ECN 6578A, Économétrie des marchés financiers, Hiver 2022

Cours 5

William McCausland

2023-02-05

Plan

1. La fonction de vraisemblance : exemples bernoullien, poissonien
2. Maximisation de vraisemblance : exemples simples
3. Maximum de vraisemblance : propriétés (un peu de théorie)
4. Le modèle EGARCH
5. Estimation des modèles GARCH, quelques résultats

Éléments de l'analyse maximum de vraisemblance

- Quantités pertinentes :
  - θ, un vecteur de paramètres inconnus,
  - $y = (y_1, \ldots, y_T)$, un vecteur aléatoire des variables observables,
  - y°, le vecteur observé.
- Fonctions pertinentes :
  - $f(y|\theta)$, la densité conditionnelle des données (modèle),
  - $\mathcal{L}(\theta; y) = f(y|\theta)$, la vraisemblance,
  - $\mathcal{L}(\theta; y^\circ) = f(y^\circ|\theta)$, la vraisemblance réalisée.

Le modèle Bernoulli

- Supposons que les y_i sont iid Bernoulli avec probabilité $\theta \in [0,1]$:
$$f(y_i|\theta) = \begin{cases} \theta & y_i = 1 \\ (1-\theta) & y_i = 0 \end{cases}$$
$$= \theta^{y_i}(1-\theta)^{1-y_i}$$
- On observe $y = (y_1, \ldots, y_n)$; la fonction de masse de probabilité est

$$f(y|\theta) = \prod_{i=1}^{n} f(y_i|\theta) = \prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i} = \theta^{n_1}(1-\theta)^{n_0},$$

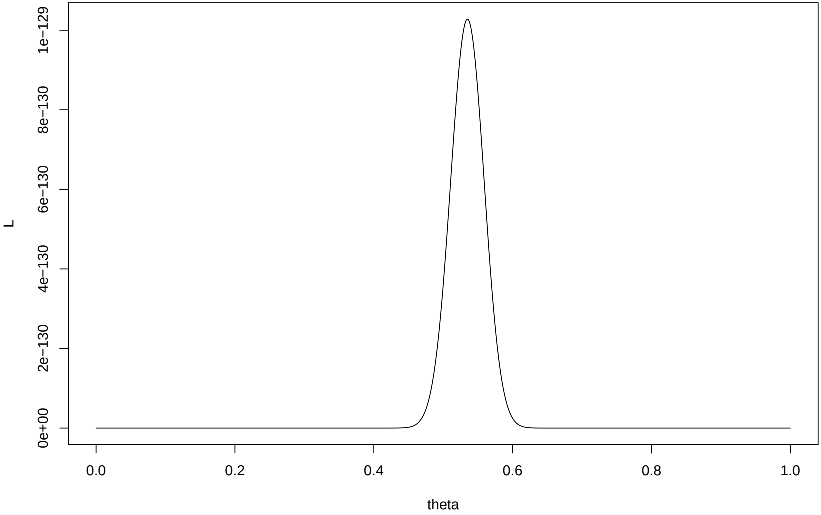
  où
  - $n_1 = \sum_{i=1}^{n} y_i$ est le nombre de fois qu'on observe 1, et
  - $n_0 = n - \sum_{i=1}^{n} y_i$ est le nombre de fois qu'on observe 0.

Deux intérpretations de la même expression

- Deux façons de dénoter la même expression :
  - Fonction de masse de probabilité $f(y|\theta) = \theta^{n_1}(1-\theta)^{n_0}$.
  - Fonction de vraisemblance $\mathcal{L}(\theta; y) = \theta^{n_1}(1-\theta)^{n_0}$.
- $f(y|\theta)$ donne, pour θ fixe, les probabilités relatives de plusieurs séquences $(y_1, \ldots, y_n)$.
- $\mathcal{L}(\theta; y)$ donne, pour y fixe (le vecteur des données observées) une note (ou évaluation) à chaque valeur θ pour la qualité de sa prévision des données observées.
- Soit $L(\theta; y) = \log \mathcal{L}(\theta; y)$, la log-vraisemblance.

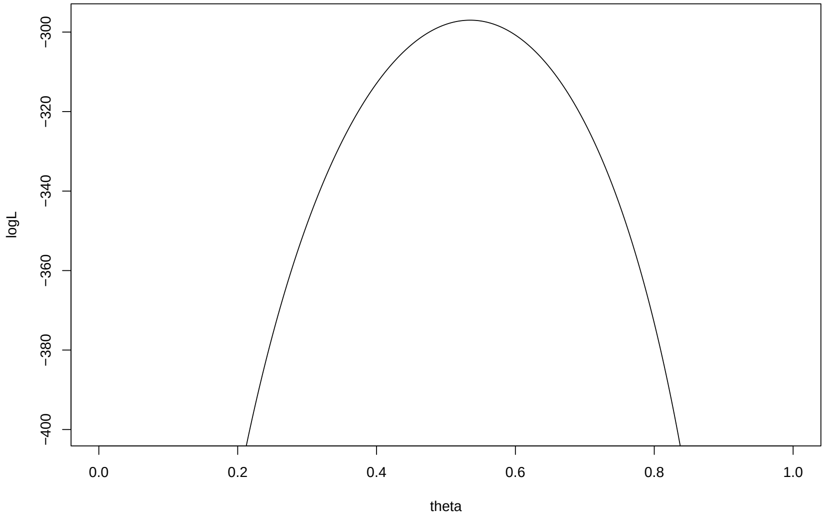
La vraisemblance Bernoulli pour $n_0 = 200$, $n_1 = 230$

```
n_0 = 200; n_1 = 230; theta = seq(0, 1, by=0.001)
L = theta^n_1 * (1-theta)^n_0
plot(theta, L, type='l')
```



