Chapitre 1
Propriétés et Faits Stylisés des Séries Financières
(2023-2024)

Économétrie Financière Avancée

Olivier DARNÉ

Propriétés et Faits Stylisés

Les séries de prix d'actifs financiers et de rentabilités présentent généralement un certain nombre de propriétés similaires suivant leur périodicité

Notations

- p_t : prix d'un actif financier (ou portefeuille) à une date t (cours de clôture)
- r_t : le logarithme de la rentabilité correspondante
- $r_t = \ln(p_t) - \ln(p_{t-1})$
- $r_t = \ln(1 + R_t)$ avec $R_t = \frac{p_t - p_{t-1}}{p_{t-1}}$

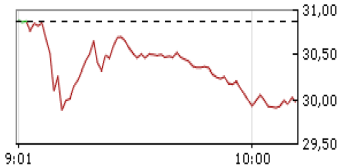
Propriétés et Faits Stylisés

Remarques

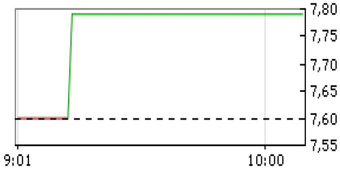
▶ r_t : rentabilités en temps continu (marché très liquide) et mesurent une variation dynamique du prix d'un actif entre le début et la fin de la période.
Dans l'intervalle de temps [t-1 ; t] il existe une infinité de points

▶ R_t : rentabilités en temps discret (marché peu liquide) et mesurent une variation statique du prix d'un actif entre le début et la fin de la période.
Dans l'intervalle de temps [t-1 ; t] rien ne se passe

Titre très liquide



Titre peu liquide



- Equivalence discret-continu :

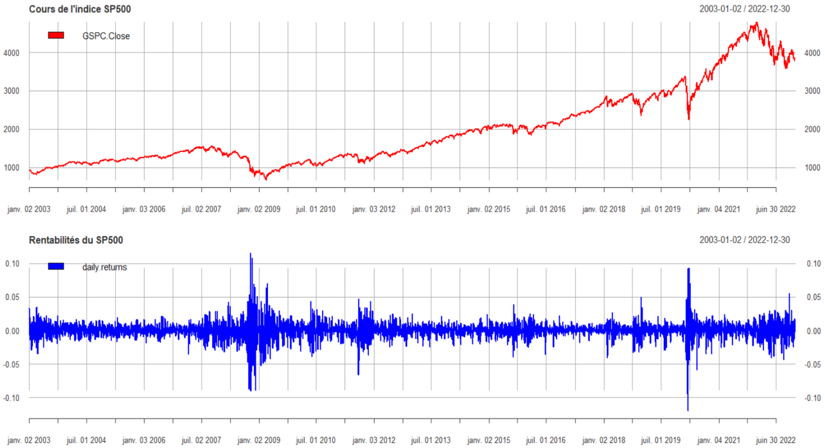
$$
\begin{aligned}
p_t &= p_{t-1}(1+R_t) \\
p_t &= p_{t-1}e^{r_t} \\
p_{t-1}(1+R_t) &= p_{t-1}e^{r_t} \\
r_t &= \ln(1+R_t)
\end{aligned}
$$

Par conséquent, si les rentabilités sont petites alors : $r_t \cong R_t$

En effet, lorsque la rentabilité discrète est proche de zéro, en utilisant un développement limité à l'ordre 1 de $\ln(1+R_t)$ au voisinage de 0, le taux continu tend vers le taux discret

Autrement dit : DL_1 de $\ln(1+x)$: $\ln(1+x) = x$

Considérons l'indice S&P500 observé en cours de clôture (ajusté des dividendes) ainsi que les rentabilités quotidiennes correspondates sur la période du 01/01/2003 au 01/01/2023



Charpentier (2002) distingue ainsi **8 principales propriétés** des actifs financiers

1. Stationnarité des rentabilités
2. Autocorrélation des carrés des rentabilités
3. Leptokurticité ou queues de distribution épaisses
4. Asymétrie perte/gain
5. Clusters de volatilité
6. Queues épaisses conditionnelles
7. Effet de levier
8. Saisonnalité

Propriété 1 : Stationnarité des rentabilités

Les processus stochastiques p_t associés aux prix d'actif sont généralement non stationnaires au sens de la stationnarité du second ordre, tandis que les processus associés aux rentabilités r_t sont compatibles avec la propriété de stationnarité au second ordre.

