

# ECN 6578A, Économétrie des marchés financiers, Hiver 2020

## Cours 5

William McCausland

2020-01-31

# Plan

1. Maximum de vraisemblance : l'exemple le plus simple
2. Maximum de vraisemblance : un peu de théorie
3. La vraisemblance d'un ARCH(1), de la semaine passée.
4. Prédiction de la volatilité avec un ARCH(m), de la semaine passée.

# Le modèle binomial

- Supposons que les  $y_i$  sont iid Bernoulli avec probabilité  $\theta \in [0, 1]$  :

$$\begin{aligned}f(y_i|\theta) &= \begin{cases} \theta & y_i = 1 \\ (1 - \theta) & y_i = 0 \end{cases} \\ &= \theta^{y_i}(1 - \theta)^{1-y_i}\end{aligned}$$

- On observe  $y = (y_1, \dots, y_n)$  ; 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

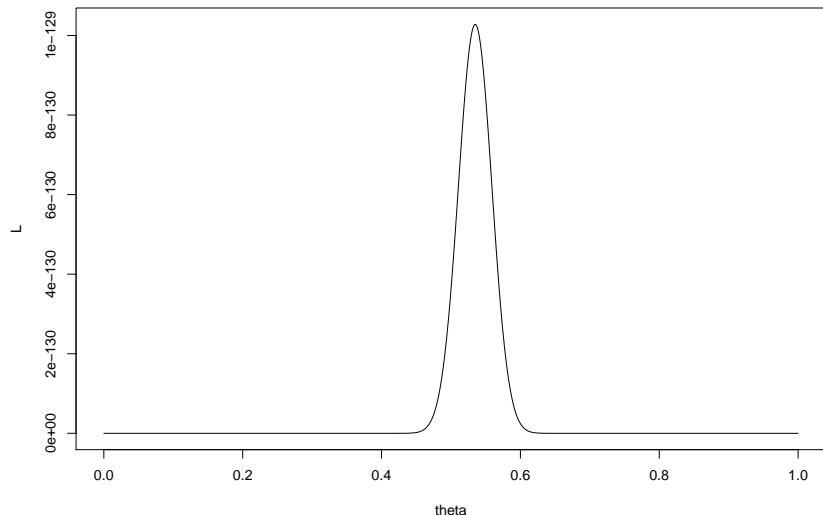
- $n_1 = \sum_{i=1}^n y_i$  est le nombre de fois qu'on observe 1, et
- $n_0 = n - \sum_{i=1}^n y_i$  est le nombre de fois qu'on observe 0.

# Deux interprétations 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  $\theta$  fixe, les probabilités relatives de plusieurs séquences  $(y_1, \dots, y_n)$ .
- ▶  $\mathcal{L}(\theta; y)$  donne, pour  $y$  fixe (le vecteur des données observées) une note (ou évaluation) à chaque valeur  $\theta$  pour la qualité de sa prévision des données observées.
- ▶ Soit  $L(\theta; y) = \log \mathcal{L}(\theta; y)$ , la log-vraisemblance.

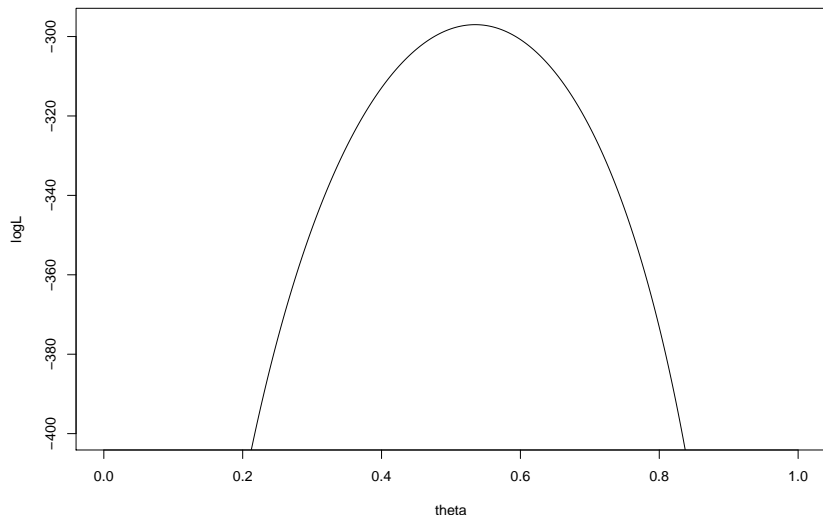
## La vraisemblance Bernoulli pour $n_0 = 200$ , $n_1 = 230$

```
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)))
```



