Time Series lecture 13

Non-Linear Time Series

Sofia Olhede

EPFL

May 21, 2025

Lecture outline

1. Motivation
2. ARCH
3. Diagnostics

Motivation

- So far, we have looked at linear time series models.
- These models are very useful, but they are not always appropriate.
- In this lecture, we will look at non-linear time series models.
- In particular, we will focus on models that are used to handle volatility clustering.
- This is often motivated by financial time series .
- In this context, one typically looks at the log-returns of a stock price:

$$r_t = \log\left(\frac{X_t}{X_{t-1}}\right) = \log(X_t) - \log(X_{t-1})$$

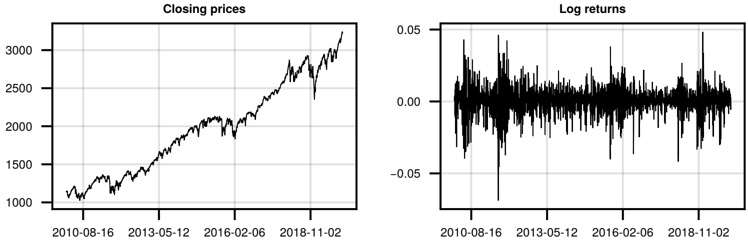

Figure: S&P 500 stock index prices from 01/01/2010 to 01/01/2020.

ACF and PACF

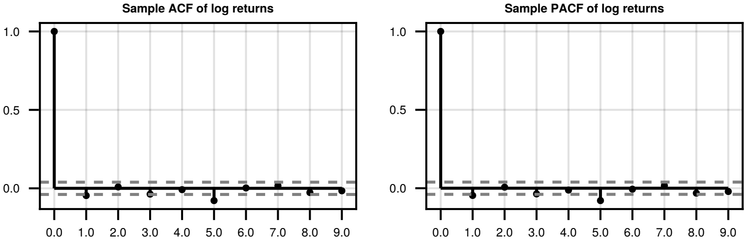

Figure: acf and pacf of the log-returns from 01/01/2010 to 01/01/2020.

ACF and PACF of the squared log-returns

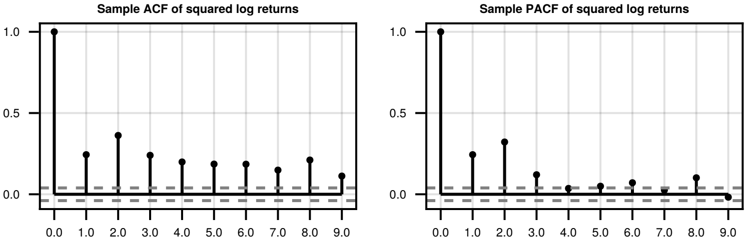

Figure: acf and pacf of the square of the log-returns from 01/01/2010 to 01/01/2020.

Marginal distribution of the log-returns

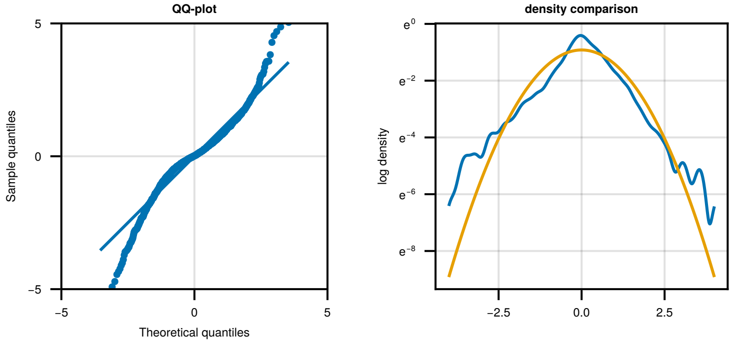

Figure: qq plot of standardised log returns (left) and log density (right) both compared to a standard normal.

ARCH

Definition 13.1 (ARCH models)

We will say that a time series $\{r_t\}$ is AutoRegressive Conditionally Heteroscedastic (ARCH) model if it takes the form

$$r_t = \sigma_t \varepsilon_t, \qquad t \in \mathbb{Z}$$

where $\{\varepsilon_t\}$ is a white noise process with mean zero and variance one, and

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2$$

is the conditional variance of the process. The parameters $\alpha_0, \alpha_1 > 0$.

- So we are doing the same thing as before, but for the variance of the noise process, not the process itself.
- The idea is that this will allow the conditional variance to change over time.