- $p_t \sim I(1)$
- $r_t \sim I(0)$

Rappel : Un processus $(x_t,\ t \in \mathbb{Z})$ est dit stationnaire au second ordre, ou stationnaire au sens faible, ou stationnaire d'ordre deux, si les trois conditions suivantes sont satisfaites

- $\forall t \in \mathbb{Z}, E(x_t^2) < \infty$
- $\forall t \in \mathbb{Z}, E(x_t) = m$, indépendant de t
- $\forall (t,h) \in \mathbb{Z}^2, cov(x_t, x_{t+h}) = E[(x_{t+h} - m)(x_t - m)] = \gamma(h)$, indépendant de t

- La première condition $E(x_t^2) < \infty$ garantit l'existence (ou la convergence) des moments d'ordre deux

- La seconde condition $E(x_t) = m$ porte sur le moment d'ordre un et signifie tout simplement que l'espérance du processus x_t doit être indépendante du temps

- La troisième condition porte sur les moments d'ordre deux résumés par la fonction d'autocovariance $\gamma(h)$: la fonction d'autocovariance du processus x_t doit être indépendante du temps et ne doit dépendre que l'ordre des retards h.

  Cela implique en particulier que la variance $\gamma(0)$ du processus x_t doit être constante au cours du temps.

Les tests de racine unitaire permettent de déterminer la stationnarité (stochastique) des prix et des rentabilités

- tests de racine unitaire de Dickey-Fuller augmenté (ADF) ou Philips-Perron
- test de stationnarité de KPSS

Propriété 2 : Autocorrélation des carrés des rentabilités

La série r_t^2 associée aux carrés des rentabilités présente généralement de fortes autocorrélations tandis que les autocorrélation de la série des rentabilités r_t sont souvent très faibles (hypothèse de bruit blanc)

- $cov(r_t^2, r_{t-k}^2) \neq 0$
- $cov(r_t, r_{t-k}) = 0$

Pour déterminer l'autocorrélation dans les rentabilités r_t ou les rentabilités au carré r_t^2 on peut utiliser les tests de Box et Pierce (1970) ou de Ljung et Box (1978)

Le **test de Ljung-Box** permet de tester l'hypothèse nulle qu'il n'y a pas d'autocorrélation jusqu'à l'ordre k du processus z_t (r_t ou r_t^2) :

$$\mathrm{H}_0 : \rho_1 = \rho_2 = \cdots = \rho_k$$

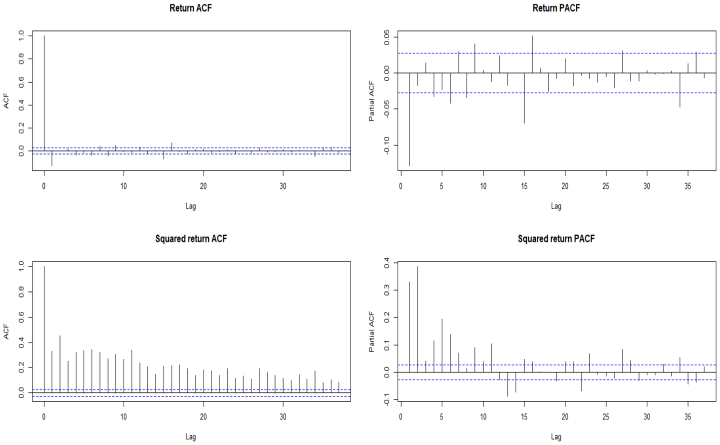
La statistique de test est :

$$Q_{LB}(k) = T(T+2)\sum_{j=1}^{k} \frac{\rho_j^2}{T-j} \sim \chi^2_{1-\alpha}(k)$$

où ρ_j est le coefficient d'autocorrélation d'ordre j, avec j=1,..., k.

$$\rho_k = \frac{\sum_{t=k+1}^{T}(z_t-\bar{z})(z_{t-k}-\bar{z})/(T-k)}{\sum_{t=k+1}^{T}(z_t-\bar{z})^2/T}$$

On vérifie sur nos données relatives à l'indice S&P500 l'absence de corrélations des rentabilités r_t ainsi que la présence de corrélations des rentabilités aux carrés r_t^2
(décroissance positive lente $\Rightarrow$ présence de persistence dans la volatilité)



Propriété 3 : Leptokurticité ou queues de distribution épaisses

L'hypothèse de Normalité des rentabilités est généralement rejetée. Les queues des distributions empiriques des rentabilités sont généralement plus épaisses que celles d'une loi Gaussienne. On parle alors de distribution leptokurtique.