# La fonction de vraisemblance pour une séries chronologique

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

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

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

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

- ▶ Pourquoi la log-vraisemblance et non juste la vraisemblance?
  - ▶ Pas de underflow numérique (soutassement arithmétique)
  - ▶ Plus facile à maximiser (la dérivée d'une somme est la somme des dérivées)

# 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

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



# Maximum de vraisemblance : conditions de régularité

- ▶ Définitions :

- ▶  $\theta$  est le vecteur des paramètres ;  $\Theta$ , l'ensemble de toutes les valeurs possibles de  $\theta$ .
- ▶  $r$  est le vecteur (aléatoire) des données.

- ▶ Conditions informelles de régularité :

1. Le modèle est correct pour une valeur  $\theta = \theta_0 \in \Theta$ .
2. La vraie valeur  $\theta_0$  est dans l'intérieur de  $\Theta$ .
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 :

1.  $\hat{\theta} = \arg \max_{\theta} L(\theta; r)$  existe, est unique.
2.  $\hat{\theta} \rightarrow_p \theta_0$  (loi de grands nombres)
3.  $\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d N(0, \mathcal{I}(\theta_0)^{-1})$  (théorème central limite)
4.  $E_{r|\theta} \left[ \frac{\partial L(\theta; r)}{\partial \theta} \right] = 0$ , alors  $\mathcal{I}(\theta) = \text{Var}_{r|\theta} \left[ \frac{\partial L(\theta; r)}{\partial \theta} \right]$ .
5.  $\mathcal{I}(\theta) = E_{r|\theta} \left[ -\frac{\partial^2 L(\theta; r)}{\partial \theta \partial \theta^\top} \right]$ .

Problèmes restants :

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

## Comment trouver $\hat{\theta}$ !

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

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

- Processus itératif pour trouver  $\hat{\theta}$ :  $\theta_1, \theta_2, \dots$ ,
- Approximation quadratique local autour de  $\theta_k$  :

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

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

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

- $\tilde{s}(\theta_{k+1}) = 0$  définit  $\theta_{k+1}$  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 vraisemblance est loin de quadratique.
- Solution de BHHH : choisir une valeur scalaire  $\lambda_k$  et calculer

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

1. Calculez  $s(\theta_k)$ ,  $H(\theta_k)$ .
  2. Trouvez une bonne valeur de  $\lambda_k$  (recherche linéaire)
- Des fois, on utilise souvent, au lieu de  $H(\theta)$ ,

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

- Avec  $r_t$  indépendants, une loi de grands nombres donne

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

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

- ▶ On utilise  $H(\hat{\theta})$  ou  $\hat{H}(\hat{\theta})$  au lieu de  $\mathcal{I}(\theta_0)$ , qui est inconnu.
- ▶ Convergence de  $\hat{\theta}$  à  $\theta_0$ .
- ▶ Convergence de  $\hat{H}(\cdot)$  à  $H(\cdot)$
- ▶ Convergence de  $H(\theta_0)$  à  $\mathcal{I}(\theta_0) = E[H(\theta_0)]$ .
- ▶ Ensemble, convergence de  $\hat{H}(\hat{\theta})$  à  $\mathcal{I}(\theta_0)$ .