ARCH

- Informally, we write the information available at time F_t.
- We are interested in the properties of r_t given F_{t-1}.
- The conditional mean is given by

$$\mathbb{E}\left[r_t \mid F_{t-1}\right] = \mathbb{E}\left[\sigma_t \varepsilon_t \mid F_{t-1}\right] = \sigma_t \mathbb{E}\left[\varepsilon_t \mid F_{t-1}\right] = 0$$

- We now compute the conditional variance

$$\begin{aligned}
\operatorname{Var}\left(r_t \mid F_{t-1}\right) &= \mathbb{E}\left[r_t^2 \mid F_{t-1}\right] - \mathbb{E}\left[r_t \mid F_{t-1}\right]^2 \\
&= \mathbb{E}\left[\left(\sigma_t \varepsilon_t\right)^2 \mid F_{t-1}\right] \\
&= \sigma_t^2 \mathbb{E}\left[\varepsilon_t^2 \mid F_{t-1}\right] \\
&= \sigma_t^2 \\
&= \alpha_0 + \alpha_1 r_{t-1}^2.
\end{aligned}$$

ARCH: mean and variance

- We can also use this to compute unconditional expectations and variances.
- We can compute the unconditional expectation using the law of iterated expectation:

$$\begin{aligned}\mathbb{E}[r_t] &= \mathbb{E}[\mathbb{E}[r_t \mid F_{t-1}]] \\ &= \mathbb{E}[0] \\ &= 0\end{aligned}$$

ARCH: covariance

- We can also compute the marginal variance

$$\begin{aligned}\mathrm{Var}(r_t) &= \mathbb{E}\left[r_t^2\right] - \mathbb{E}[r_t]^2 \\ &= \mathbb{E}\left[\mathbb{E}\left[r_t^2 \mid F_{t-1}\right]\right] - 0^2 \\ &= \mathbb{E}\left[\alpha_0 + \alpha_1 r_{t-1}^2\right] \\ &= \alpha_0 + \alpha_1 \mathrm{Var}(r_{t-1})\end{aligned}$$

- Now assuming shift invariance, we can write

$$\mathrm{Var}(r_t) = \frac{\alpha_0}{1-\alpha_1}.$$

- We need $\alpha_1 < 1$ to get a stationary process.

- The autocorrelation for $\tau > 0$ is given by

$$\begin{aligned} \mathrm{Cov}\left(r_t, r_{t+\tau}\right) &= \mathbb{E}\left[r_{t+\tau} r_t\right] - \mathbb{E}\left[r_{t+\tau}\right]\mathbb{E}\left[r_t\right] \\ &= \mathbb{E}\left[\sigma_{t+\tau}\varepsilon_{t+\tau} r_t\right] - 0^2 \\ &= \mathbb{E}\left[\mathbb{E}\left[\sigma_{t+\tau}\varepsilon_{t+\tau} r_t \mid F_{t+\tau-1}\right]\right] \\ &= \mathbb{E}\left[r_t \sigma_{t+\tau}\mathbb{E}\left[\varepsilon_{t+\tau} \mid F_{t+\tau-1}\right]\right] \\ &= 0 \end{aligned}$$

- So actually, ARCH is a type of white noise!

Isserlis' Theorem

Theorem 13.2 (Isserlis' Theorem)

If $(X_1, \dots X_N)$ *is a zero-mean multivariate normal random vector and* N *is even, then*

$$\mathbb{E}\{X_1 \dots X_N\} = \sum_{p \in P_N} \prod_{\{i,j\} \in p} \text{Cov}\{X_i, X_j\}$$

where P_N *is the set of all partitions of* $\{1, \dots, N\}$ *into pairs.*

A special case of this theorem notes that for $(X_1, \dots X_4,)$ zero-mean multivariate Gaussian:

$$\begin{aligned}\mathbb{E}(X_1X_2X_3X_4) &= \mathbb{E}\{X_1X_2\}\,\mathbb{E}\{X_3X_4\} \\ &+ \mathbb{E}\{X_1X_3\}\,\mathbb{E}\{X_2X_4\} \\ &+ \mathbb{E}\{X_1X_4\}\,\mathbb{E}\{X_2X_3\}\end{aligned}$$

