Y, teta)
{
  Jc <- JChapeau(Y, teta)
  Ic <- IChapeau(Y, teta)
  Jcm1 <- ginv(Jc)
  1/length(Y)*Jcm1%%Ic%%Jcm1
}
EstimARCH <- function(Beta0, Alpha0, Y)
{
  Teta0 <- c(Beta0, Alpha0)
  Aux <- optim(par=Teta0, fn = logpmv, gr=gradlogpmv,
    method='BFGS', control=list(trace=0), lower = -Inf,
    upper = Inf, hessian = FALSE, Y)
  names(Aux$par) <- c("beta0", "beta1", "c", "a1")
  VarAsympt <- Vas(Y, Aux$par)
  toto <- as.data.frame(VarAsympt)
  names(toto) <- c("beta0", "beta1", "c", "a1")
  row.names(toto) <- c("beta0", "beta1", "c", "a1")
  VarAsympt <- as.matrix(toto)
  BetaChapeau <- getbeta(Aux$par)
  AlphaChapeau <- getalpha(Aux$par)
  Test<-Aux$par/sqrt(diag(VarAsympt))
  PValeur <- (1-pnorm(Test))*2
  BetaTest <- getbeta(Test)
}

```

```

AlphaTest <- getalpha(Test)
BetaPValeur <- getbeta(PValeur)
AlphaPValeur <- getalpha(PValeur)
TT <- length(Y)
Residuals.1 <- Y[2:TT] - BetaChapeau['beta0'] - BetaChapeau['beta1']*Y[1:(TT-1)]
hht <- 0*c(1:(TT-1))
hht[1] <- AlphaChapeau['c']
for (tt in 2:(TT-1))
{   hht[tt] <- AlphaChapeau['c'] + AlphaChapeau['a1']*Residuals.1[tt-1]^2
}
Residuals.2 <- Residuals.1/sqrt(hht)
Res<-list(BetaC=BetaChapeau, AlphaC=AlphaChapeau, TestBeta=BetaTest,
          TestAlpha=AlphaTest, PValBeta=BetaPValeur, PValAlpha=AlphaPValeur,
          VarAsympt=VarAsympt, Residuals.1=Residuals.1, Residuals.2=Residuals.2,
          LogVrais=-Aux$value)
class(Res) <- "AR1ARCH1"
return(Res)
}
print.AR1ARCH1 <- function(x, ...)
{
  Aux <- matrix(c(x$BetaC, x$AlphaC),nrow=4, ncol=1)
  colnames(Aux) <- "Estimateur"
  rownames(Aux) <- c("Beta0", "Beta1", "Alpha0", "Alpha1")
  print(Aux)
  cat("\n")
  cat(sprintf("Log-vraisemblance=%f\n", x$LogVrais))
}
summary.AR1ARCH1 <- function(x, ...)
{
  Aux <- matrix(c(x$BetaC, x$AlphaC),nrow=4, ncol=1)
  colnames(Aux) <- "Estimateur"
  rownames(Aux) <- c("Beta0", "Beta1", "Alpha0", "Alpha1")
  Aux1 <- matrix(c(sqrt(diag(x$VarAsympt))), ncol=1)
  colnames(Aux1) <- "Ecart-type"
  Aux2 <- matrix(c(x$TestBeta, x$TestAlpha), ncol=1)
  colnames(Aux2) <- "Student"
  Aux3 <- matrix(c(x$PValBeta, x$PValAlpha), ncol=1)
  colnames(Aux3) <- "P Valeur"
  print(cbind(Aux, Aux1, Aux2, Aux3))
  cat("\n")
  cat(sprintf("Log-vraisemblance=%f\n", x$LogVrais))
}
ProcDeuxEtapes <- function(Y, X, p)
## Modèle de régression avec résidus ARCH(p)
{   TT <- length(Y)
    Res1 <- lm(Y~1+X)

```

```

BetaTilda <- Res1$coefficients
EpsTilda <- as.vector(Res1$residuals)
XX <- matrix(data=0, ncol=p, nrow=TT-p)
for (i in 1:p)
{
  XX[,i]<-EpsTilda[(p+1-i):(TT-i)]^2
}
Res2 <- lm(EpsTilda[(p+1):TT]^2~1+XX)$coefficients

##calcul de la variance conditionnelle
ht <- c(1:TT)
for (tt in 1:p)
{
  ht[tt] <- Res2[1]
  if (tt > 1)
  {
    for (i in 1:(tt-1))
    {
      ht[tt] <- ht[tt]+Res2[i+1]*EpsTilda[tt-i]^2
    }
  }
}
for (tt in (p+1):TT)
{
  ht[tt] <- Res2[1]
  if (tt > 1)
  {
    for (i in 1:p)
    {
      ht[tt] <- ht[tt]+Res2[i+1]*EpsTilda[tt-i]^2
    }
  }
}

#Recalcul des betas
Res1 <- lm(Y~1+X, weights=1/sqrt(ht))
BetaTildaTilda <- Res1$coefficients
for (i in 1:length(BetaTildaTilda))
{
  names(BetaTildaTilda)[i] <- paste('beta',i-1,sep="")
}

EpsTildaTilda <- Res1$residuals
XX <- matrix(data=0, ncol=p, nrow=TT-p)
for (i in 1:p)
{
  XX[,i]<-EpsTilda[(p+1-i):(TT-i)]^2
}
Res2 <- lm(EpsTildaTilda[(p+1):TT]^2~1+XX, weights=1/sqrt(2*ht[(p+1):TT]^2))
ctildatilda <- Res2$coefficients[1]
names(ctildatilda) <- "c"
atildatilda <- Res2$coefficients[2:(p+1)]
for (i in 1:p)
{
  names(atildatilda)[i] <- paste('a',i,sep="")
}
list(BetaTildaTilda=BetaTildaTilda, ctildatilda=ctildatilda, atildatilda=atildatilda)

```