La **Kurtosis** (ou coefficient d'aplatissement) est la mesure standard de l'épaisseur des queues de distributions fondée sur le moment centré d'ordre 4 (μ_4) :

$$\textbf{Kurtosis} : \text{Kur} = \frac{T^{-1}\sum_{i=1}^{T}(y_i-\bar{y})^4}{\left[T^{-1}\sum_{i=1}^{T}(y_i-\bar{y})^2\right]^2} = \frac{\mu_4}{\mu_2^2} \sim \mathcal{N}\left(3, \frac{24}{T}\right)$$

$$\textbf{Excès de Kurtosis} : \text{Excess} \;\; \text{Kur} = \frac{\mu_4}{\mu_2^2} - 3 = Kur - 3 \sim \mathcal{N}\left(0, \frac{24}{T}\right)$$

Le principe du **test de la Kurtosis** est le suivant :

- H_0 : $Kur = 3$ ou $Excess\ Kur = 0 \Rightarrow$ distribution Normale
- H_1 : $Kur \neq 3$ ou $Excess\ Kur \neq 0 \Rightarrow$ distribution non Normale

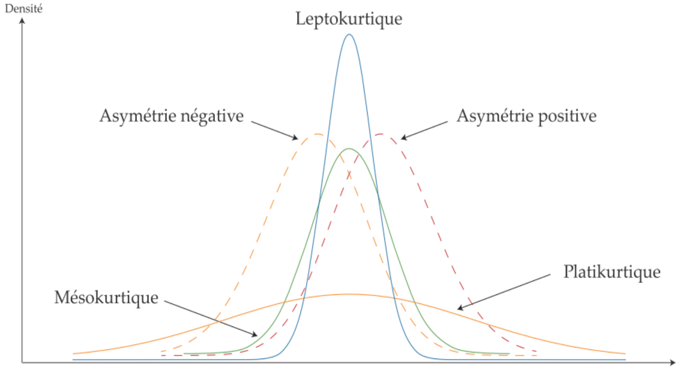
La statistique de la Kurtosis est donnée par

$$\nu_2 = \frac{Kur - 3}{\sqrt{24/T}} = \frac{Excess\ Kur}{\sqrt{24/T}} \sim \mathcal{N}\left(0,1\right)$$

RDD :

- si $|\nu_2| < VC_{\alpha\%}$, alors H_0 est acceptée au seuil de $\alpha\%$ et donc la distribution est Normale ou dite mésokurtique
- si $|\nu_2| > VC_{\alpha\%}$, alors H_0 est rejetée au seuil de $\alpha\%$ (risque d'erreur) et donc la distribution est non Normale.
  - si $\nu_2 < 0$ alors la distribution est platikurtique $\Rightarrow$ distribution plus aplatie et avec des queues de distribution étroites (*thin tail*)
  - si $\nu_2 > 0$ alors la distribution est leptokurtique $\Rightarrow$ distribution plus élevée et avec des queues de distribution épaisses (*fat tail* ou *heavy tail*)

Propriétés et Faits Stylisés



Propriété 4 : Asymétrie ou dissymétrie des rentabilités (pertes/gains)

La distribution des cours est généralement asymétrique : il y a plus de mouvements forts à la baisse qu'à la hausse.

Le **Skewness** (ou coefficient d'asymétrie) est une mesure de l'asymétrie de la distribution de la série autour de sa moyenne :

$$\mathrm{Skew} = \left(\frac{\left[T^{-1} \sum_{i=1}^{T} (y_i - \bar{y})^3 \right]^2}{\left[T^{-1} \sum_{i=1}^{T} (y_i - \bar{y})^2 \right]^3} \right)^{1/2} = \left(\frac{\mu_3^2}{\mu_2^3} \right)^{1/2} = \frac{\mu_3}{\sigma(x)^3} \sim \mathcal{N}\left(0, \frac{6}{T}\right)$$

Le principe du **test du Skewness** est le suivant :

- H_0 : $Skew = 0 \Rightarrow$ distribution symétrique
- H_1 : $Skew \neq 0 \Rightarrow$ distribution asymétrique

La statistique du Skewness est donnée par

$$\nu_1 = \frac{Skew - 0}{\sqrt{6/T}} \sim \mathcal{N}\left(0,1\right)$$

RDD :