ARCH: higher order moments

- Often, financial time series exhibit "fat tails", which means a kurtosis greater than 3.
- We can compute the fourth-order moments of the ARCH process

$$\begin{aligned}\mathbb{E}\left[r_t^4 \mid F_{t-1}\right] &= \mathbb{E}\left[\sigma_t^4 \varepsilon_t^4 \mid F_{t-1}\right] \\ &= \sigma_t^4 \mathbb{E}\left[\varepsilon_t^4 \mid F_{t-1}\right] \\ &= 3\sigma_t^4 \mathbb{E}\left[\varepsilon_t^2 \mid F_{t-1}\right]^2 \\ &= 3\sigma_t^4 \\ &= 3\left(\alpha_0 + \alpha_1 r_{t-1}^2\right)^2\end{aligned}$$

where Isserlis' Theorem gives the fourth moment.

ARCH: higher order moments

Again assuming shift invariance, the unconditional fourth moment is

$$\begin{aligned}\mathbb{E}\left[r_t^4\right] &= \mathbb{E}\left[\mathbb{E}\left[r_t^4 \mid F_{t-1}\right]\right] \\ &= 3\mathbb{E}\left[\left(\alpha_0 + \alpha_1 r_{t-1}^2\right)^2\right] \\ &= 3\left[\alpha_0^2 + 2\alpha_0\alpha_1\mathbb{E}\left[r_t^2\right] + \alpha_1^2\mathbb{E}\left[r_t^4\right]\right] \\ &= 3\left[\alpha_0^2 + 2\alpha_0\alpha_1\frac{\alpha_0}{1-\alpha_1} + \alpha_1^2\mathbb{E}\left[r_t^4\right]\right]\end{aligned}$$

Re-arrangement gives

$$\mathbb{E}\left[r_t^4\right] = \frac{3\alpha_0^2\left(1+\alpha_1\right)}{\left(1-\alpha_1\right)\left(1-3\alpha_1^2\right)}$$

Which is finite if we require $0 \leq \alpha_1^2 < 1/3$.

Then the kurtosis is given by

$$\frac{\mathbb{E}\left[r_t^4\right]}{\mathsf{Var}\left(r_t\right)^2} = \frac{3\alpha_0^2\left(1+\alpha_1\right)}{\left(1-\alpha_1\right)\left(1-3\alpha_1^2\right)}\frac{\left(1-\alpha_1\right)^2}{\alpha_0^2}$$
$$= 3\frac{1-\alpha_1^2}{1-3\alpha_1^2}$$

which is greater than 3 if $\alpha_1 > 0$.

The autocorrelation of the squares

- One can also show (see the exercises) that

$$\text{Corr}\left(r_{t+\tau}^2, r_t^2\right) = \alpha_1^{|\tau|}$$

- This means that we have a geometric decay in the autocovariance.
- This is sometimes fine, but we might want longer memory than this.

Parameter estimation

- The parameters of the ARCH model can be estimated using maximum likelihood estimation (MLE).
- The idea is to construct a likelihood function based on the conditional distribution of the data given the parameters.
- In general, you can write

$$L(\theta \mid r_1, \ldots, r_n) = \prod_{t=1}^{n} f(r_t | r_{t-1}, \ldots, r_1, \theta)$$

- This idea is central when we construct likelihood functions for time series models.

ARCH parameter estimation

- Remember that the arch model is

$$r_t \mid F_{t-1} \sim \mathcal{N}\left(0, \alpha_0 + \alpha_1 r_{t-1}^2\right)$$

- So we just need to deal with the base case
- We know the marginal variance, so this is fairly easy here
- Then we have the scheme

$$r_1 \sim \mathcal{N}\left(0, \frac{\alpha_0}{1-\alpha_1}\right)$$
$$r_t \mid r_{t-1}, \ldots, r_1 \sim \mathcal{N}\left(0, \alpha_0 + \alpha_1 r_{t-1}^2\right) \quad \text{for } 2 \leq t \leq n$$

So lets fit a model and see

- In this case we get $\alpha_0 = 6.663 \times 10^{-5}$, and $\alpha_1 = 0.2384$
- Lets look at a simulation from the fitted model

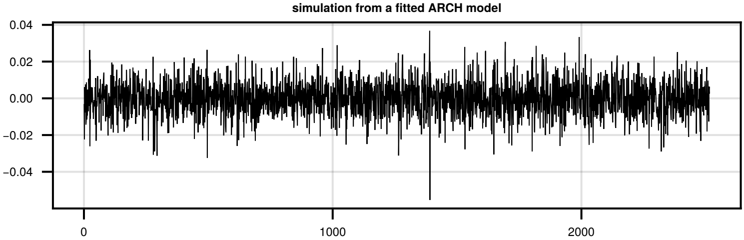

Figure: Simulated realisation of the ARCH process.