La log vraisemblance Bernoulli pour $n_0 = 200$, $n_1 = 230$

```
logL = n_1 * log(theta) + n_0 * log(1-theta)
plot(theta, logL, type='l', ylim=c(-400, max(logL)))
```



Le modèle poissonien

- Supposez que les y_i sont iid Poisson avec moyenne $\theta > 0$.
- La fonction de masse de probabilité de y_i est

$$f(y_i|\theta) = e^{-\theta} \frac{\theta^{y_i}}{y_i!}.$$

- On observe le vecteur aléatoire $y = (y_1, \ldots, y_n)$; la fonction de masse de probabilité de y est

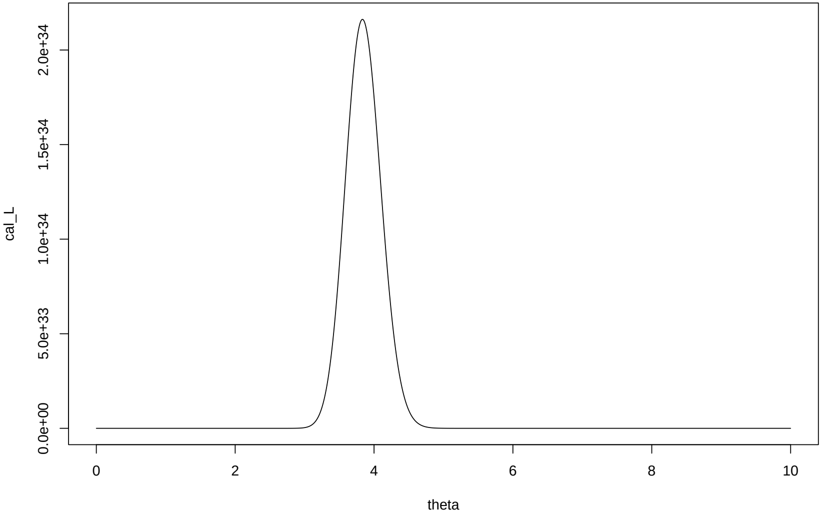
$$f(y|\theta) = \prod_{i=1}^{n} f(y_i|\theta) = \prod_{i=1}^{n} e^{-\theta} \frac{\theta^{y_i}}{y_i!} = \left[\prod_{i=1}^{n} \frac{1}{y_i!}\right] e^{-n\theta} \theta^{\sum_{i=1}^{n} y_i}.$$

- Pour simplifier un facteur qui importe peu,

$$c \equiv \left[\prod_{i=1}^{n} \frac{1}{y_i!}\right].$$

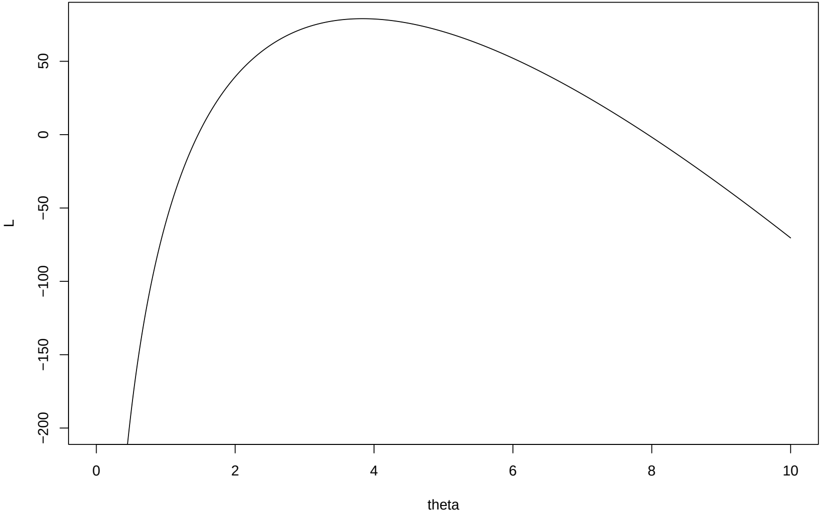
Vraisemblance poissonienne pour $n = 60$, $\sum_{i=1}^{n} y_i = 230$

```
n = 60; somme_y = 230; theta = seq(0, 10, by=0.001)
cal_L = exp(-n*theta) * theta^somme_y
plot(theta, cal_L, type='l')
```