```

}

calculomega <- function(Eps, H)
{
  TT <- length(Eps)
  Omega <- matrix(data=0, nrow = H, ncol = H)
  for (tt in (H+2):TT)
  {
    Omega <- Omega + (Eps[(tt-H):(tt-1)]*Eps[tt]^2) %*% t(Eps[(tt-H):(tt-1)])
  }
  Omega / (TT-H-1)
}

test.portmanteau <- function(Eps, H)
{
  Omega <- calculomega(Eps, H)
  Omegam1 <- ginv(Omega)
  MonGamma <- acf(Eps, lag.max = H, type = "correlation", plot = FALSE)$acf[2:(H+1)]
  QH <- length(Eps)*t(MonGamma) %*% Omegam1 %*% MonGamma
  P.Value <- 1-pchisq(QH,H)
  Res <- list(QH=QH, P.Value = P.Value, gamma=MonGamma, df=H)
  class(Res) <- "portmanteau"
  return(Res)
}

print.portmanteau<-function(x, ...)
{
  Aux1 <- matrix(seq(1,x$df), ncol=1)
  Aux2 <- matrix(x$gamma, ncol=1)
  Aux <- cbind(Aux1, Aux2)
  colnames(Aux)<-c("Retard","Autocorrélation")
  rownames(Aux)<-rep(" ", x$df)
  print(Aux)
  Aux <- cbind(x$QH, x$P.Value, x$df)
  colnames(Aux) <- c("QH", "P-Valeur", "D.D.L")
  rownames(Aux)<-" "
  print(Aux)
}

test.score <- function(Y, X, p)
## Modèle de régression avec résidus ARCH(p)
{
  TT <- length(Y)
  if (length(X) == 0) { Eps <- Y - mean(Y) } else { Eps <- lm(Y~1+X)$residuals}
  XX <- matrix(data=0, ncol=p, nrow=TT-p)
  for (i in 1:p)
  {
    XX[,i]<-Eps[(p+1-i):(TT-i)]^2
  }
  R2 <- summary(lm(Eps[(p+1):TT]^2~1+XX))$r.squared
  Stat = TT*R2
}

```

```

P.Value = 1 - pchisq(Stat, p)
Res<-list(Stat=TT*R2, P.Value=P.Value, df=p)
class(Res)<-"test.score"
return(Res)
}
print.test.score<-function(x, ...)
{
  Aux <- cbind(x$Stat, x$P.Value, x$df)
  colnames(Aux)<-c("Stat", "P-Valeur", "D.D.L.")
  rownames(Aux)<-" "
  print(Aux)
}

```

2.5.2 résultats

```

> library(MASS)
> source("AR-ARCH.R")
> #Rt contient les rentabilités journalières du CAC 40
> TT <- length(Rt)
> # Test du score pour le modèle AR(1)-ARCH(1)
> ResTestScore <- test.score(Rt[2:TT], Rt[1:(TT-1)], 1)
> print(ResTestScore)
      Stat P-Valeur D.D.L.
147.9260      0      1
>
> # Estimation des paramètres
>
> # Initialisation grace à la procédure en 2 étapes
> Teta0 <- ProcDeuxEtapes(Rt[2:TT],Rt[1:(TT-1)], 1)
> Beta0 <- Teta0$BetaTildaTilda
> Alpha0 <- c(Teta0$ctildatilda, Teta0$atildatilda)
> Teta0b <- c(Beta0, Alpha0)
> names(Teta0b) <- c("Beta0", "Beta1", "Alpha0", "Alpha1")
> print(Teta0b)
      Beta0      Beta1      Alpha0      Alpha1
0.0003083719 0.0261300351 0.0001348942 0.1954747879
>
> # Maximum de vraisemblance
>
> Res <- EstimARCH(Beta0, Alpha0, Rt)
> summary(Res)
      Estimateur  Ecart-type  Student    P Valeur
Beta0  0.0003758029 1.818244e-04  2.066845 3.874874e-02
Beta1  0.0515530183 2.113987e-02  2.438663 1.474171e-02
Alpha0 0.0001339734 5.732059e-06 23.372641 0.000000e+00
Alpha1 0.2026949994 2.850588e-02  7.110638 1.155076e-12

```

```
Log-vraisemblance=18244.329399
> cat("\n")

> # Etude des résidus
> # Test du portmanteau sur les résidus de la régression
> ResTestPort1 <- test.portmanteau(Res$Residuals.1, 20)
> print(ResTestPort1)
  Retard Autocorrélation
      1    -0.001251963
      2    -0.017067741
      3    -0.041499517
      4     0.015583881
      5    -0.030950404
      6    -0.019615139
      7    -0.019427721
      8     0.019179902
      9    -0.001154291
     10     0.007823751
     11     0.005873862
     12    -0.016164606
     13     0.012710129
     14     0.022060552
     15     0.032786739
     16    -0.006792242
     17    -0.020630568
     18    -0.019114552
     19    -0.026498194
     20     0.014105126
      QH P-Valeur D.D.L
707108209      0      20
> ResTestPort2 <- test.portmanteau(Res$Residuals.2, 20)
> print(ResTestPort2)
  Retard Autocorrélation
      1    -0.0008606933
      2    -0.0161040489
      3    -0.0346629906
      4     0.0194900951
      5    -0.0273763475
      6    -0.0104610637
      7    -0.0228480559
      8     0.0155412211
      9     0.0033037683
     10     0.0053453776
     11     0.0036091329
     12    -0.0083628657
     13     0.0112579240
```

