GARCH models

Mathieu Rosenbaum

École Polytechnique

January 2024

Plan

Introduction

GARCH models

- GARCH (Generalized Autoregressive Conditionally Heteroskedastic) models are the non iid models which are the most commonly used in finance, notably for risk management and regulation.
- Before introducing them, we recall some classical stylized facts of financial data.

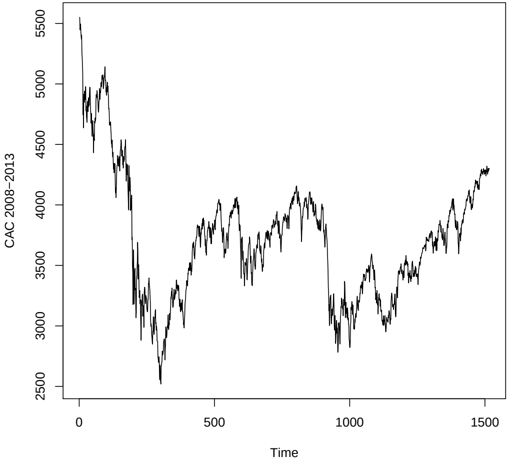
Plan

Empirical properties of financial returns

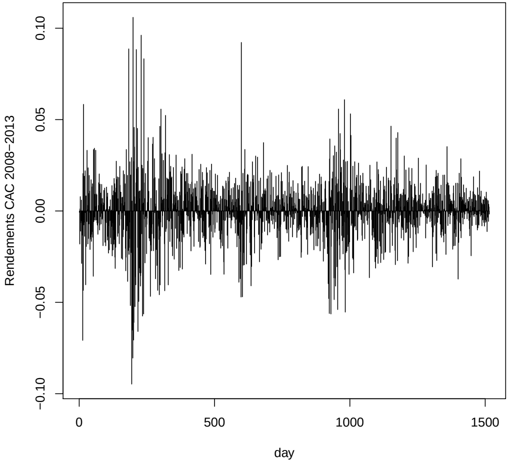
Framework

- We are interested in modelling low frequency (daily) returns of financial assets.
- $$r_t = \frac{p_t - p_{t-1}}{p_{t-1}} \text{ or } \varepsilon_t = \log(p_t/p_{t-1}),$$
  with p_t the price of the asset at time t and 1 being typically one day.
- The two returns defined above are very close and we use one or the other depending on the application.
- A good model must satisfy the following classical properties of low frquency financial data.

Empirical properties of financial returns



Empirical properties of financial returns



Empirical properties of financial returns

Stationarity

- Contrary to (p_t), (ε_t) is in general stationary at least at order 2.
- This property generalized the iid assumption and is crucial to do estimation and forecasting.

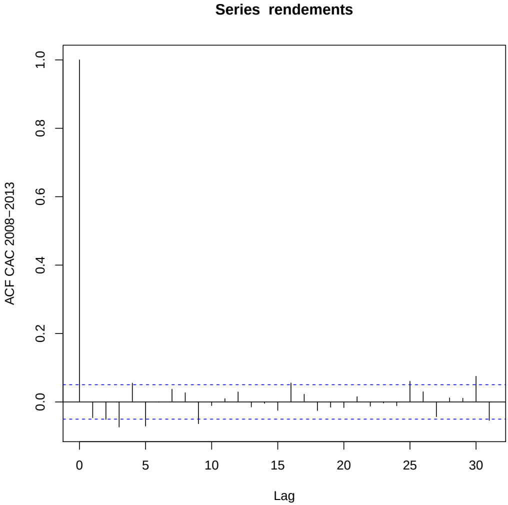
Definition (Stationarity)

- A process $(X_t)_{t\in\mathbb{N}}$ is said to be strictly stationary if for any $k \in \mathbb{N}$ and any $h \in \mathbb{N}$, $(X_1, \dots, X_k)$ and $(X_{1+h}, \dots, X_{k+h})$ have the same law.
- It is stationary at order 2 if for any $t \in \mathbb{N}$ and $h \in \mathbb{N}$,