Log vraisemblance poissonienne, $n = 60$, $\sum_{i=1}^{n} y_i = 230$

```r
L = -n*theta + somme_y*log(theta)
plot(theta, L, type='l', ylim=c(-200, max(L)))
```



La fonction de vraisemblance pour une séries chronologique

- La vraisemblance en général pour un modèle qui donne la densité $f(r_1, \ldots, r_T, \theta)$.

$$\mathcal{L}(\theta; r) = f(r_1|\theta) f(r_2|r_1, \theta) \cdots f(r_T|r_1, \ldots, r_{T-1}, \theta)$$

- Chaque densité $f(r_t|r_1, \ldots, r_{t-1}, \theta)$ est un genre de prévision conditionnelle de r_t sachant $r_1, \ldots, r_{t-1}$ et θ.
- La log vraisemblance est

$$L(\theta; r) = \sum_{t=1}^{T} \log f(r_t|r_1, \ldots, r_{t-1}, \theta).$$

- Pourquoi la log-vraisemblance et non juste la vraisemblance?
  - Pas de dépassement ou soupassement numérique (overflow/underflow)
  - Plus facile à maximiser (la dérivée d'une somme est la somme des dérivées, la log-vraisemblance est plus souvent concave)

Exemple : évaluation de la log vraisemblance GARCH(1,1)

- Rappel : $L(\theta; r) = \sum_{t=1}^{T} f(r_t | r_1, \ldots, r_{t-1}, \theta)$.
- Juste avant l'itération t, la valeur σ_t^2 est disponible.
- À l'itération t,

1. On calcule le terme $\log f(r_t | r_1, \ldots, r_{t-1}, \theta)$ de la log vraisemblance. Dans le cas gaussien,

   $$\log f(r_t | r_1, \ldots, r_{t-1}, \theta) = -\tfrac{1}{2}(\log 2\pi + \log \sigma_t^2) - \tfrac{1}{2} r_t^2 / \sigma_t^2.$$

   et plus en général,

   $$\log f(r_t | r_1, \ldots, r_{t-1}, \theta) = -\log \sigma_t + \log f_\epsilon(r_t / \sigma_t | \theta),$$

   où $f_\epsilon(\epsilon | \theta)$ et la densité des ϵ_t.
2. On calcule la valeur σ_{t+1}^2 :

   $$\sigma_{t+1}^2 = \alpha_0 + \alpha_1 r_t^2 + \beta_1 \sigma_t^2.$$

Maximum de la vraisemblance Bernoulli

- Vraisemblance : $\mathcal{L}(\theta; y) = \theta^{n_1}(1-\theta)^{n_0}$.
- Log vraisemblance : $L(\theta; y) = n_1 \log(\theta) + n_0 \log(1-\theta)$
- Deux dérivées de la log vraisemblance :

$$\frac{\partial L(\theta; y)}{\partial \theta} = \frac{n_1}{\theta} - \frac{n_0}{1-\theta}$$

$$\frac{\partial^2 L(\theta; y)}{\partial \theta^2} = -\frac{n_1}{\theta^2} - \frac{n_0}{(1-\theta)^2} < 0.$$

- La valeur qui maximise la vraisemblance et la log-vraisemblance est

$$\hat{\theta} = \frac{n_1}{n_0 + n_1} = \frac{n_1}{n}.$$

Maximum de la vraisemblance poissonienne

- Vraisemblance : $\mathcal{L}(\theta; y) = ce^{-n\theta}\theta^{\sum_{i=1}^{n} y_i}$.
- Log vraisemblance : $L(\theta; y) = \log c - n\theta + \left(\sum_{i=1}^{n} y_i\right)\log\theta$.
- Deux dérivées de la log vraisemblance :

$$\frac{\partial L(\theta; y)}{\partial\theta} = -n + \frac{\sum_{i=1}^{n} y_i}{\theta}$$

$$\frac{\partial^2 L(\theta; y)}{\partial\theta^2} = -\frac{\sum_{i=1}^{n} y_i}{\theta^2} < 0.$$

- La valeur $\hat{\theta}$ (souvent vue comme une variable aléatoire) qui maximise la vraisemblance et la log-vraisemblance est

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

- Pour $n = 60$ et $\sum_{i=1}^{n} y_i = 230$, $\hat{\theta} = \frac{23}{6} \approx 3.833$.

Maximum de vraisemblance : conditions de régularité

- Définitions :
  - θ est le vecteur des paramètres ; Θ, l'ensemble de toutes les valeurs possibles de θ.
  - r est le vecteur (aléatoire) des données.
- Conditions informelles de regularité :
  1. Le modèle est correct pour une valeur $\theta = \theta_0 \in \Theta$.
  2. La vraie valeur θ_0 est dans l'intérieur de Θ.
  3. Identification :

     $$\theta \neq \theta_0 \Rightarrow f(\cdot|\theta) \neq f(\cdot|\theta_0).$$

  4. $L(\theta; r) \equiv \log f(r|\theta)$ a toujours un maximum global unique.
  5. Le gradient de $L(\theta; r)$ est toujours borné.
  6. La matrice $\mathcal{I}(\theta)$ suivante (matrice d'information de Fisher) est définie positive:

     $$\mathcal{I}(\theta) = E_{r|\theta}\left[\frac{\partial L(\theta; r)}{\partial \theta^\top}\frac{\partial L(\theta; r)}{\partial \theta}\right].$$

Maximum de vraisemblance : résultats

Résultats : (Soit $\hat{\theta} \equiv \arg\max_\theta L(\theta; r)$, qui existe et est unique.)