- si $|\nu_1| < VC_{\alpha\%}$, alors H_0 est acceptée au seuil de $\alpha\%$ et donc la distribution est symétrique $\Rightarrow$ le titre a autant de pertes que de gains sur la période étudiée (si la moyenne des rentabilités est proche de zéro)
- si $|\nu_1| > VC_{\alpha\%}$, alors H_0 est rejetée au seuil de $\alpha\%$ (risque d'erreur) et donc la distribution est asymétrique :
  - si $\nu_1 < 0$ alors asymétrie négative $\Rightarrow$ les rentabilités négatives (pertes) sont plus importantes que les rentabilités positives (gains)
  - si $\nu_1 > 0$ alors asymétrie positive $\Rightarrow$ les rentabilités positives (gains) sont plus importantes que les rentabilités négatives (pertes)

Le **test de Jarque-Bera** ou test de Normalité du Chi-deux (χ^2) mesure la différence entre le Skewness et le Kurtosis de la série et celles d'une distribution Normale.

Le principe du test de Jarque-Bera est le suivant :

- H_0 : $Skew = 0$ et $Kur = 3 \Rightarrow$ distribution Normale
- H_1 : $Skew \neq 0$ ou $Kur \neq 3 \Rightarrow$ distribution non Normale

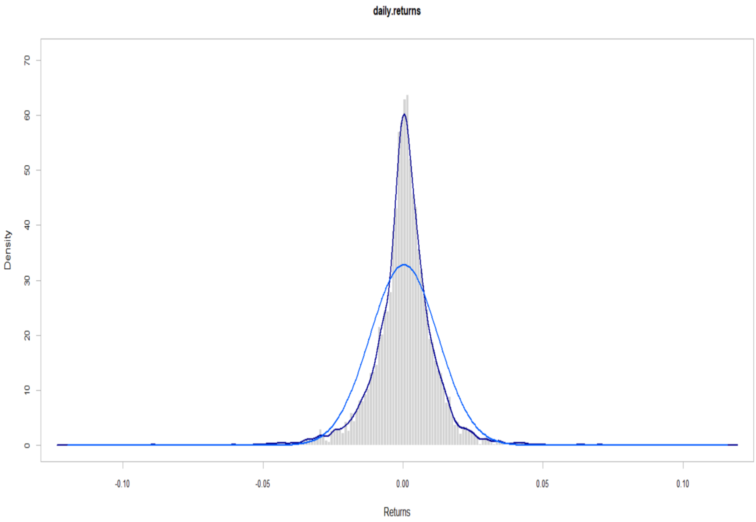
▶ La statistique de test est un multiplicateur de Lagrange (test de nullité jointe)

$$\mathrm{JB} = \frac{T-k}{6}\mathrm{Skew}^2 + \frac{T}{24}(\mathrm{Kur}-3)^2 \sim \chi^2_{1-\alpha}(2)$$

où k représente le nombre de coefficients estimés du modèle.

RDD :

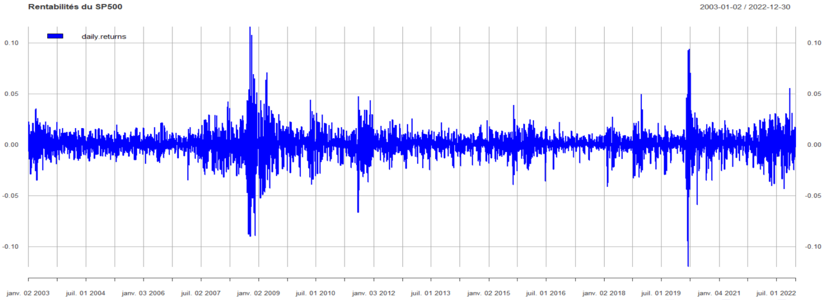
- si $JB > \chi^2_{0.95}(2)$ ou si la p-value < 0.05 alors on rejette H_0 de Normalité des résidus au seuil de 5%.
- si $JB < \chi^2_{0.95}(2)$ ou si la p-value > 0.05 alors on accepte H_0 de Normalité des résidus au seuil de 5%.


daily.returns
Density
0
10
20
30
40
50
60
70
-0.10
-0.05
0.00
0.05
0.10
Returns

Propriété 5 : Clusters de Volatilité

On observe empiriquement que de fortes variations des rentabilités sont généralement suivies de fortes variations. On assiste ainsi à un regroupement des extrêmes en cluster ou paquets de volatilités.