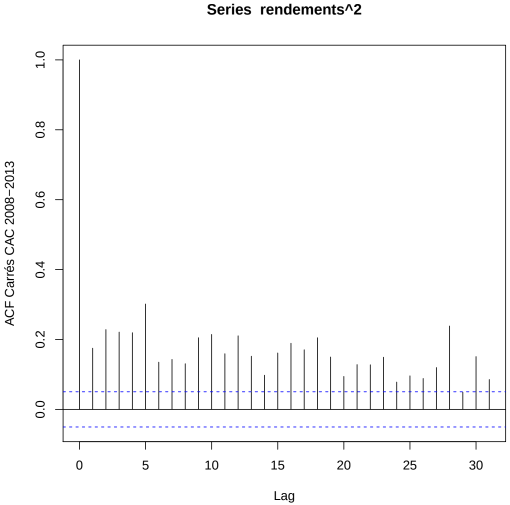
$$\mathbb{E}[X_t^2] < +\infty,\ \mathbb{E}[X_t] = m,\ \mathrm{Cov}[X_t, X_{t+h}] = \gamma(h),$$

where $m \in \mathbb{R}$ and γ is the autocovariance function.
- Stationarity at order 2 is essentially weaker than strict stationarity but implies existence of a variance.

Empirical properties of financial returns



Empirical properties of financial returns



Empirical properties of financial returns

White noise behaviour

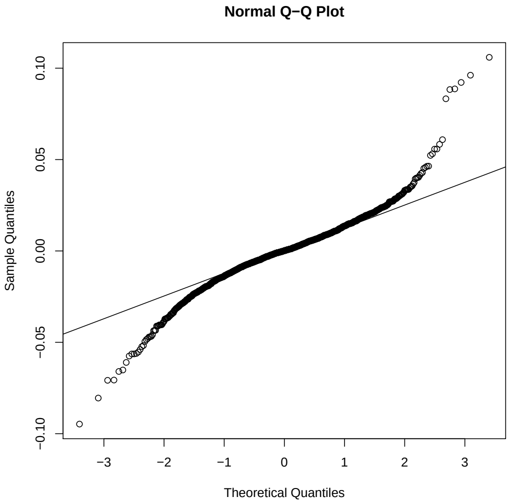
- (ε_t) is in general close to a white noise, that is a stationary process at order 2 with $\gamma(h) = 0$ for $\neq 0$.
- However, there is significant autocorrelation in the time series of (ε_t^2). This is not compatible with an iid white noise.

Empirical properties of financial returns

Volatility clustering

- Large values of ε_t^2 tend to be followed by high values of ε_t^2.
- High volatility regimes and low volatility regimes.
- This is not incompatible with stationary returns, and in particular with a constant variance of the returns.
- However the conditional variance of ε_t (conditional on $\varepsilon_{t-1},\ \varepsilon_{t-2}\ldots$) seems not constant.
- Conditional heteroskedasticity not incompatible with marginal homoskedasticity.

Empirical properties of financial returns



Empirical properties of financial returns

Fat tails

- Goodness of fits testing procedure with respect to the Gaussian distribution are rejected.
- Tails are fatter than Gaussian ones.
- More extreme events than in a Gaussian setting.
- In particular, the Kurtosis $\mathbb{E}[(X - \mathbb{E}[X])^4]/(\mathrm{Var}[X])^2$ is higher than 3.

Other properties

- Leverage effect : large negative returns increase volatility more than small ones.
- Some seasonalities : Monday and Friday are specific...

Empirical properties of financial returns

ARMA processes

- The previous properties are not well reproduced by the ARMA(p,q) classically used in time series modelling : for any $t \in \mathbb{Z}$:

$$X_t + \sum_{i=1}^{p} a_i X_{t-i} = \mu + \varepsilon_t + \sum_{j=1}^{q} b_j \varepsilon_{t-j},$$

  with (ε_t) a white noise and X_t stationary at order 2.
- ARMA type processes are interesting when we focus on the autocovariances of a process.
- However financial time series behave as a white noise.
- A specificity of financial time series is that $\mathrm{Var}[\varepsilon_t|\varepsilon_{t-1},\varepsilon_{t-2},\ldots]$ is not constant.
- This is not compatible with ARMA type processes.