1. $\hat{\theta} \to_p \theta_0$ (loi de grands nombres)
2. $\sqrt{n}(\hat{\theta} - \theta_0) \to_d N(0, \mathcal{I}(\theta_0)^{-1})$ (théorème central limite)
3. $E_{r|\theta}\left[\frac{\partial L(\theta;r)}{\partial\theta}\right] = 0$, alors $\mathcal{I}(\theta) = \mathrm{Var}_{r|\theta}\left[\frac{\partial L(\theta;r)}{\partial\theta}\right]$.
4. $\mathcal{I}(\theta) = E_{r|\theta}\left[-\frac{\partial^2 L(\theta;r)}{\partial\theta\partial\theta^\top}\right]$.

Problèmes restants :

1. Il faut trouver $\hat{\theta}$.
2. La variance asymptotique $\mathcal{I}(\theta_0)^{-1}$ dépend de θ_0, qui est inconnu.
3. L'espérance dans l'expression de $\mathcal{I}(\theta)$ est difficile à évaluer analytiquement.

Exemple Bernoulli

- Un cas rare où les calculs analytiques sont faisables.
- La moyenne du score :

$$E_{y|\theta}\left[\frac{\partial L}{\partial \theta}\right] = E_{y|\theta}\left[\frac{n_1}{\theta} - \frac{n_0}{(1-\theta)}\right] = \frac{n\theta}{\theta} - \frac{n(1-\theta)}{(1-\theta)} = 0$$

- La matrice d'information de Fisher :

$$\begin{aligned}\mathcal{I}(\theta) = E_{y|\theta}\left[-\frac{\partial^2 L}{\partial \theta^2}\right] &= E_{y|\theta}\left[\frac{n_1}{\theta^2} + \frac{n_0}{(1-\theta)^2}\right] \\ &= \frac{n\theta}{\theta^2} + \frac{n(1-\theta)}{(1-\theta)^2} = \frac{n}{\theta(1-\theta)}.\end{aligned}$$

- La variance de $\hat{\theta}$ (exacte, pas asymptotique) :

$$\mathrm{Var}[\hat{\theta}] = \mathrm{Var}\left[\frac{n_1}{n}\right] = \frac{1}{n^2} n \mathrm{Var}[y_i] = \frac{1}{n}(\theta - \theta^2) = \frac{\theta(1-\theta)}{n}.$$

Exemple poissonien

- Un autre cas rare où les calculs analytiques sont faisables.
- La matrice d'information de Fisher : ($E[y_i] = \theta$, $\mathrm{Var}[y_i] = \theta$)

$$\mathcal{I}(\theta) = E_{y|\theta}\left[-\frac{\partial^2 L}{\partial\theta^2}\right] = E_{y|\theta}\left[\frac{\sum_{i=1}^n y_i}{\theta^2}\right] = \frac{n\theta}{\theta^2} = \frac{n}{\theta}.$$

- La variance de $\hat{\theta}$ (exacte, pas asymptotique) :

$$\mathrm{Var}[\hat{\theta}] = \mathrm{Var}\left[\frac{\sum_{i=1}^n y_i}{n}\right] = \frac{1}{n^2} n\mathrm{Var}[y_i] = \frac{\theta}{n}.$$

- Pour $n = 60$ et $\sum_{i=1}^n y_i = 230$, $\mathrm{Var}[\hat{\theta}]$ est de $(0.2528)^2$ pour $\theta = \hat{\theta} \approx 3.833$, $(0.2236)^2$ pour $\theta = 3$ et $(0.2739)^2$ pour $\theta = 4.5$.

Comment trouver $\hat{\theta}$ I

- Gradient (score) et hessienne de la log-vraisemblance :

$$s(\theta) \equiv \frac{\partial L(\theta; r)}{\partial \theta^\top}, \quad H(\theta) \equiv \frac{\partial^2 L(\theta; r)}{\partial \theta \partial \theta^\top}.$$

- On utilise un processus séquentiel pour trouver $\hat{\theta}$: $\theta_1, \theta_2, \ldots,$
- Expansion quadratique de Taylor autour de θ_k :

$$\tilde{L}(\theta; r) = L(\theta_k; r) + s(\theta_k)^\top (\theta - \theta_k) + \frac{1}{2}(\theta - \theta_k)^\top H(\theta_k)(\theta - \theta_k).$$

- Le gradient $\tilde{s}(\theta)$ de $\tilde{L}(\theta; r)$:

$$\tilde{s}(\theta) = s(\theta_k) + H(\theta_k)(\theta - \theta_k)$$

- La condition $\tilde{s}(\theta_{k+1}) = 0$ définit la mise à jour θ_{k+1} de la méthode Newton :

$$\theta_{k+1} = \theta_k - H(\theta_k)^{-1} s(\theta_k).$$

Comment trouver $\hat{\theta}$ II

- Problème de non-convergence si la forme de la log vraisemblance est loin de quadratique et négative définie.