What about the ACF and PACF?

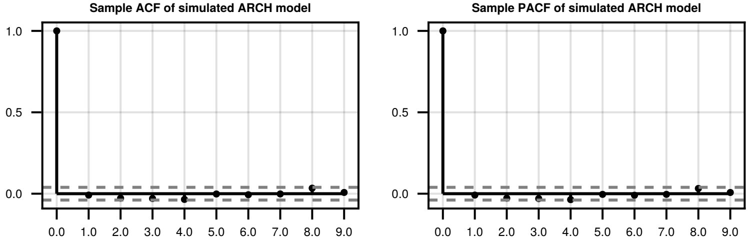

Figure: Simulated realisation of the ARCH process.

And the square?

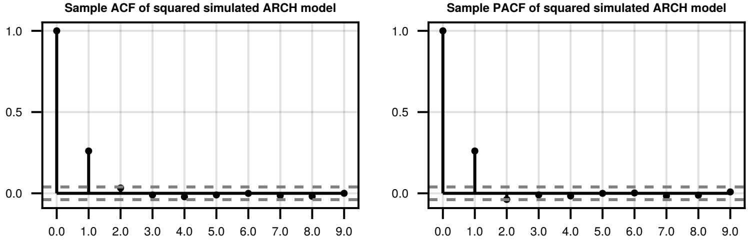

Figure: Simulated realisation of the ARCH process.

Recall the ACF and PACF of the squared log-returns

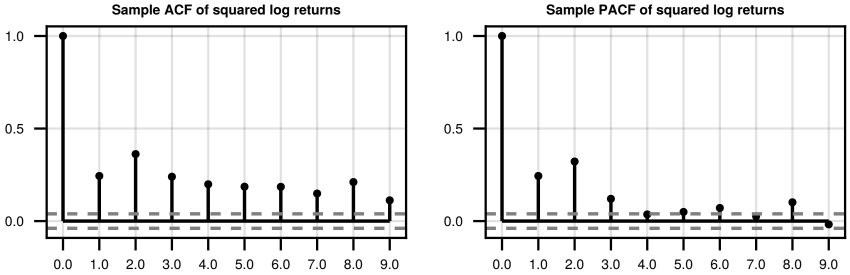

Figure: acf and pacf of the square of the log-returns from 01/01/2010 to 01/01/2020.

- They do not look similar.

What happened?

- Recall that the ARCH model is of the form

$$r_t = \sigma_t \varepsilon_t$$
$$\sigma_t = \sqrt{\alpha_0 + \alpha_1 r_{t-1}^2}$$

where ε_t is a white noise process.

- This means that if we are in a volatile period, we will have a large value of σ_t, and hence a large value of r_t on average.
- But, this does not persist if we get one small value of r_t by chance.

Solution 1: ARCH(p)

Definition 13.3 (ARCH(p) models)

We will say that a time series $\{r_t\}$ is an ARCH(p) model if it takes the form

$$r_t = \sigma_t \varepsilon_t, \qquad t \in \mathbb{Z}$$

where $\{\varepsilon_t\}$ is a white noise process with mean zero and variance one, and

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \ldots + \alpha_p r_{t-p}^2$$

is the conditional variance of the process. The parameters satisfy $\alpha_0, \alpha_p > 0$ and $\alpha_j \geq 0$ for $1 \leq j < p$.

- The idea now is that we can have a longer memory of the past.
- But, this might require a lot of parameters.

Returning to the ACF and PACF of the squared log returns

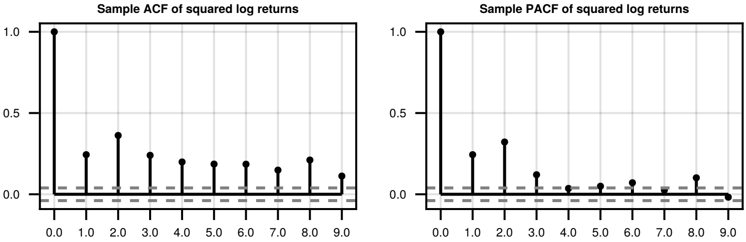

Figure: acf and pacf of the square of the log-returns from 01/01/2010 to 01/01/2020.