Plan

GARCH(p, q) processes

Introduction

- ARCH models (Autoregressive Conditionally Heteroskedastic) : Engle, 1982.
- GARCH models (Generalized Autoregressive Conditionally Heteroskedastic) : Bollerslev, 1986.
- Reproduce the previous properties.
- Can be easily estimated.
- Enable us to easily do forecasting.

GARCH(p, q) processes

Definition (GARCH(p, q))

- Let (η_t) be a sequence of iid random variables with density such that $\mathbb{E}[\eta_t] = 0$ and $\mathrm{Var}[\eta_t] = 1$ (building blocks).
- (ε_t) is a GARCH(p, q) process with respect to (η_t) if for any $t \in \mathbb{Z}$:

$$\varepsilon_t = \sigma_t \eta_t, \quad \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2,$$

  with $\omega > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, σ_t^2 measurable with respect to $\sigma(\varepsilon_u,\ u < t)$, η_t independent of $\sigma(\varepsilon_u,\ u < t)$.
- In our modelling ε_t : return, σ_t : volatility.

GARCH(p, q) processes

Remarks

- If $p = 0$: ARCH(q).
- $\mathbb{E}[\varepsilon_t] = \mathbb{E}[\mathbb{E}[\sigma_t \eta_t | \varepsilon_u,\ u < t]] = \mathbb{E}[\sigma_t]\mathbb{E}[\eta_t] = 0.$
- $\mathbb{E}[\varepsilon_t | \varepsilon_u,\ u < t] = \sigma_t \mathbb{E}[\eta_t | \varepsilon_u,\ u < t] = 0.$
- $\mathrm{Var}[\varepsilon_t | \varepsilon_u,\ u < t] = \mathbb{E}[\varepsilon_t^2 | \varepsilon_u,\ u < t] = \sigma_t^2.$
- σ_t^2 represents the conditional variance of the returns.
- $\mathrm{Cov}[\varepsilon_t, \varepsilon_{t-h}] = \mathbb{E}[\varepsilon_t \varepsilon_{t-h}] = \mathbb{E}[\eta_t]\mathbb{E}[\sigma_t \varepsilon_{t-h}] = 0.$
- ε_t is a white noise (non iid).

GARCH(p, q) processes

ARMA representation

- Let $\nu_t = \varepsilon_t^2 - \sigma_t^2 = \varepsilon_t^2 - \mathbb{E}[\varepsilon_t^2 | \varepsilon_u,\ u < t]$. We can show that ν_t is a white noise.

$$\varepsilon_t^2 - \nu_t = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j (\varepsilon_{t-j}^2 - \nu_{t-j})$$

$$\varepsilon_t^2 = \omega + \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j (\nu_{t-j}),$$

  with the convention $\alpha_i = 0$ for $i > q$, $\beta_j = 0$ for $j > p$.
- ε_t^2 admits an ARMA(p, q) representation.
- This will be useful for estimation of parameters.
- For the GARCH($1, 1$) it implies for example that $\mathrm{Corr}[\varepsilon_t^2, \varepsilon_{t-h}^2] = c(\alpha_1 + \beta_1)^k$ with c not depending on k.

Plan

Stationary solution

Question

- Does a stationary, non anticipative process satisfying the GARCH equations exist ? (non anticipative : ε_t must be a measurable function of the η_{t-s}, $s \geq 0$. Actually we even impose that σ_t^2 is measurable with respect to the η_{t-s}, $s > 0$).
- We start by the GARCH$(1,1)$ model, the most widely used in practice :

$$\varepsilon_t = \sigma_t \eta_t, \quad \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{1}$$

  with $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$.
- We set $a(z) = \alpha z^2 + \beta$.