- Une recherche linéaire est plus robuste : choisir une valeur scalaire λ_k et calculer

$$\theta_{k+1} = \theta_k - \lambda_k H(\theta_k)^{-1} s(\theta_k).$$

  1. Calculez $s(\theta_k)$, $H(\theta_k)$.
  2. Trouvez une bonne valeur de λ_k (recherche linéaire)

- Des fois, on utilise souvent, au lieu de $H(\theta)$, une approximation

$$\hat{H}(\theta) = -\sum_{t=1}^{T} \frac{\partial \log f(r_t|r_1,\ldots,r_{t-1},\theta)}{\partial \theta^\top} \frac{\partial \log f(r_t|r_1,\ldots,r_{t-1},\theta)}{\partial \theta}.$$

- Une loi de grands nombres donne

$$\hat{H}(\theta_0) \to_p E[s(\theta_0)s(\theta_0)^\top] = \mathcal{I}(\theta_0) = -E[H(\theta_0)].$$

Approximation de $\mathcal{I}(\theta_0)$

- Rappelons que $\mathcal{I}(\theta_0)^{-1}$ est la variance asymptotique de l'estimateur MV.
- Cependant, θ_0 et $\mathcal{I}(\theta_0)$ sont inconnus.
- On utilise $-H(\hat{\theta})$ ou $-\hat{H}(\hat{\theta})$ au lieu de $\mathcal{I}(\theta_0)$, qui est inconnu.
- Heureusement, on a
  - Convergence de $\hat{\theta}$ à θ_0.
  - Convergence de $-\hat{H}(\theta_0)$ ou $-H(\theta_0)$ à $\mathcal{I}(\theta_0) = E[-H(\theta_0)]$.
  - Ensemble : convergence de $-H(\hat{\theta})$ ou $-\hat{H}(\hat{\theta})$ à $\mathcal{I}(\theta_0)$.

Le modèle EGARCH

Le modèle EGARCH(1,1) :

$$a_t = \sigma_t \epsilon_t \quad \ln \sigma_t^2 = \alpha \ln \sigma_{t-1}^2 + (1-\alpha)\alpha_0 + g(\epsilon_t) \quad \epsilon_t \sim \mathrm{iid}(0,1),$$

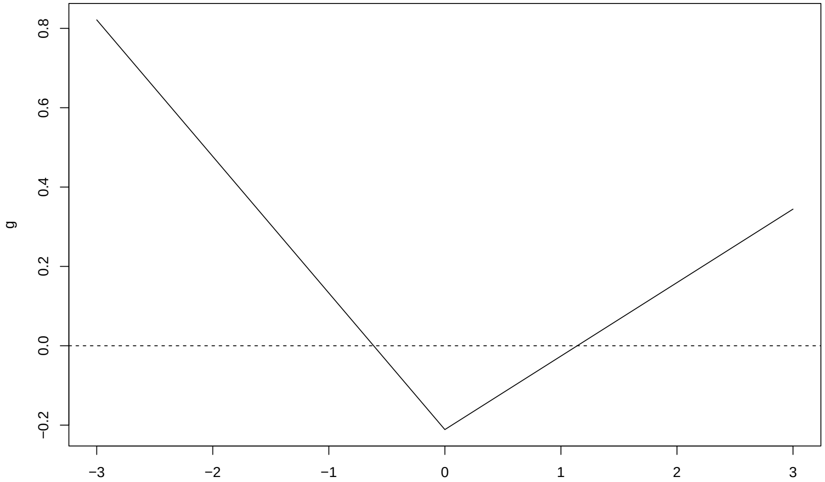
où $g(\epsilon) = \theta\epsilon + \gamma[|\epsilon| - E[|\epsilon|]]$.

Notes :

- $E[\epsilon_t] = 0$, $E[|\epsilon_t| - E[|\epsilon_t|]] = 0$, $E[g(\epsilon_t)] = 0$.
- Par exemple, si $\epsilon_t \sim N(0,1)$, $E[|\epsilon_t|] = \sqrt{2/\pi}$
- $\ln \sigma_t^2$ est un processus AR(1), puisque $g(\epsilon_t)$ est un bruit blanc.
- Pour $\theta < 0$, il y a un effet de levier.
- Pas besoin de contraintes sur les coefficients pour assurer la positivité de la volatilité, grace à la spécification logarithmique.