## 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)

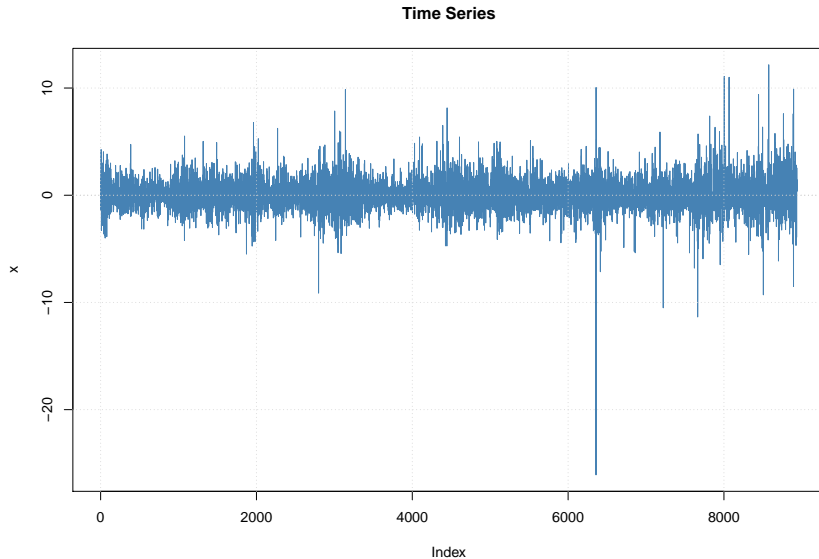
# mu_t : ARMA(1, 0), sigma_t : GARCH(1, 1) gaussien
agn = garchFit(~arma(1,0)+garch(1,1), cond.dist='norm', data=r)

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

# GARCH(1, 1) t de Student avec asymétrie
gst = garchFit(~garch(1,1), cond.dist='sstd', data=r)
```

# Données IBM journalière, $r_t$

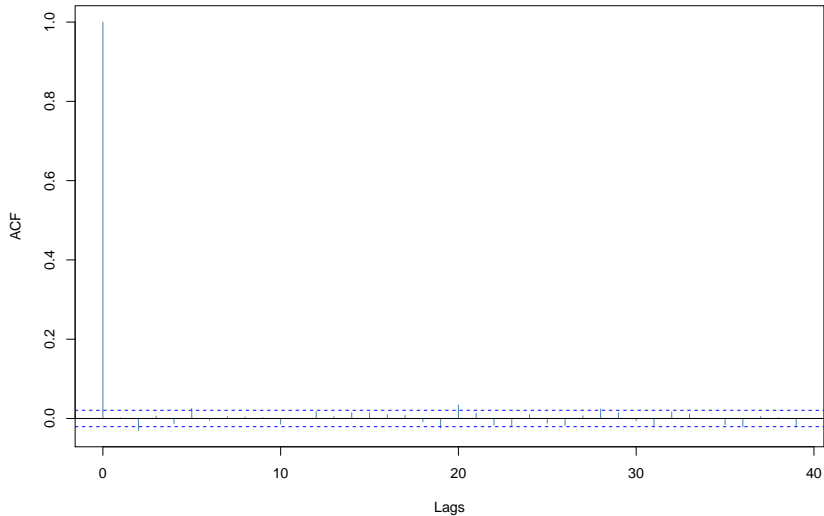
```
plot(gn, which=1)
```



$ACF(r_t)$