```

14      0.0193194634
15      0.0219735177
16     -0.0062899555
17     -0.0176937169
18     -0.0106221297
19     -0.0267569004
20      0.0192829180
      QH  P-Valeur D.D.L
19.22904 0.5069859    20
> ResTestScore1 <- test.score(Res$Residuals.1[2:(TT-1)], Res$Residuals.1[1:(TT-2)], 1)
> print(ResTestScore1)
      Stat P-Valeur D.D.L.
148.8607      0      1
> ResTestScore2 <- test.score(Res$Residuals.2[2:(TT-1)], Res$Residuals.2[1:(TT-2)], 1)
> print(ResTestScore2)
      Stat P-Valeur D.D.L.
1.050465 0.3054001      1
> ResTestScore2.b <- test.score(Res$Residuals.2[2:(TT-1)], Res$Residuals.2[1:(TT-2)], 2)
> print(ResTestScore2.b)
      Stat P-Valeur D.D.L.
124.4367      0      2
> library(fSeries)
> basicStats(Res$Residuals.1)
      round.ans..digits...6.
nobs                4716.000000
NAs                  0.000000
Minimum              -0.076808
Maximum              0.067686
1. Quartile          -0.006963
3. Quartile           0.007242
Mean                 -0.000043
Median                0.000037
Sum                  -0.202894
SE Mean               0.000189
LCL Mean              -0.000413
UCL Mean              0.000327
Variance              0.000168
Stdev                 0.012967
Skewness              -0.115191
Kurtosis              2.930713