Cette propriété remet en cause l'hypothèse d'homoscédasticité généralement adoptée en économétrie linéaire $\Rightarrow$ hétéroscédasticité (voir Annexe)



Le **test LM ARCH** ou **test du multiplicateur de Lagrange** (LM) de Engle (1982) renvoie au concept d'une hétéroscédasticité conditionnelle, cad l'idée selon laquelle la variance en t dépend de ses valeurs passées (modèles ARCH).

Ce test est basé sur la régression des résidus au carré sur les résidus au carré décalés jusqu'à l'ordre q :

$$e_t^2 = \beta_0 + \sum_{i=1}^{q} \beta_i e_{t-i}^2 + u_t$$

où e_t sont les résidus du modèle.

Le principe du **test LM-ARCH(q)** est le suivant :

- H_0 : $\beta_1 = \beta_2 = \cdots = \beta_q = 0 \Rightarrow$ homoscédasticité
- H_1 : $\beta_i \neq 0$ $(0 < i \leq q) \Rightarrow$ hétéroscédasticité conditionnelle

On calcule la statistique LM-ARCH(q) $= TR^2$ où T est le nombre d'observations de e_t et R^2 est le coefficient de détermination de la régression précédente.

Sous l'hypothèse nulle d'homoscédasticité, la statistique est

$$\text{LM-ARCH}(q) = TR^2 \sim \chi^2(q)$$

RDD

- si LM-ARCH(q) $< \chi^2(q)$ ou si la p-value > 0.05 alors H_0 est acceptée au seuil de 5% $\Rightarrow$ homoscédasticité
- si LM-ARCH(q) $< \chi^2(q)$ou si la p-value < 0.05 alors H_0 est rejetée au seuil de 5% $\Rightarrow$ hétéroscédasticité conditionnelle

Remarques

- Le test de Ljung-Box sur les résidus au carré permet aussi de tester l'hétéroscédasticité conditionnelle (McLeod et Li, 1983)
- Tse (2001) propose aussi un test pour l'hétéroscédasticité conditionnelle : $T(m) \sim \chi^2(m)$.

Propriété 6 : Queues épaisses conditionnelles

Même une fois corrigée du phénomène de volatilité clustering (par exemple avec des modèles ARCH / GARCH), la distribution des résidus demeure leptokurtique même si la kurtosis est plus faible que dans le cas non conditionnel

Propriété 7 : Effet de levier

Il existe une asymétrie entre l'effet des valeurs passées négatives et l'effet des valeurs passées positives sur la volatité des cours ou de rentabilités.

Les baisses de cours tendent à engendrer une augmentation de la volatilité supérieure à celle induite par une hausse des cours de même ampleur.

Cette propriété n'est pas à confondre avec celle d'asymétrie de la distribution des cours ou des rentabilités (propriété 4).

Il s'agit ici d'une asymétrie de la relation liant les valeurs passés des cours ou rentabilités à la volatilité de ces derniers.

Effet de l'asymétrie : les rentabilités passées négatives des actifs financiers augmentent plus fortement la volatilité que les rentabilités passées positives
$\Rightarrow$ les rentabilités sont négativement corrélées avec les variations de leur volatilité.

Plusieurs explications de ce phénomène sont avancées :

- Effet de levier (*leverage effect*) : la baisse du prix de l'action (mauvaise nouvelle) d'une entreprise endettée aggrave son ratio de solvabilité (ratio dette / capitaux propres) $\Rightarrow$ incertitude sur l'avenir de l'entreprise lié à ses revenus futurs $\Rightarrow$ augmente le risque spécifique (idiosynchratique) $\Rightarrow$ augmente la volatilité de l'action (Black, 1976 ; Christie, 1982)

- Effet rétroactif (*feedback effect*) : une volatilité en hausse incite les investisseurs à exiger une prime de risque excédentaire pour rémunérer davantage les actifs qui deviennent plus risqués $\Rightarrow$ une hausse du taux de rendement exigé $\Rightarrow$ un repli immédiat des cours (Campbell et Hentschel, 1992 ; Bekaert et Wu, 2000)

Propriété 8 : Saisonnalité

Les rentabilités présentent de nombreux phénomènes de saisonnalité (effets week-end, effet janvier etc), aussi appelé anomalies boursières.