```
plot(gn, which=4)
```

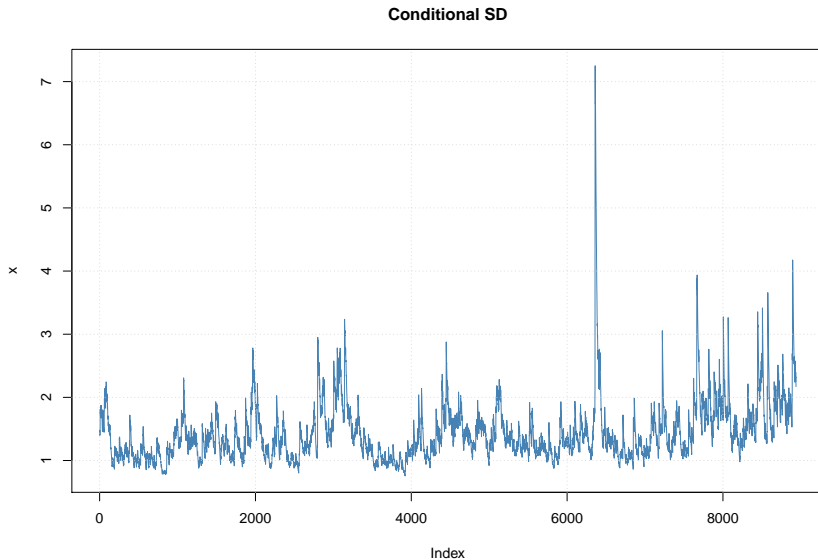
ACF of Observations





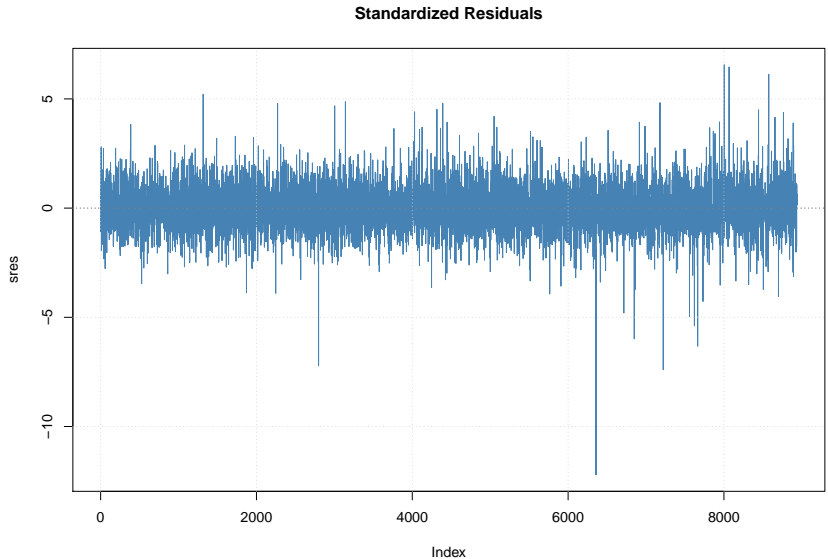
# GARCH(1,1) gaussien, $\hat{\sigma}_t$

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