> basicStats(Res$Residuals.2)
      round.ans..digits...6.
nobs                4716.000000
NAs                  0.000000
Minimum              -6.476226
Maximum              5.846098

```

1. Quartile	-0.561133
3. Quartile	0.580467
Mean	-0.006409
Median	0.002588
Sum	-30.226356
SE Mean	0.014578
LCL Mean	-0.034989
UCL Mean	0.022170
Variance	1.002236
Stdev	1.001117
Skewness	-0.238632
Kurtosis	2.757685

Le modèle obtenu est le suivant :

$$\begin{aligned} m_t &= 0.00038 + 0.05155 R_{t-1} \\ h_t &= 0.00013 + 0.20269 (R_t - m_{t-1})^2 \end{aligned}$$

Les paramètres sont tous significativement (à 5%) non nuls. Les statistiques élémentaires ne font pas apparaître une diminution sensible des coefficients d'asymétrie et d'aplatissement des résidus standardisés (la série `Res.Residuals.2`). Le modèle choisi ici n'est pas suffisant pour modéliser correctement l'évolution des rentabilités journalières du CAC.

2.6 Processus à mémoire longue

L'objectif de ce paragraphe est de montrer une autre caractéristique importante du processus des rentabilités boursières.

Tout d'abord, définissons un processus à *mémoire longue*.

Définition 17 *Un processus Y_t est un processus à mémoire longue s'il existe un nombre réel $\alpha \in]0, 1[$ et une constante $c > 0$ tels que :*

$$\lim_{h \rightarrow \infty} \frac{\rho(h)}{h^{-\alpha}} = c$$

où $\rho(h)$ est la fonction d'autocorrélation de Y_t .

Cela signifie que la décroissance des coefficients d'autocorrélation est lente (hyperbolique) ($\rho(h) \sim \frac{c}{h^\alpha}$ quand $h \rightarrow \infty$). Pour un modèle ARMA stationnaire, on a $\rho(h) < c a^h$ où $a \in]0, 1[$: la décroissance des coefficients d'autocorrélation est géométrique.

2.6.1 Comment obtenir des processus stationnaires à mémoire longue?

On part de la constatation suivante :

Considérons un bruit blanc (u_t). Ce processus est stationnaire (au second ordre). On peut définir le processus ARIMA(0, d=0, 0) (X_{1t}) par $(Id - L)^0 X_{1t} = X_{1t} = u_t$. Ce processus est à mémoire courte (dans ce cas précis, il est même sans mémoire puisque tous les coefficients

d'autocorrélations sont nuls.). On peut aussi définir le processus ARIMA(0, d=1, 0) (X_{2t}) par $(Id - L)^1 X_{2t} = u_t$. Ce processus est intégré d'ordre 1, il est non stationnaire et donc à mémoire "infinie" parce que toute perturbation de X_{2t} a des conséquences permanentes sur le futur de la série.

L'idée est alors de définir un **"bruit blanc fractionnaire"** de la manière suivante :

$$\Delta^d X_t = (Id - L)^d X_t = u_t$$

avec $d \in \mathbb{R}$ et d'utiliser le développement binomial