Solution 2: GARCH

Definition 13.4 (GARCH(p,q) models)

We will say that a time series $\{r_t\}$ is a GARCH(p,q) model if it takes the form

$$r_t = \sigma_t \varepsilon_t, \qquad t \in \mathbb{Z}$$

where $\{\varepsilon_t\}$ is a Gaussian white noise process with mean zero and variance one, and

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \ldots + \alpha_p r_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_q \sigma_{t-q}^2$$

is the conditional variance of the process. The parameters satisfy $\alpha_0, \alpha_p, \beta_q > 0$ and $\alpha_j, \beta_j \geq 0$ for $1 \leq j < p$.

- Often low order GARCH models are used, such as GARCH(1,1).
- This is far more parsimonious than the ARCH(p) model.

Properties of GARCH

- GARCH processes are also mean-zero and white noise (see the exercises)
- GARCH processes enable us to model more dependent volatility
- For a GARCH(1,1), $\alpha + \beta < 1$ implies stationarity
- Often in practice the parameters do sum to a value close to one
- We could replace the normal noise with some other iid mean-zero unit-variance noise
- For a GARCH(1,1), the fourth moment exists if and only if

$$3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$$

Proposition 13.5

A GARCH(p,q) has a weakly stationary solution if and only if

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

In this case, its variance is given by

$$\frac{\alpha_0}{1 - \sum_{j=1}^{p} \alpha_j - \sum_{j=1}^{q} \beta_j}$$

- In general, $\sum_{j=1}^{p} \alpha_j + \sum_{j=1}^{q} \beta_j < 1$ is sufficient for strict stationarity
- For certain classes of noise (including Gaussian) $\sum_{j=1}^{p} \alpha_j + \sum_{j=1}^{q} \beta_j = 1$ also has a strictly stationary solution, just not a weakly stationary one

Lets fit a GARCH(1,1) model

- In this case, we get

$$\alpha_0 = 3.686 \times 10^{-6},\ \alpha_1 = 0.1713,\ \beta_1 = 0.7897$$

- Look at a simulation, this is much better!

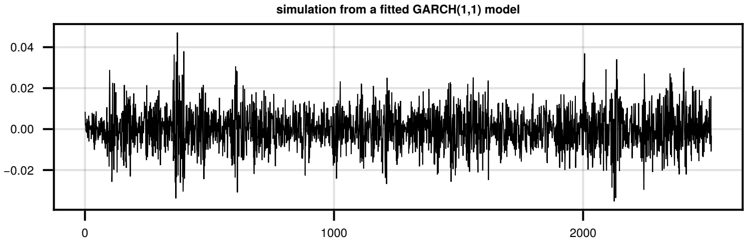

Figure: Simulated realisation of the GARCH(1,1) process.

Now lets check the ACF and PACF

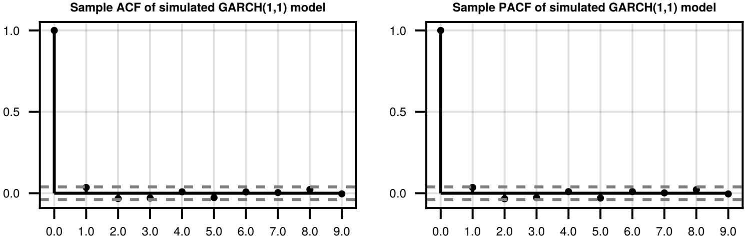

Figure: Simulated realisation of the GARCH(1,1) process.

And the square?

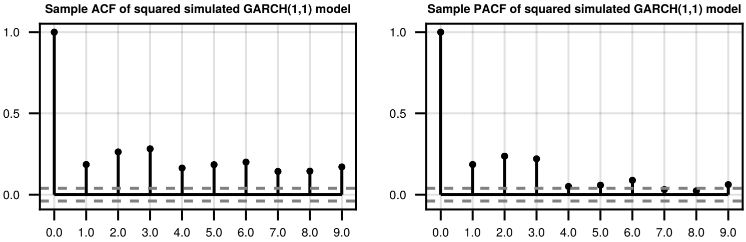

Figure: Simulated realisation of the GARCH(1,1) process.