Effet janvier : les rentabilités boursières sont en moyenne plus élevées en janvier, notamment pour les sociétés à plus faible capitalisation boursière

- de nombreux investisseurs modifient la composition de leurs portefeuilles en fin d'année pour des raisons fiscales
- de nombreux de gestionnaires remanient leur portefeuille en fin d'année, se "débarrassant" des valeurs dont les perspectives sont jugées peu intéressantes $\Rightarrow$ baisse des cours en décembre, le rattrapage intervenant le mois suivant, en janvier.

Effet week-end ou effet lundi : les rentabilités boursières du lundi sont significativement plus faibles que lors des autres jours

- publication de mauvaises nouvelles d'une société le week-end après la clôture du vendredi $\Rightarrow$ réaction (et non sur-réaction) à l'ouverture le lundi
- la réorganisation des portefeuilles institutionnels en début de semaine
- le "blues" du lundi

$$r_t = \beta_1 + \beta_2 D_{2,t} + \beta_3 D_{3,t} + \beta_4 D_{4,t} + \beta_5 D_{5,t} + \varepsilon_t$$

- $D_{2,t}, D_{3,t}, D_{4,t}, D_{5,t}$: variables dichotomiques (valeurs 0 ou 1) pour les mardi, mercredi, jeudi et vendredi
- β_1 : rentabilité attendue pour le lundi
- éviter le *dummy variable trap* : multicolinéarité $\Rightarrow$ biase l'estimation MCO
- $\beta_2, \ldots, \beta_5$: différence entre la rentabilité attendue le lundi et celle de chacun des autres jours de la semaine

Analyse exploratoire des données

Toute étude sur les rentabilités et de manière générale doit débuter par une analyse des statistiques descriptives basiques. R propose différents packages :

Le package **fBasics** avec la fonction `basicStats()`

```
library(fBasics)
basicStats(rGSPC)
```

Le package **PerformanceAnalytics** avec la fonction `table.Stats()`

```
library(PerformanceAnalytics)
table.Stats(rGSPC*100)
table.Distributions(rGSPC*100)
```

Le package **DescTools** permet de faire le test de normalité de Jarque-Bera avec la fonction `JarqueBeraTest()`

```
library(DescTools)
JarqueBeraTest(rGSPC, robust = FALSE, method = "chisq")
```

Le package **stats** permet de faire les tests d'autocorrélation de Box-Pierce et Ljung-Box avec la fonction `Box.test()`

```
Box.test(rGSPC, lag = 10, type = c("Box-Pierce", "Ljung-Box"), fitdf = 0)
Box.test(rGSPC, lag = 10, type = "Ljung-Box", fitdf = 0)
```

Le package **FinTS** permet de faire le LM ARCH d'Engle avec la fonction `ArchTest()`

```
library(FinTS)
library(e1071)
ArchTest(base2,lag=5)
ArchTest(base2,lag=10)
```

Ces propriétés sont difficiles, voir impossible, à reproduire à partir de modèles ARMA linéaires classiques

Le théorème central de l'analyse des séries temporelles : le théorème de Wold (1954) indique que tout processus faiblement stationnaire peut être réécrit sous la forme d'une moyenne mobile infinie de processus de type bruits blancs, cad sous la forme d'une combinaison linéaire d'une séquence de variables aléatoires non corrélées dans le temps

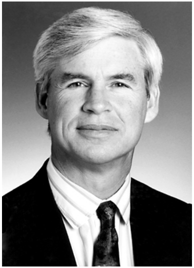
$\Rightarrow$ l'hypothèse de processus ARMA stationnaires ne permet pas de prendre en compte

- l'autocorrélation des rentabilités au carré
- les clusters de volatilités
- les queues de distribution épaisses

Approches ARCH-GARCH

Le premier article à proposer une modélisation ARCH-GARCH est celui de Robert Engle (1982) (prix Nobel 2003) :

Engle R.F. (1982), *AutoRegressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation*. Econometrica, 50, 987-1008. [**pdf**]



Objectif : prendre en compte les variances conditionnelles dépendant du temps
$\Rightarrow$ remise en cause de la propriété d'homoscédasticité

Analyse traditionnelle de la prévision des modèles linéaires (Box et Jenkins). Soit un processus AR(1) stationnaire :

$$X_t = \theta X_{t-1} + \varepsilon_t \qquad \text{avec } \varepsilon_t \sim i.i.d.\ N(0, \sigma_\varepsilon^2)$$

- $E(X_{t+1}) = 0$
- $E(X_{t+1}|\Omega_t) = \theta X_t$ avec $\Omega_t = X_t, X_{t-1}, \ldots$