La fonction $g(\epsilon)$ de l'équation (3.31) (un exemple)

```
eps = seq(-3, 3, by=0.01)
theta = -0.0795; gamma = 0.2647
g = theta * eps + gamma * (abs(eps) - sqrt(2/pi))
plot(eps, g, type='l'); abline(h=0, lty=2)
```

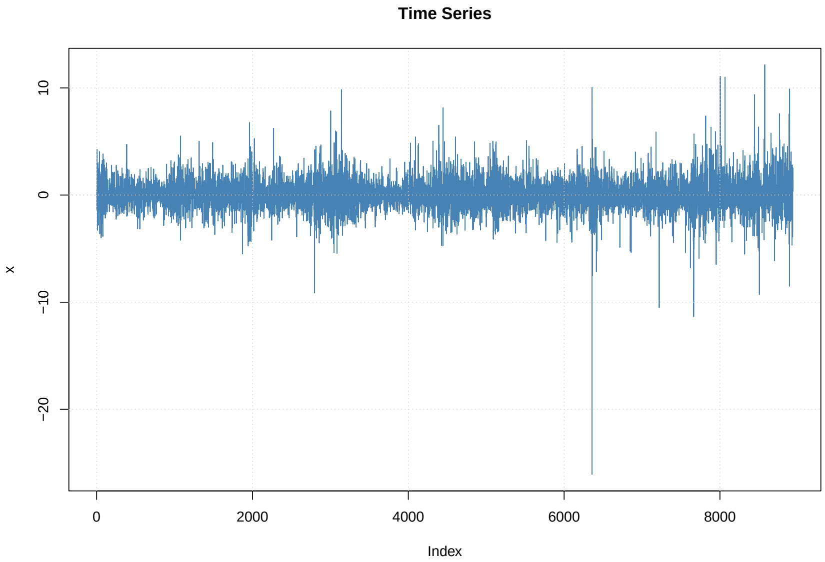


Ajustement de plusieurs modèles GARCH (code)

```
library(fGarch)
# Séries IBM journalière, log rendements 1962-97
r = scan('d-ibmln.txt')

# GARCH(1, 1) gaussien
gn = garchFit(~garch(1,1), cond.dist='norm', data=r)

# mu_t : ARMA(1, 0), sigma_t : GARCH(1, 1) gaussien
agn = garchFit(~arma(1,0)+garch(1,1), cond.dist='norm', dat

# GARCH(1, 1) t de Student
gt = garchFit(~garch(1,1), cond.dist='std', data=r)
```

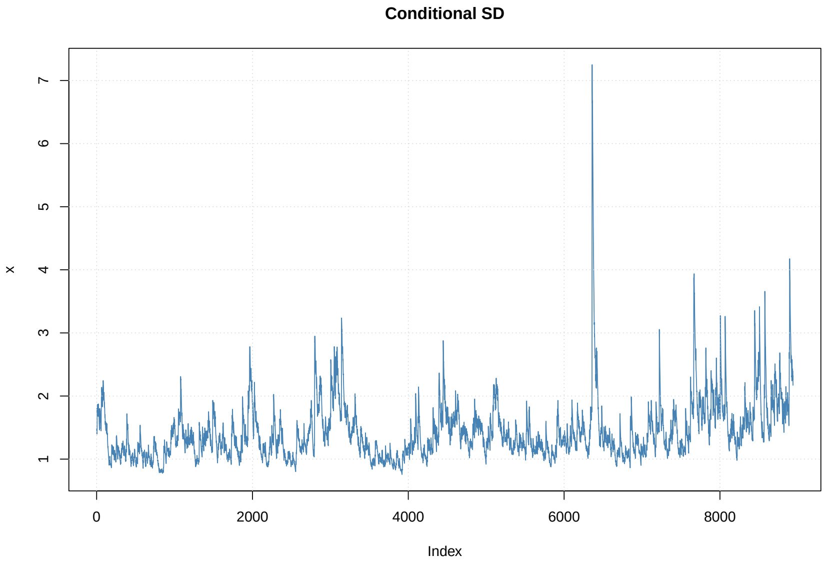
Données IBM journalière, r_t

```r
plot(gn, which=1)
```



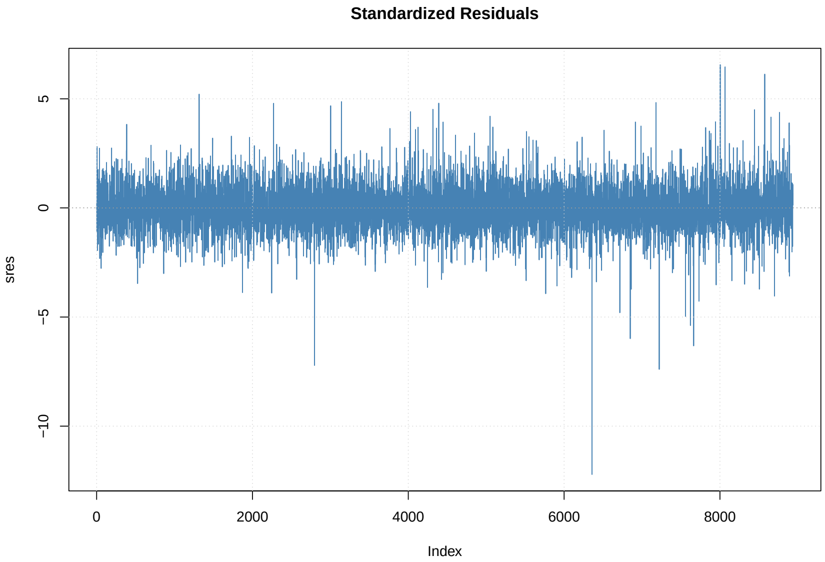
GARCH(1,1) gaussien, $\hat{\sigma}_t$

```
plot(gn, which=2)
```