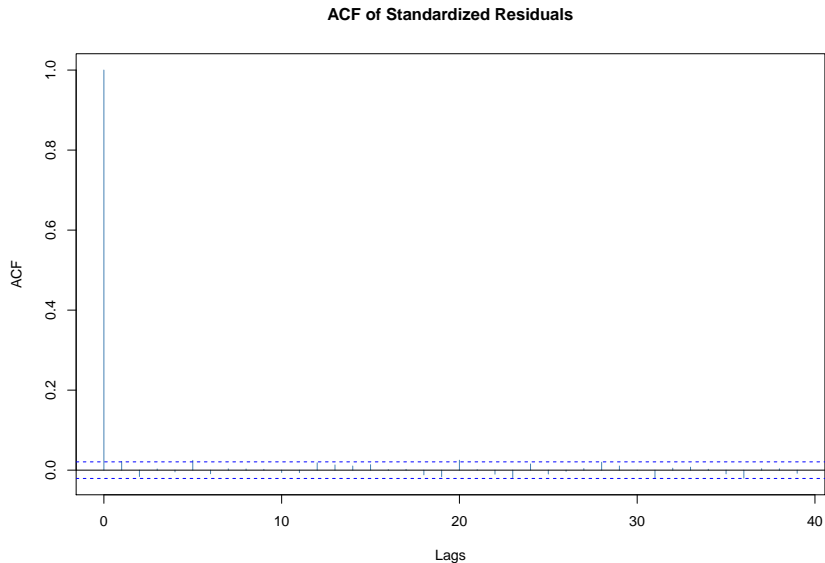
# GARCH(1,1) gaussien, $\hat{\epsilon}_t$

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



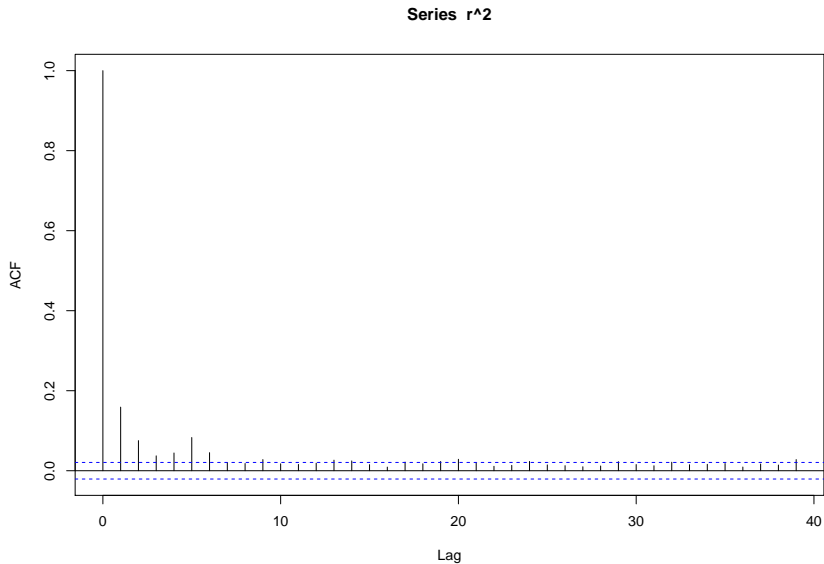
# GARCH(1,1) gaussien, $ACF(\hat{\epsilon}_t)$