Puisque seule l'espérance conditionnelle varie dans le temps, l'amélioration des prévisions issues de modèles de séries temporelles provient clairement de l'exploitation de l'information contenue dans l'espérance conditionnelle du processus

L'espérance conditionnelle $E(X_{t+1}|\Omega_t)$ est le prédicteur optimal au sens du critère MSE (*mean square error*)

L'idée d'Engle (1982) est de tenir compte des autres moments conditionnels de ce processus

Or, pour un processus AR(1) les moments d'ordre 2 (non conditionnel et conditionnel) sont constants, quelle que soit la date de la prévision

- $E(X_{t+1}^2) = \sigma_\varepsilon^2/(1-\theta^2)$
- $E(X_{t+1}^2|\Omega_t) = \sigma_\varepsilon^2$ avec $\Omega_t = X_t, X_{t-1}, \ldots$

Avec de tels modèles on est donc incapable de mesurer d'éventuels changements dans les variances des erreurs de prévision même si l'on souhaite que celles-ci soient affectées par l'évolution passée

Le principe général proposé par Engle (1982) consiste à supposer que la variance dépend de l'ensemble informationnel dont on dispose (évolue dans le temps)

$\Rightarrow$ Spécification ARCH(q) où le carré des perturbations ε_t^2 suit un processus autorégressif d'ordre q

- $V(X_{t+1}) = c$
- $V(X_{t+1}^2 | \Omega_t) = f(X_t, X_{t-1}, \ldots, X_1; \theta)$ avec θ un ensemble de paramètres

Les modèles ARCH sont donc des modèles autorégressifs conditionnellement hétéroscédastiques

Les outliers

La présence d'outliers ou *jumps* (*large shocks* ou *black swans*) est très importante dans les séries financières et peut être due à différents événements économiques, financiers, politiques, naturels . . .

Leur présence peut entraîner des effets indésirables en économétrie financière, comme, par exemple, sur :

- les tests de normalité, skewness et kurtosis
- les tests d'hétéroscédasticité conditionnelle
- l'estimation des paramètres des modèles de volatilité
- la prévision de la volatilité
- les mesures de risque

Il existe principalement 2 approches pour traiter des outliers :

(1) utiliser des estimateurs robustes à la présence d'outliers

(2) corriger les données des outliers et ensuite utiliser les estimateurs classiques

Pour corriger les données des outliers il existe plusieurs méthodes, principalement basées sur l'analyse d'intervention de Box et Tiao (1975), transposée aux modèles de volatilité de type GARCH

Boudt, Peterson et Croux (2008) proposent une approche qui ne nécessite par de modélisation de la volatilité.
Leur procédure se déroule en 3 étapes sur les rentabilités $r_t = \{r_1, \ldots, r_T\}$

Step 1: Classement des observations dans l'ordre de leur extrême

La distance de Mahalanobis au carré permet de mesurer le niveau extrême des rentabilités r_t

$$d_t^2 = (r_t - \mu)' \Sigma^{-1} (r_t - \mu)$$

avec μ et Σ les estimations de la moyenne et de la matrice de variance-covariance d'un sous-ensemble des données de taille $\lfloor (1-\alpha)T \rfloor$, avec $\lfloor . \rfloor$ la partie entière.

Ces estimations seront robustes contre les α plus extrêmes rentabilités

On obtient $d^2_{(1)}, \ldots, d^2_{(T)}$ la séquence ordonnée des distances de Mahalanobis au carré estimées tel que $d^2_{(i)} \leq d^2_{(i+1)}$

Distance de Mahalanobis : mesure de distance basée sur la corrélation entre des variables indépendantes (espace multidimensionnel)
La distance de Mahalanobis est souvent utilisée pour la détection de données aberrantes ou bien pour déterminer la cohérence (similitude) entre deux types de données
La distance euclidienne, correspond à la distance de Mahalanobis dans le cas où la matrice de covariance est la matrice identité, cad que les variables sont centrées réduites et indépendantes (les variables indépendantes ne son pas corrélées)

Step 2: Identification des outliers

Une observation est qualifiée d'outlier si

$$d_t^2 > d^2_{\lfloor (1-\alpha)T \rfloor} \qquad \text{et} \qquad d_t^2 > \chi^2_{n;0.999}$$

où

- $d^2_{\lfloor (1-\alpha)T \rfloor}$: le $(1-\alpha)$ quantile empirique
- $\chi^2_{n;0.999}$: le 99.9% quantile d'un chi-deux à n degré de liberté