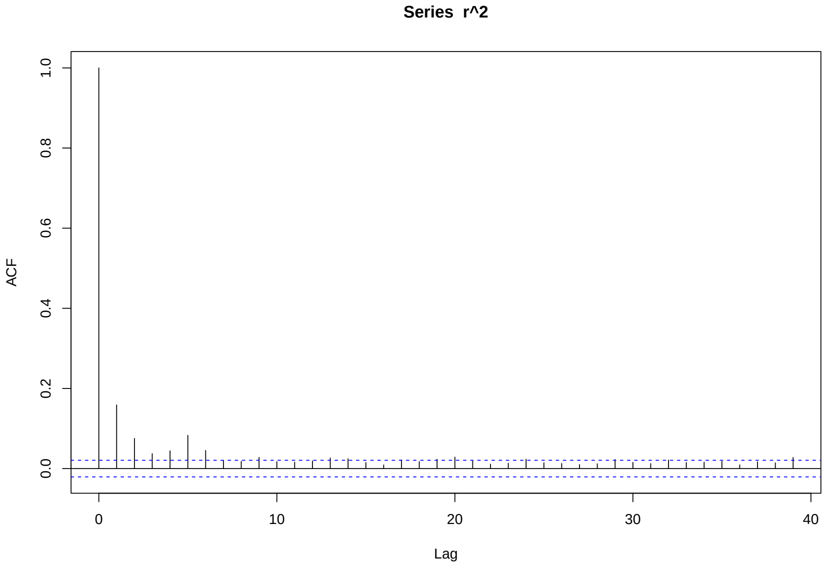
GARCH(1,1) gaussien, $\hat{\epsilon}_t$

```
plot(gn, which=9)
```



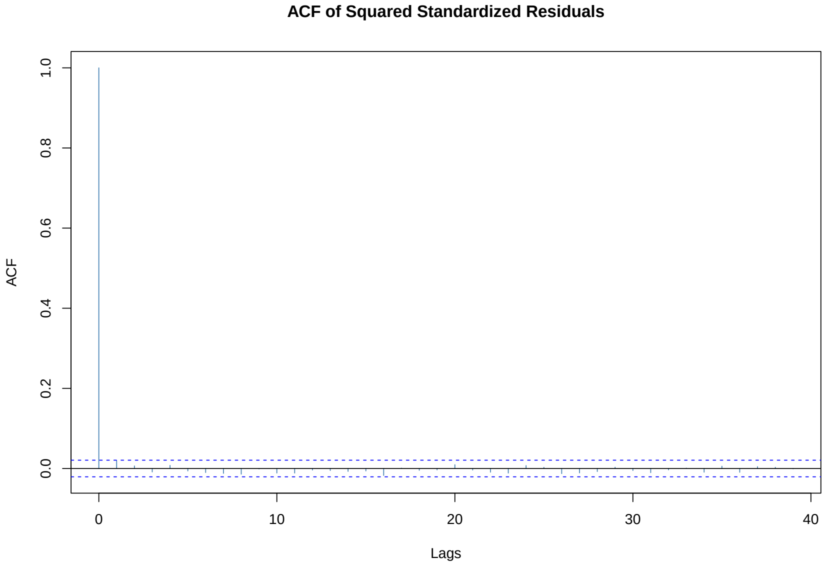
ACF(r_t^2)

```r
acf(r^2)
```



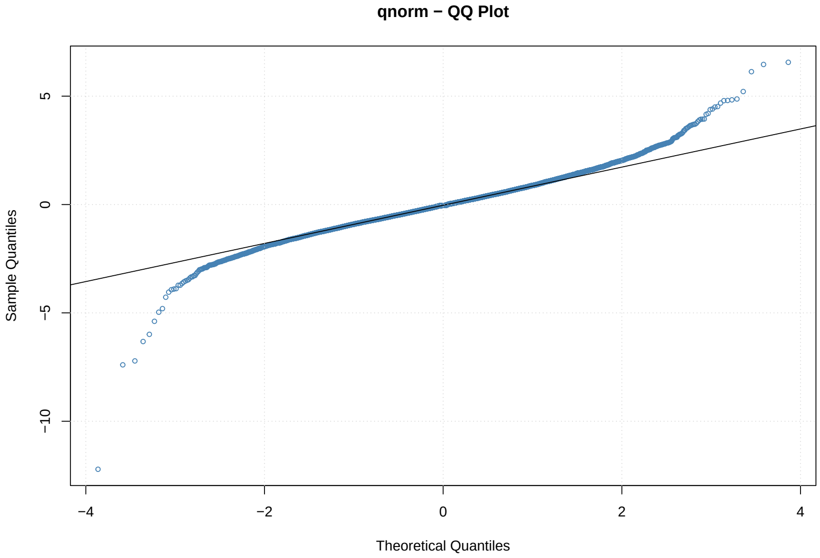
GARCH(1,1) gaussien, ACF($\hat{\epsilon}_t^2$)

```
plot(gn, which=11)
```