$$\Delta^d = (Id - L)^d = Id - dL - \frac{d(1-d)}{2!}L^2 - \frac{d(1-d)(2-d)}{3!}L^3 + \dots = \sum_{j=0}^{+\infty} \pi_j L^j$$

où $\pi_0 = 1$ et pour $j > 0$, $\pi_j = \prod_{k=1}^j \frac{k-1-d}{k}$.

On obtient la propriété suivante :

Théorème 2 Si X_t est un bruit blanc fractionnaire, $\Delta^d X_t = u_t$, on a

1. si $d < \frac{1}{2}$, X_t est stationnaire et possède une représentation moyenne mobile infinie :

$$X_t = \Psi(L) u_t = \sum_{k=0}^{+\infty} \psi_k u_{t-k}$$

où

$$\psi_k = \prod_{j=0}^{k-1} \frac{d+j}{j+1}$$

et

$$\lim_{k \rightarrow \infty} \psi_k = \frac{k^{d-1}}{\Gamma(d)}$$

2. si $d > -\frac{1}{2}$, X_t est inversible et possède une représentation autorégressive infinie :

$$\Pi(L) X_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} = u_t$$

où

$$\pi_j = \prod_{k=1}^j \frac{k-1-d}{k}$$

et

$$\lim_{j \rightarrow \infty} \pi_j = \frac{k^{-d-1}}{\Gamma(-d)}$$

3. si $-\frac{1}{2} < d < \frac{1}{2}$ et $\text{var}(u_t) = 1$, alors la fonction d'autocovariance $\gamma(h)$ vérifie

$$\gamma(h) \sim \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} h^{2d-1}, \text{ quand } h \rightarrow \infty$$

où $\Gamma(x) = x \Gamma(x-1) = \int_0^{+\infty} t^{x-1} e^{-t} dt$ est la fonction gamma d'Euler.

On constate donc que si $d \in]-1/2, 1/2[$, le bruit blanc fractionnaire est stationnaire et inversible, les coefficients ψ_k et π_j décroissent à une vitesse hyperbolique et pas géométrique comme dans le cas des modèles ARMA. Ce processus est donc à stationnaire, inversible et à mémoire longue.

Il suffit maintenant de généraliser la définition pour obtenir un processus ARFIMA (processus ARMA fractionnaire).

Définition 18 *Un processus (X_t) stationnaire admet une représentation ARFIMA(p, d, q), s'il existe*

- deux polynômes retards $\Phi(L)$ de degré p et $\Theta(L)$ de degré q ,
- un bruit blanc u_t ,

tels que

$$\begin{aligned}\Delta^d \varepsilon_t &= u_t \\ \Phi(L) X_t &= \Theta(L) \varepsilon_t\end{aligned}$$

Autrement dit, un processus ARFIMA est un processus ARMA pour lequel on a remplacé le bruit blanc par un bruit blanc fractionnaire. Le graphique FIG. (2.5) montre une trajectoire et la fonction d'autocorrélation d'un bruit blanc fractionnaire avec $d = 0.45$ (ou encore d'un modèle ARFIMA(0, 0.45, 0))

2.6.2 Exposant de Hurst

Hurst (1951) a proposé une statistique permettant de détecter le phénomène de mémoire longue dans les séries chronologiques. Cette statistique appelée *analyse R/S* conduit à estimer un paramètre, *l'exposant de Hurst*, qui permet d'identifier la "nature de la mémoire" (longue ou courte) des séries chronologiques.