Stationary solution

Theorem (Strictly stationary solution)

- Let $\gamma = \mathbb{E}[\log(a(\eta_t))]$.
- If $-\infty \leq \gamma < 0$, the series

$$h_t = \Big(1 + \sum_{i=1}^{+\infty} a(\eta_{t-1}) \dots a(\eta_{t-i})\Big)\omega$$

  converges almost surely and the process ε_t defined by $\varepsilon_t = \sqrt{h_t}\eta_t$ is the unique strictly stationary solution of (1).
- This solution is non-anticipative and ergodic (see proof).
- If $\gamma \geq 0$ there is no strictly stationary solution.

Stationary solution

Remarks

- $\gamma = \mathbb{E}[\log(a(\eta_t))]$ exists in $[-\infty, +\infty)$ almost surely as

$$\mathbb{E}[\max(\log, 0)(a(\eta_t))] \leq \mathbb{E}[a(\eta_t)] = \alpha + \beta.$$

- The condition of the theorem implies $\beta < 1$.
- If $\alpha + \beta < 1$, by Jensen

$$\mathbb{E}[\log(a(\eta_t))] \leq \log \mathbb{E}[a(\eta_t)] = \log(\alpha + \beta) < 0.$$

  Therefore the condition is satisfied. It is a necessary condition.

- Proof of the theorem : on tablet.

Stationary solution

Theorem (Stationary solution at order 2)

- If $\alpha + \beta \geq 1$, there is no non-anticipative stationary solution at order 2 of the GARCH$(1,1)$ equations.
- If $\alpha + \beta < 1$, the solution of (1) defined in the previous theorem is stationary at order 2 (it is even a white noise).
- If $\alpha + \beta < 1$, there is no other non-anticipative stationary solution at order 2.
- In the GARCH(p,q) case the condition becomes

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1.$$

- Proof of the theorem : on tablet.

Plan

Fat tails

Kurtosis

- We continue to show that GARCH processes reproduce the listed empirical properties.
- The conditional moments of order k are proportional to σ_t^{2k} : $\mathbb{E}[\varepsilon_t^{2k}|\varepsilon_u,\ u<t] = \sigma_t^{2k}\mathbb{E}[\eta_t^{2k}]$. Thus,

$$\frac{\mathbb{E}[\varepsilon_t^4|\varepsilon_u,\ u<t]}{\big(\mathbb{E}[\varepsilon_t^2|\varepsilon_u,\ u<t]\big)^2} = \text{Kurtosis}(\eta_t).$$

- For the unconditional kurtosis :

$$\frac{\mathbb{E}[\varepsilon_t^4]}{\big(\mathbb{E}[\varepsilon_t^2]\big)^2} = \frac{\mathbb{E}\big[\mathbb{E}[\varepsilon_t^4|\varepsilon_u,\ u<t]\big]}{\big(\mathbb{E}\big[[\varepsilon_t^2|\varepsilon_u,\ u<t]\big]\big)^2} = \frac{\mathbb{E}[\sigma_t^4]}{(\mathbb{E}[\sigma_t^2])^2}\text{Kurtosis}(\eta_t) > \text{Kurtosis}(\eta_t).$$

- The random nature of volatility implies fatter distribution tails.

Plan

Forecasting

Forecasting formulas

- GARCH processes are very convenient to make variance forecasts.
- For any $h > 0$ and $t \in \mathbb{Z}$,
  $\mathbb{E}[\varepsilon_{t+h}|\varepsilon_u,\ u < t] = \mathbb{E}\big[\mathbb{E}[\varepsilon_{t+h}|\varepsilon_u,\ u < t+h]|\varepsilon_u,\ u < t\big] = 0.$

$$\mathbb{E}[\varepsilon^2_{t+h}|\varepsilon_u,\ u < t] = \mathbb{E}[\sigma^2_{t+h}|\varepsilon_u,\ u < t]$$
$$= \omega + \sum_{i=1}^{q} \alpha_i \mathbb{E}[\varepsilon^2_{t+h-i}|\varepsilon_u,\ u < t] + \sum_{j=1}^{p} \beta_j \mathbb{E}[\sigma^2_{t+h-j}|\varepsilon_u,\ u < t].$$