```
plot(gn, which=10)
```



# GARCH(1,1) gaussien, $ACF(r_t^2)$

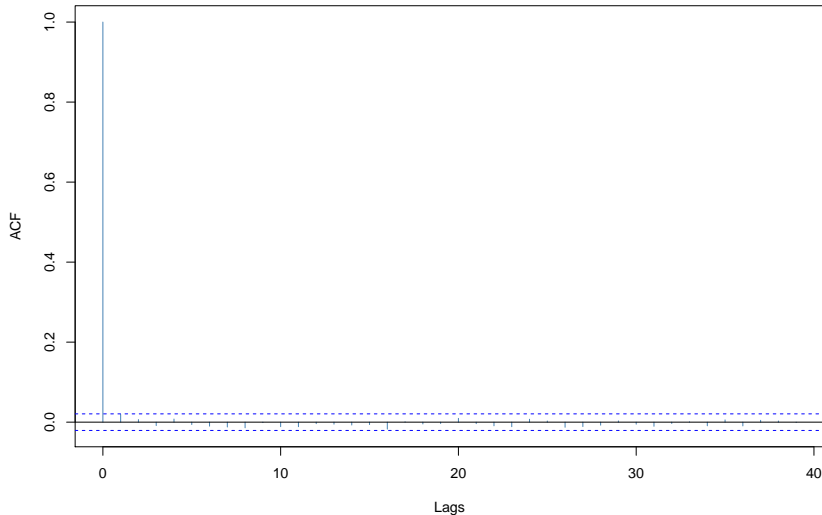
```
acf(r^2)
```



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

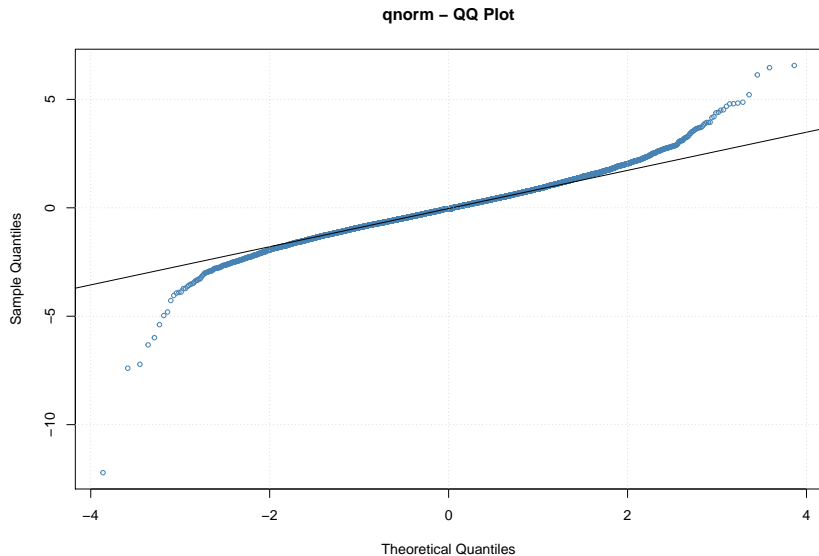
```
plot(gn, which=11)
```

ACF of Squared Standardized Residuals



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

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

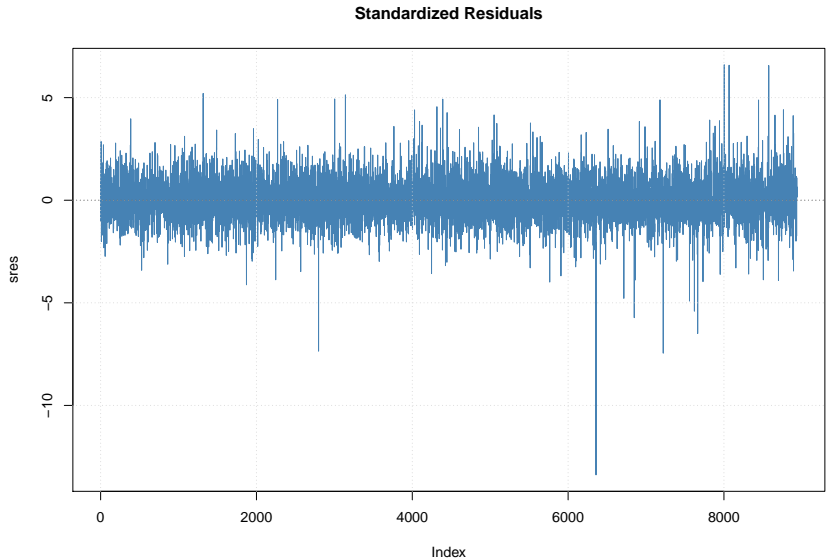


## GARCH(1,1) gaussien, sommaire des résultats

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

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

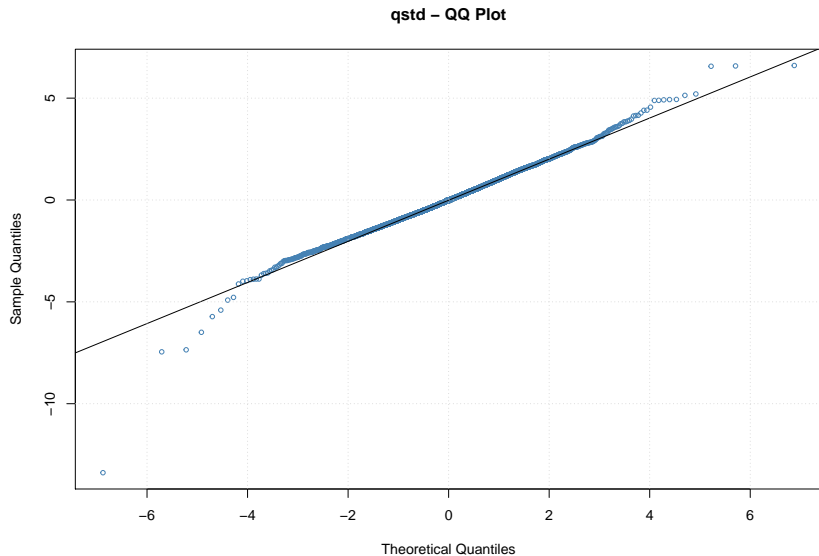
```
plot(gt, which=9)
```





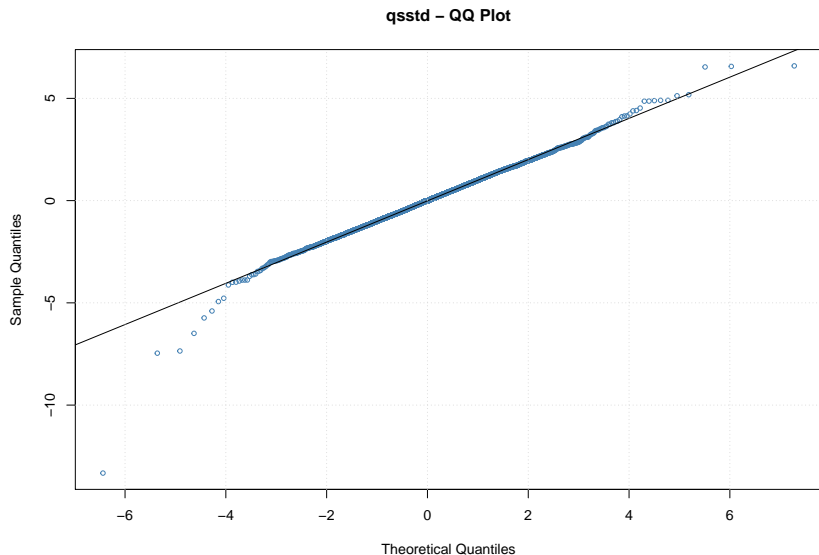
# GARCH(1,1) $t$ de Student, graphique Q-Q pour $\hat{\epsilon}_t$

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



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

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



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

- ▶ Le modèle capture mieux l'aplatissement conditionnel.
- ▶ Le modèle asymétrique capture un peu mieux l'asymétrie conditionnel.
- ▶ Le modèle ne capture pas bien les (mettons) 10 valeurs les plus extrêmes (sur  $\approx 9000$ )