Définition 19 *Si X_t est une série chronologique, on définit la statistique R/S, notée Q_T par :*

$$Q_T = \frac{\max_{1 \leq k \leq T} \sum_{j=1}^k (X_j - \bar{X}_T) - \min_{1 \leq k \leq T} \sum_{j=1}^k (X_j - \bar{X}_T)}{S_T}$$

où

$$\begin{aligned}\bar{X}_T &= \frac{1}{T} \sum_{t=1}^T X_t \\ S_T^2 &= \frac{1}{T} \sum_{t=1}^T (X_t - \bar{X}_T)^2\end{aligned}$$

Proposition 12

$$\exists H \in]0, 1[, Q_T = O(T^H) \text{ quand } T \rightarrow +\infty$$

L'exposant de Hurst est ainsi donné par

$$H \sim \frac{\ln Q_T}{\ln T}$$

Série	Estimateur \hat{H}	Écart-type de l'estimateur $s(\hat{H})$	Intervalle de confiance à 95% pour d
R_t	0.5314	0.0172	[-0.0024, 0.0652]
R_t^2	0.8354	0.0133	[0.3093, 0.3614]
$ R_t $	0.8617	0.0148	[0.3326, 0.3907]
$ R_t ^{1.65}$	0.8471	0.0138	[0.3200, 0.3741]

TAB. 2.1 – *Exposants de Hurst pour des transformations non linéaires des rentabilités du CAC*

2.6.3 Exposant de Hurst et paramètre d'intégration fractionnaire

On peut montrer la proposition suivante :

Proposition 13 *Pour un modèle ARFIMA(p, d, q), on a la relation :*

$$d = H - \frac{1}{2}$$

- si $\frac{1}{2} < H < 1$, le processus ARFIMA est stationnaire à mémoire longue. Les autocorrélations sont positives et décroissent lentement vers 0 quand le retard augmente.
- si $H = \frac{1}{2}$, le processus se réduit à un processus ARMA et ne présente pas de structure de dépendance à long terme.
- si $0 < H < \frac{1}{2}$, le processus est dit anti-persistant. Les autocorrélations alternent de signe.

La librairie *fSeries* de R propose 9 méthodes pour estimer l'exposant de Hurst. Appliquées aux rentabilités journalières du CAC 40, aucune de ces méthodes ne conduit à une valeur estimée significativement (au seuil de 5%) différente de $\frac{1}{2}$. Le processus des rentabilités boursières n'est donc pas à mémoire longue (ce qui est, comme nous l'avons déjà vu, conforme à l'hypothèse d'efficience).

En fait, le phénomène de mémoire “longue” se produit sur des transformations non linéaires des rentabilités. Tout d'abord, on effectue l'étude de l'autocorrélation d'ordre 1 entre les puissance non entière des rentabilités. Appliquée à celles du CAC 40, cela donne le graphique FIG. (2.6). Le maximum de $\rho(|R_t|^\alpha, |R_{t-1}|^\alpha)$ est atteint pour $\alpha = 1.65$, vaut 0.1785 (et est significativement non nul au niveau 0.00...%). Ce phénomène est “universel”, pour les grands indices internationaux (les titres individuels, les portefeuilles), on constate d'une part, que les puissances non entières de deux rentabilités journalières consécutives sont significativement positivement corrélées, et que le maximum est atteint pour un puissance α en général comprise entre 1 et 2.