- For $i > h$: $\mathbb{E}[\sigma^2_{t+h-i}|\varepsilon_u,\ u < t] = \sigma^2_{t+h-i}$,
  $\mathbb{E}[\varepsilon^2_{t+h-i}|\varepsilon_u,\ u < t] = \varepsilon^2_{t+h-i}$.
- For $i = h$: $\mathbb{E}[\sigma^2_{t+h-i}|\varepsilon_u,\ u < t] = \sigma^2_{t+h-i}$,
  $\mathbb{E}[\varepsilon^2_{t+h-i}|\varepsilon_u,\ u < t] = \sigma^2_{t+h-i}$.
- For $i < h$: $\mathbb{E}[\varepsilon^2_{t+h-i}|\varepsilon_u,\ u < t] = \mathbb{E}[\sigma^2_{t+h-i}|\varepsilon_u,\ u < t]$.
- The predictions are then computed recursively.

Plan

Parameters estimation of GARCH(p, q)

Quasi-conditional maximum likelihood

- Most efficient and accurate approach to estimate $\theta = (\omega, \alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p)$.
- We compute the conditional likelihood as if the (η_t) are $\mathcal{N}(0,1)$ (but we do not do this assumption !).

$$\varepsilon_t = \sigma_t \eta_t, \quad \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2.$$

Parameters estimation of GARCH(p, q)

Quasi-conditional maximum likelihood

- Given starting values $\varepsilon_0^2, \dots, \varepsilon_{1-q}^2$, $\tilde{\sigma}_0^2, \dots, \tilde{\sigma}_{1-p}^2$, the Gaussian likelihood at point $(\varepsilon_n, \dots, \varepsilon_1)$ writes

$$\begin{aligned} L_n(\theta, \varepsilon_n, \dots, \varepsilon_1) &= L_n(\theta, \varepsilon_n | \varepsilon_{n-1}, \dots, \varepsilon_1) * L_n(\theta, \varepsilon_{n-1}, \dots, \varepsilon_1) \\ &= \Pi_{i=1}^n \frac{1}{\sqrt{\tilde{\sigma}_t^2}} \exp\Big(- \frac{\varepsilon_t^2}{2\tilde{\sigma}_t^2,} \Big), \end{aligned}$$

with the $\tilde{\sigma}_t^2$ defined recursively for $t \geq 1$ by
$\tilde{\sigma}^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \tilde{\sigma}_{t-j}^2$.

Parameters estimation of GARCH(p, q)

Quasi-conditional maximum likelihood

- For a given value of θ, under second order stationarity assumption, the unconditional variance is a possible choice for

$$\varepsilon_0^2 = \ldots = \varepsilon_{1-q}^2 = \tilde{\sigma}_0^2, \ldots, \tilde{\sigma}_{1-p}^2 = \frac{\omega}{1 - \sum_{i=1}^q \alpha_i - \sum_{j=1}^p \beta_j}.$$

- The quasi maximum likelihood estimator $\hat{\theta}_n$ is defined by $L_n(\hat{\theta}_n) = \sup_{\theta \in \Theta} L_n(\theta)$ or equivalently

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmin}} - l_n(\theta) \text{ with } l_n(\theta) = \frac{1}{n} \sum_{t=1}^n \frac{\varepsilon_t^2}{2\tilde{\sigma}_t^2} + \log(\tilde{\sigma}_t^2).$$

- Under suitable assumptions, $\hat{\theta}_n \to_{\mathbb{P}} \theta$, $\sqrt{n}(\hat{\theta}_n - \theta) \to_{\mathcal{L}} \mathcal{N}(0, W)$, with W explicit.