## Exercices R

Pour ces questions, utilisez les données dans le fichier d-3m7008.txt. (3M, rendements journaliers, 1970-2008) Je recommande le paquet fGARCH.

1. Pour quel modèle est-ce que la vraisemblance est plus élevée? Quelles différences voyez-vous dans la séquence de volatilités estimées?
  - a. Modèle GARCH(1,1), distribution conditionnelle  $t$  de Student.
  - b. Modèle ARCH(2), distribution conditionnelle  $t$  de Student.
2. Pour quel modèle est-ce que la vraisemblance est plus élevée?
  - a. Modèle GARCH(2,1), distribution conditionnelle  $t$  de Student.
  - b. Modèle GARCH(1,2), distribution conditionnelle  $t$  de Student.
  - c. Modèle AR(1)-GARCH(1,1), distribution conditionnelle  $t$  de Student.
3. Parmi les cinq modèles ci-dessus, quel est le meilleur selon le critère AIC?

## Question théorique

1. Faites des prévision des rendement  $r_{T+1}$  et  $r_{T+2}$  pour une modèle AR(1)-GARCH(1,1). Quelle est la variance conditionnelle des erreurs de prévision? Exprime le résultat en termes des paramètres,  $r_T$  et  $\sigma_T^2$ .

# Cours 6, la semaine prochaine

## Plan préliminaire

1. Modèle EGARCH et l'effet de levier
2. Introduction à l'inférence bayésienne
3. Modèles de volatilité stochastique

## Lectures

Tsay, 3.8 intro, 3.8.1. 3.12, 12.3.1