On étudie ensuite la fonction d'autocorrélation du processus R_t^α . Pour l'indice CAC 40, on obtient l'autocorrélogramme FIG. (2.7). Celui-ci est caractéristique des séries ARFIMA stationnaires à mémoire longue $d \in]0, \frac{1}{2}[$. Ceci est confirmé lorsqu'on estime l'exposant de Hurst pour les séries $R_t^2, |R_t|, |R_t|^{1.65}$ (cf. TAB. (2.1)) : on obtient pour ces séries, des valeurs de d comprises entre 0,3 et 0,4.

2.7 Mémoire longue pour la volatilité

Ce que montre l'étude empirique précédente, c'est que les “chocs” n'ont pas de réels impacts sur l'espérance des rentabilités mais ils en ont sur la variance conditionnelle.

L'idée est de construire une classe de processus qui permet d'expliquer cette persistance des chocs sur la volatilité, de la même manière qu'un processus ARFIMA explique la persistance des chocs sur l'espérance.

On repart de la représentation ARMA d'un processus GARCH(*cf.* p. 54). On peut donc écrire :

$$\Phi(L) Y_t^2 = \alpha_0 + \Theta(L) u_t$$

On peut alors définir le modèle IGARCH(p, q) (modèle GARCH intégré : Bollerslev et Engle (1986)) comme on a défini le modèle ARIMA(p, 1, d) par rapport au modèle ARMA :

$$\Phi(L) (Id - L) Y_t^2 = \alpha_0 + \Theta(L) u_t$$

On définit ensuite le modèle GARCH fractionnaire intégré (modèle FIGARCH : Baillie, Bollerslev, et Mikkelsen (1996)), comme on est passé du modèle ARMA au modèle ARFIMA :

$$\Phi(L) (Id - L)^d Y_t^2 = \alpha_0 + \Theta(L) u_t$$

En ré-écrivant $u_t = Y_t^2 - \sigma_t^2$, il vient :

$$\sigma_t^2 = \omega_0 + \left(Id - \Theta(L)^{-1} \Phi(L) \Delta^d \right) Y_t^2$$

avec $\omega_0 = \frac{\alpha_0}{1 - \sum_{j=1}^q \beta_j}$. Le polynôme $\lambda(L) = Id - \Theta(L)^{-1} \Phi(L) \Delta^d$ s'écrivant :

$$\lambda(L) = \sum_{k=1}^{+\infty} \lambda_k L^k$$

la variance conditionnelle ne dépend effectivement que des valeurs du passé ($k > 0$).

2.7.1 Quelques résultats empiriques

On traite les rentabilités journalières du CAC 40.

modèle ARMA(1,1)-GARCH(1,1)

Coefficient(s) :

	Estimate	Std. Error	t value	Pr(> t)
K	4.362e-04	2.200e-04	1.983	0.0474 *
AR[1]	-2.813e-01	8.440e-03	-33.337	<2e-16 ***
MA[1]	-2.739e-01	8.416e-03	-32.543	<2e-16 ***
C	5.587e-05	2.378e-06	23.497	<2e-16 ***
Alpha[1]	4.474e-01	1.400e-04	3196.082	<2e-16 ***
Beta[1]	3.007e-01	1.285e-04	2339.344	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Fit:

sigma^2 estimated as: 0.0001682

sigma^2 of innovation estimated as: 0.9986
 Log Likelihood: 14038.52

Statistiques sur les résidus standardisés :

Variance	0.000168
Stdev	0.012969
Skewness	-0.136880
Kurtosis	5.954898