GARCH(1,1) gaussien, graphique Q-Q pour $\hat{\epsilon}_t$

```
plot(gn, which=13)
```

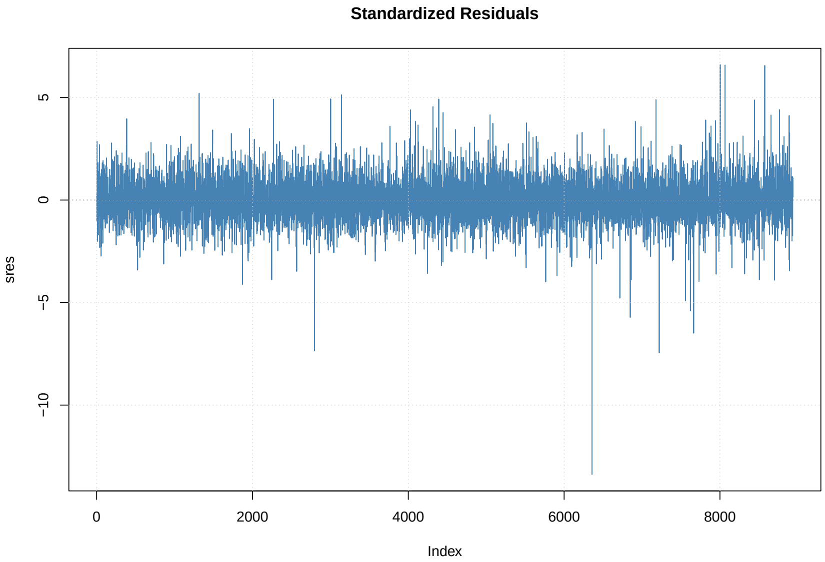


GARCH(1,1) gaussien, sommaire des résultats

- Le modèle capture bien l'autodépendence de volatilité.
- Le modèle capture mal l'asymétrie et surtout l'aplatissement conditionnel.

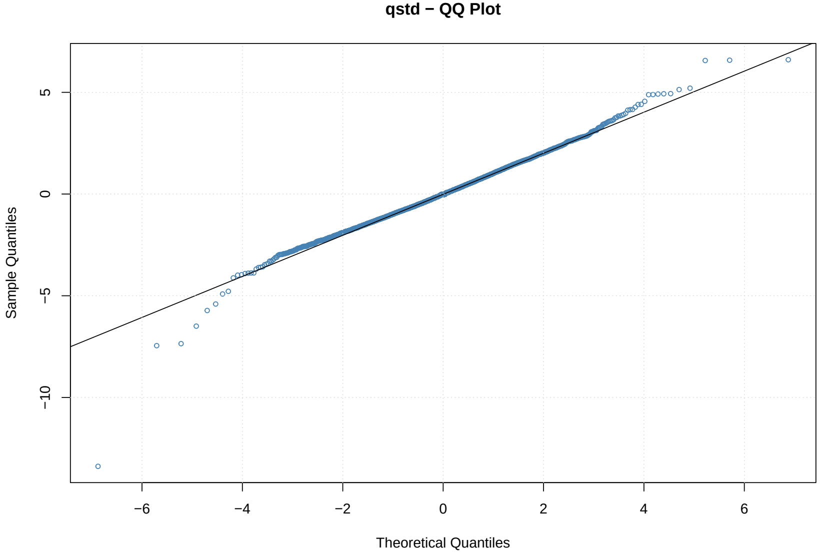
GARCH(1,1) t de Student, $\hat{\epsilon}_t$

```
plot(gt, which=9)
```



GARCH(1,1) t de Student, graphique Q-Q pour $\hat{\epsilon}_t$

```
plot(gt, which=13)
```



GARCH(1,1) t de Student, sommaire des résultats

- Le modèle capture mieux l'aplatissement conditionnel que le modèle GARCH(1,1) gaussien,
- mais pas parfaitement :
  - le modèle ne capture pas bien les (mettons) 10 valeurs les plus extrêmes (sur ≈ 9000)
  - il y a plus de valeurs extrêmes que prévu par le modèle (mauvaise spécification de l'évolution de la variance conditionnelle σ_t ou de la loi conditionnelle ou des deux?)
- Une asymétrie : les valeurs extrêmes négatives sont particulièrement extrêmes.

Cours 6, la semaine prochaine

Plan préliminaire

1. Un modèle de volatilité stochastique
2. Inférence bayésienne : un peu de théorie
3. Inférence bayésienne : un peu de computation