Step 3: Correction des outliers

Les rentabilités $\tilde{r}_t$ identifiées comme outliers sont remplacées par

$$\tilde{r}_t = r_t \sqrt{\max\left(d^2_{\lfloor (1-\alpha)T \rfloor}; \chi^2_{n;0.999}\right) / d_t}$$

La fonction `Return.clean()` du package **PerformanceAnalytics** permet d'appliquer la méthode de Boudlt et alii.
Il faut également installer le package **robustbase**

```
library(PerformanceAnalytics)
library(robustbase)
crGSPC <- Return.clean(rGSPC, method = "boudt")
```

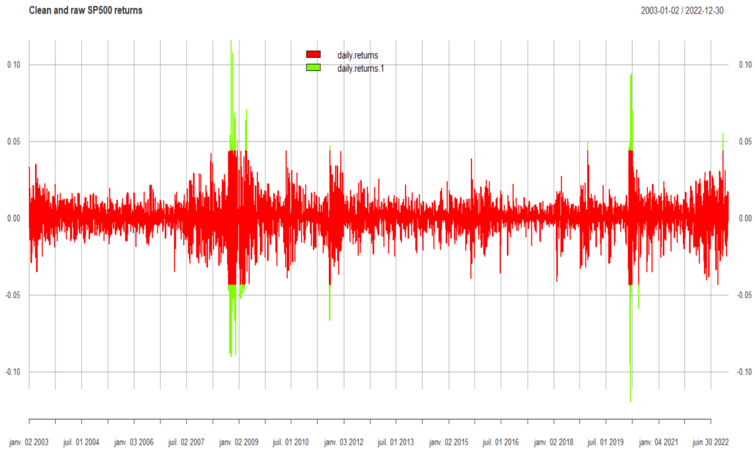
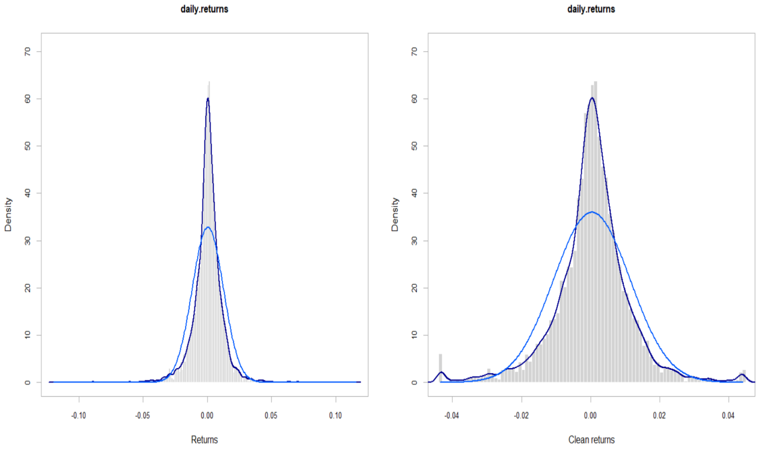


Table: Statistiques descriptives des rentabilités brutes et corrigées des outliers.

	Moyenne (%)	Mediane (%)	Min (%)	Max (%)	Ecart-type (%)	Skewness	v_1	Ex.Kurtosis	v_2
raw	0.0367	0.0702	-11.98	11.58	1.212	-0.245	-7.09	12.47	180.6
clean	0.0379	0.0702	-4.35	4.41	1.109	-0.241	-6.98	3.25	32.6

	JB	*p*-value	Q(10)	*p*-value	LM-ARCH(5)	*p*-value	LM-ARCH(10)	*p*-value
raw	32 683	0.000	127.15	0.000	1 411.4	0.000	1 529.9	0.000
clean	2 267	0.000	71.92	0.000	1 434.6	0.000	1 566.9	0.000


daily.returns
Density
0
10
20
30
40
50
60
70
-0.10
-0.05
0.00
0.05
0.10
Returns
daily.returns
Density
0
10
20
30
40
50
60
70
-0.04
-0.02
0.00
0.02
0.04
Clean returns

Python

Kevin Sheppard, dans son ouvrage *Introduction to Python for Econometrics, Statistics and Numerical Analysis*, propose diverses applications et programmes sous Python en Econométrie. [**pdf**]

`https://www.kevinsheppard.com/Python_for_Econometrics`