- This looks much better!
- The ACF and PACF of the squared log-returns are now much more similar to the simulated GARCH(1,1) process.

Diagnostics

In sample prediction

- One thing we can do is look at wether the model can predict the variance of the data well.
- In particular, our model at a given time t is

$$r_t = \mathcal{N}\left(0, \sigma_t^2\right).$$

- Therefore, we could look at the confidence intervals $\pm 1.96\hat{\sigma}_t$.

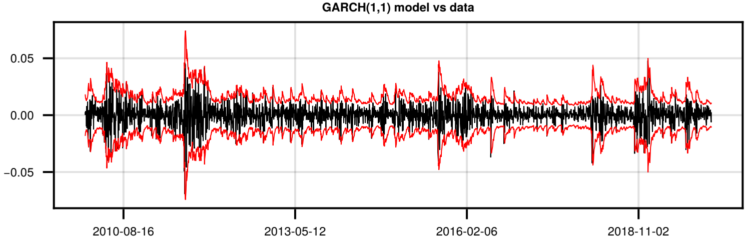

Figure: The log-returns with confidence intervals from the GARCH process.

Standardised residuals

- We can also look at the standardised residuals

$$\tilde{e}_t = \frac{r_t}{\sigma_t}.$$

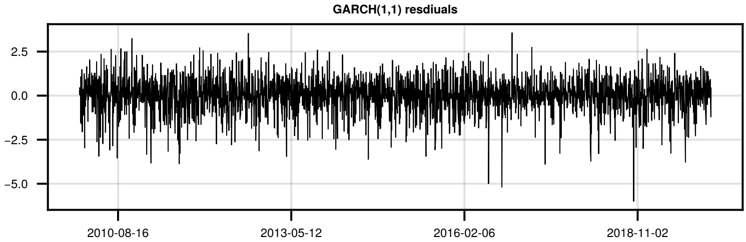

Figure: Residuals of the log-returns.

ACF and PACF of the residuals

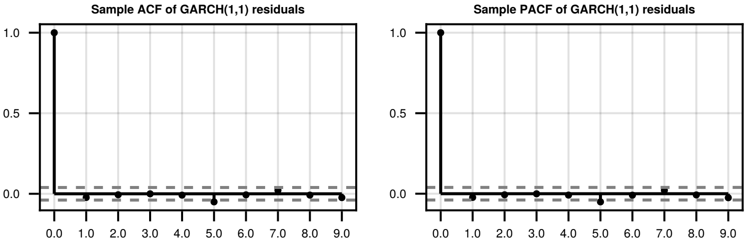

Figure: ACF and PACF of the residuals.

- We can see that the residuals are probably white noise.
- This is a good sign!

ACF and PACF of the squared residuals

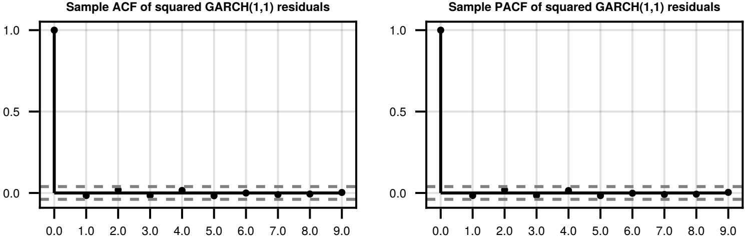

Figure: ACF and PACF of the squared residuals.

- We can see that the squared residuals are now behaving like white noise!

Residual distribution

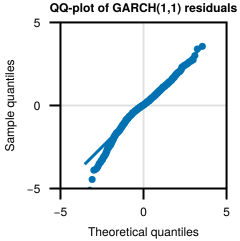

Figure: QQ plot of the residuals.

Combining GARCH and ARMA

- Since GARCH is a white noise process, we can simply make

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} - \sum_{k=0}^{q} \theta_k r_{t-k}$$

where r_t is a GARCH$(\tilde{p}, \tilde{q})$ process and the θs and ϕs satisfy the usual conditions for an ARMA(p, q).

- We just need to be careful with the likelihood, since we cannot assume iid Gaussian errors any more!

Summary

- ARCH and GARCH models can capture changes over time and volatility "clusters".
- We can model "outlier" behaviour and large shocks due to its large kurtosis.
- The GARCH model does not affect the correlation structure.
- GARCH allows us to model longer dependence in a parsimonious way.
- Positive and negative shocks affect the system in the same way.