modèle ARMA(1,1)-APARCH(1,1)

Coefficient(s):

	Estimate	Std. Error	t value	Pr(> t)
K	4.598e-04	2.016e-04	2.28	0.0226 *
AR[1]	1.738e-01	1.273e-05	13651.23	<2e-16 ***
MA[1]	1.516e-01	1.272e-04	1191.49	<2e-16 ***
C	2.235e-05	1.079e-06	20.71	<2e-16 ***
Alpha[1]	1.023e-01	1.833e-03	55.79	<2e-16 ***
Beta[1]	3.460e-01	9.161e-04	377.72	<2e-16 ***
Gamma[1]	8.457e-01	6.418e-04	1317.76	<2e-16 ***
Delta	1.760e+00	9.953e-04	1768.24	<2e-16 ***

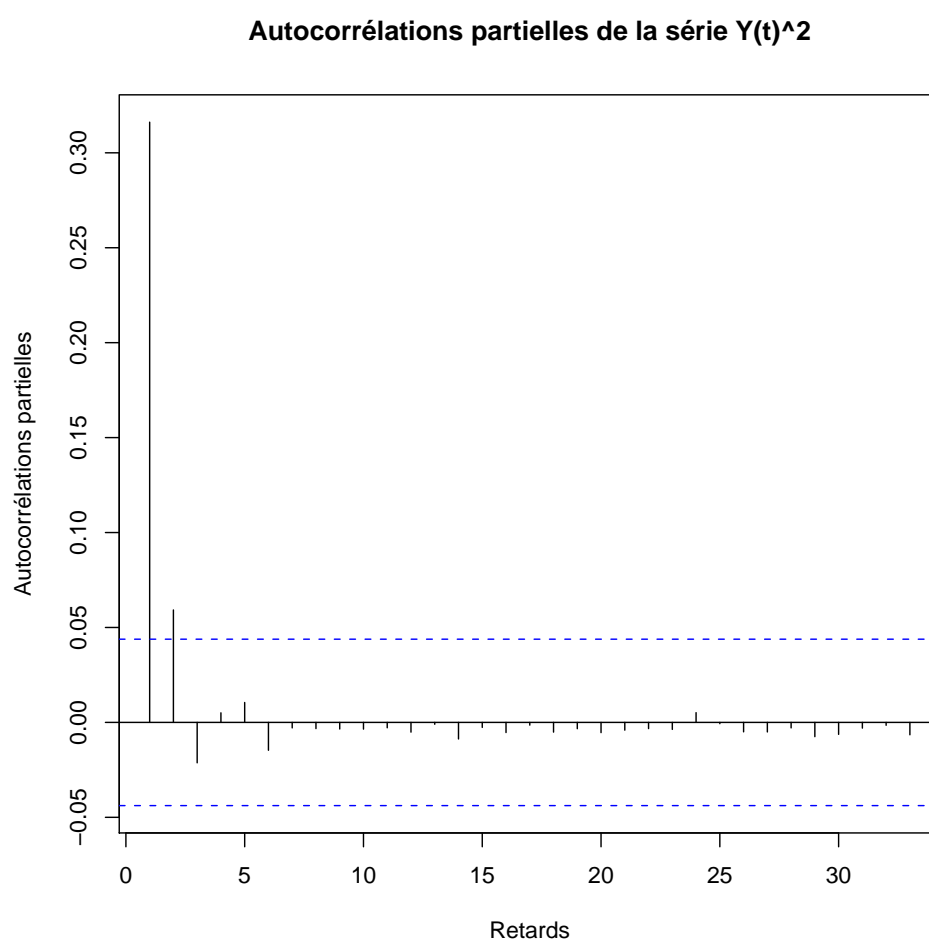
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

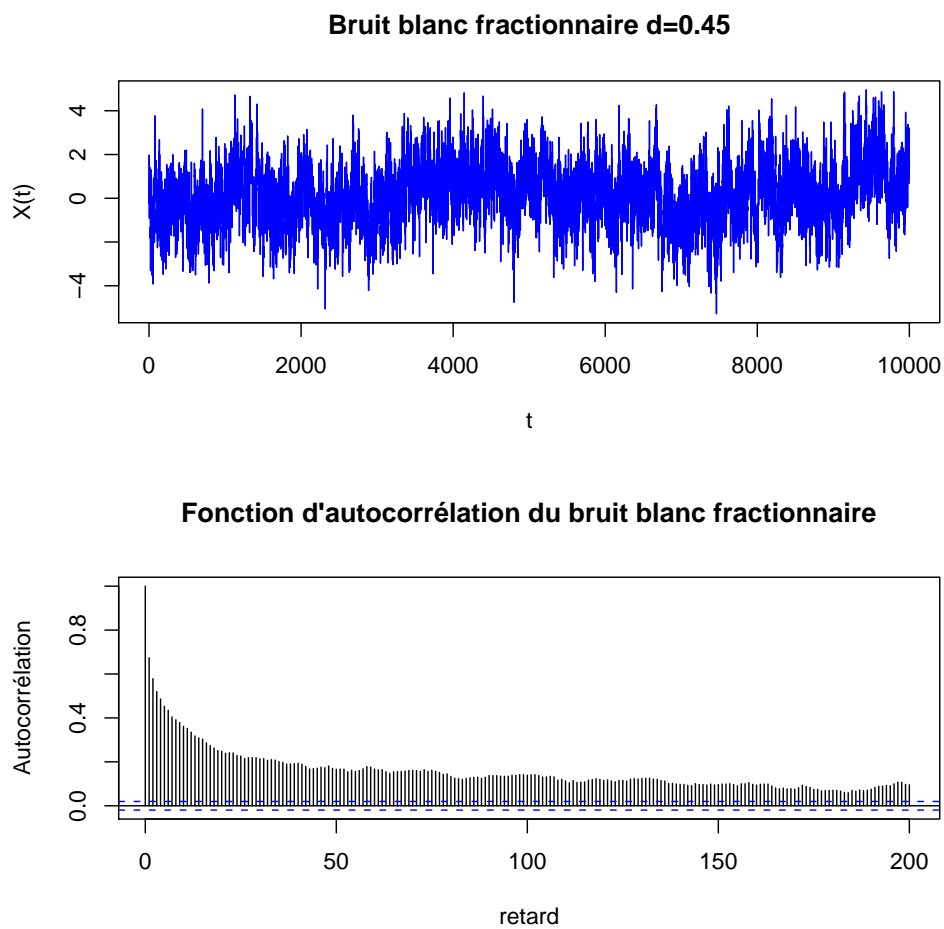
Fit:

sigma^2 estimated as: 0.0001681
 sigma^2 of innovation estimated as: 1.017
 Log Likelihood: 14287.56

Statistiques sur les résidus standardisés :

Variance	1.016778
Stdev	1.008354
Skewness	-0.381804
Kurtosis	5.709132

FIG. 2.4 – *Autocorrélations partielles de Y_t^2*

FIG. 2.5 – *Modèle ARFIMA(0, 0.45, 0)*

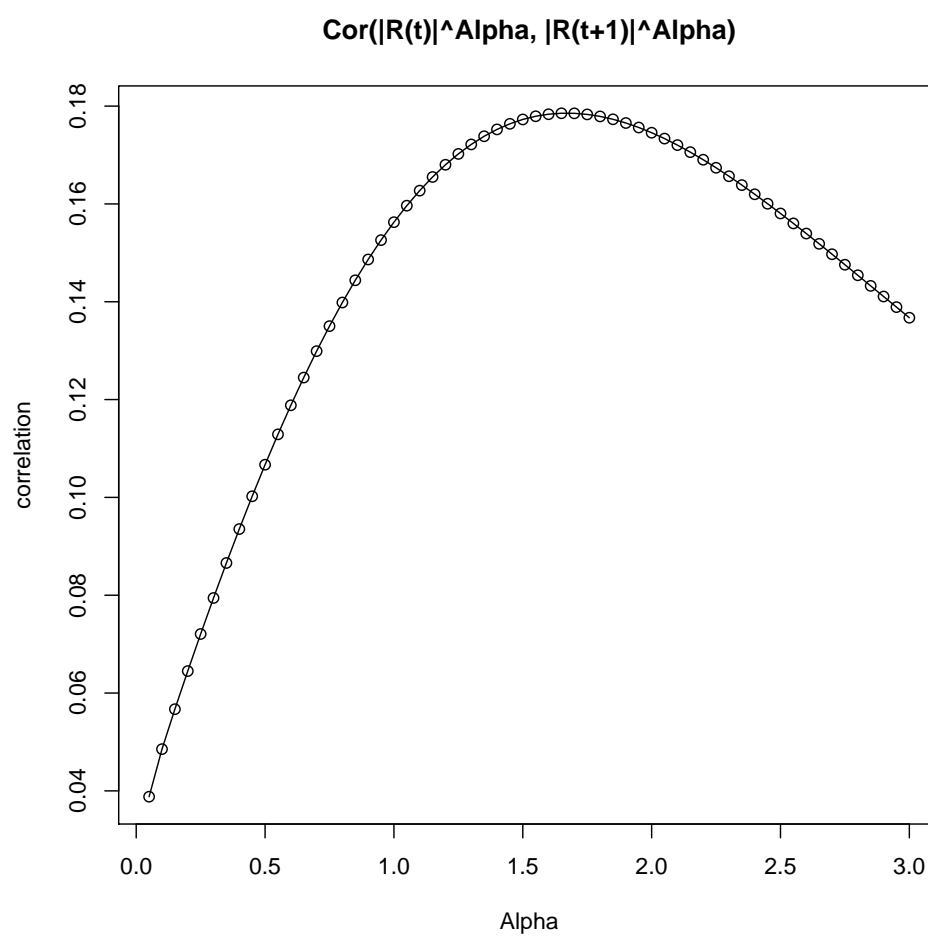


FIG. 2.6 – $\rho(|R_t|^\alpha, |R_{t-1}|^\alpha)$ en fonction de α

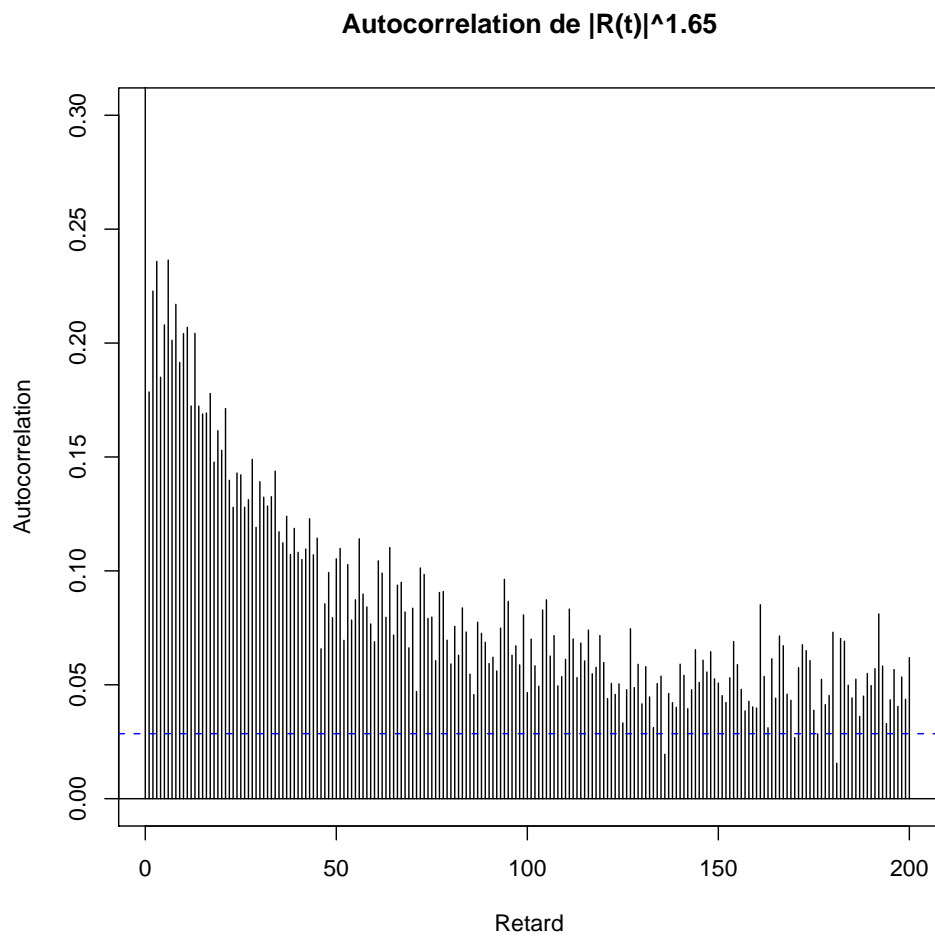
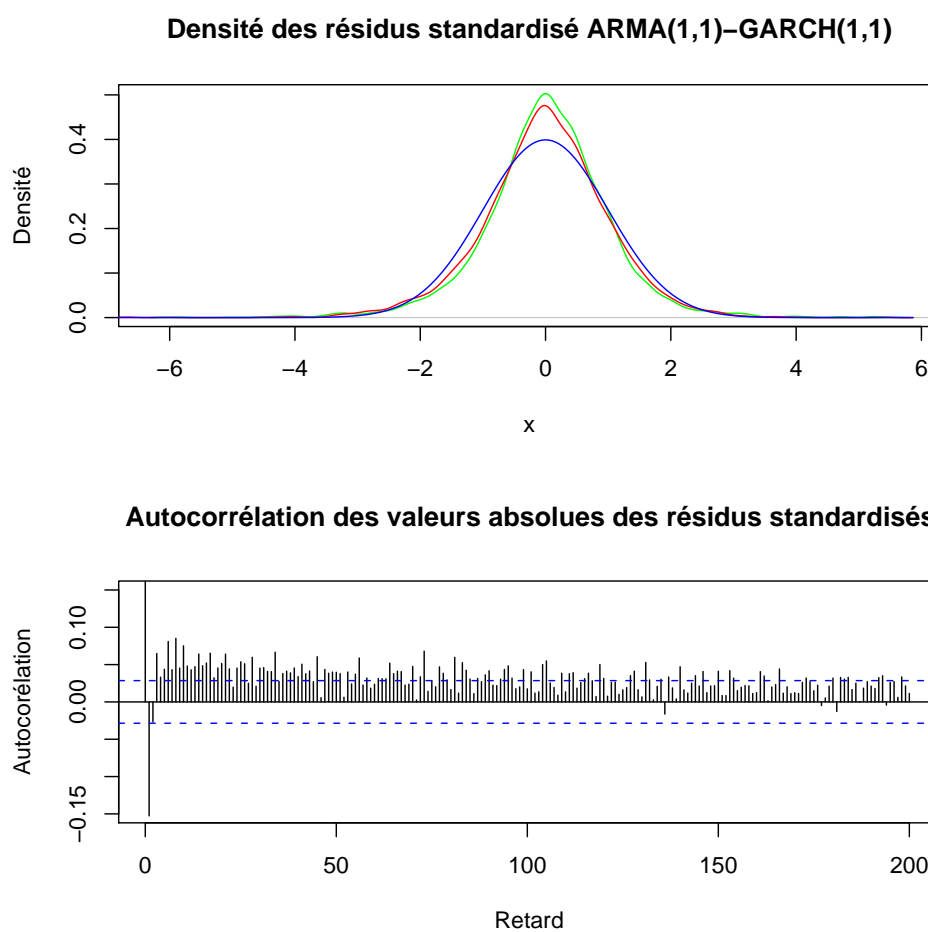


FIG. 2.7 – *Fonction d'autocorrélation de $|R_t|^{1.65}$*

FIG. 2.8 – *Résidus standardisés pour le CAC 40 modèle ARMA-GARCH*

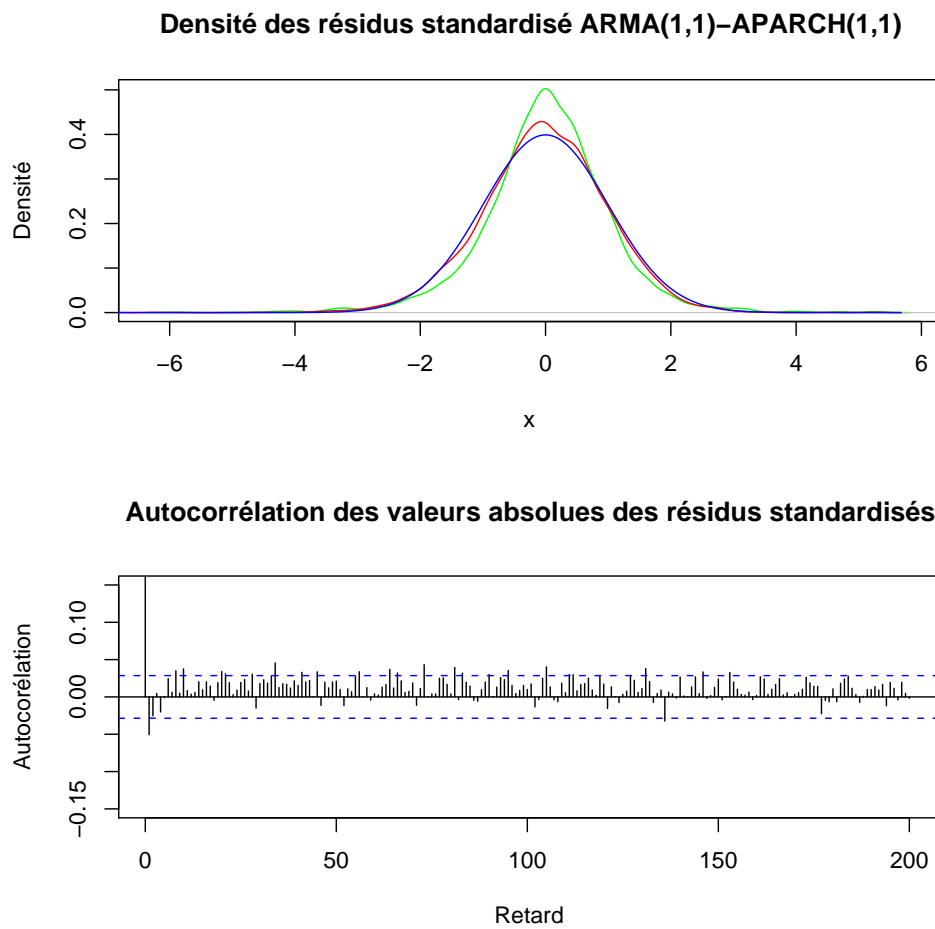


FIG. 2.9 – *Résidus standardisés pour le CAC 40 modèle ARMA-APARCH*

Références

- Baillie, R., Bollerslev, T., et Mikkelsen, H. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, Vol. 74, n° 1, p. 3–30.
- Bollerslev, T. (1986). Generalized autoregressive conditionnal heteroskedasticity. *Journal of Econometrics*, Vol. 31, p. 307–327.
- Bollerslev, T., Chou, R. Y., et Kroner, K. F. (1992). ARCH modeling in finance. *Journal of Econometrics*, Vol. 1, n° Annals, p. 5–59.
- Bollerslev, T. et Engle, R. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, n° 5, p. 1–50.
- Box, G. E. et Jenkins, G. M. (1976). *Time series analysis: forecasting and control*, Holden-Day.
- Dempster, A., Laird, N., et Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, Vol. B, n° 39, p. 1–38.
- Ding, Z., Granger, C., et Engle, R. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, Vol. 1, n° 1, p. 83–106.
- Engle, R. (1982). Autoregressive conditionnal heteroskedasticity. *Econometrica*, Vol. 50, p. 987–1008.
- Engle, R., Lilien, D., et Robbins, R. (1987). Estimating time varying risk premia in the term structure: The ARCH-M model. *Econometrica*, Vol. 55, p. 391–407.
- Hurst, H. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, n° 116, p. 770–799.
- Nelson, D. (1990). Conditionnal heteroskedasticity in asset returns: a new approach. *Econometrica*, Vol. 59, p. 347–370.