GARCH models

Matteo Barigozzi[†]

[†] London School of Economics, UK

m.barigozzi@lse.ac.uk

CIDE–Bertinoro Course for PhD students – 2014

Outline

- In the analysis of financial data, asset equity returns are typically the main variable of interest (rather than prices)
- There are at least two reasons for this:
  1. Returns are easier to interpret
  2. Returns have statistical properties which are easier to handle (e.g. stationarity)
- There are two types of return definitions: simple and log returns

Simple returns

- Let P_t be the price of an asset at period t ($t = 1, .., T$).
- The simple return is defined as the gross return

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

- In other words, R_t is the gross return generated by holding the asset for one period
- Note that since prices are nonnegative, $R_t \in [-1, +\infty)$ (corresponding to $P_t = 0$ or $P_{t-1} = 0$)

- What happens if I have been holding the asset for k periods?
- The multiperiod simple return is given by

$$R_{t-k:t} = \frac{P_t - P_{t-k}}{P_{t-k}}$$

- It is straightforward to check that

$$\begin{aligned} 1 + R_{t-k:t} &= \frac{P_t}{P_{t-k}} \\ &= \frac{P_t}{P_{t-1}} \frac{P_{t-1}}{P_{t-2}} \dots \frac{P_{t-k+1}}{P_{t-k}} \\ &= (1 + R_t)(1 + R_{t-1}) \dots (1 + R_{t-k}) \\ &= \prod_{j=0}^{k} (1 + R_{t-j}) \end{aligned}$$

- Thus, the multiperiod return can be expressed as a product of single period returns

- It is often of interest to know what is the average return on an asset after holding the asset from $t = 1$ to T

$$\overline{R}_T = \left[\prod_{t=1}^{T} (1 + R_t) \right]^{1/T} - 1$$

that is, the geometric mean of $1 + R_t$ (minus 1)

- The geometric mean is the correct average in the sense that it produces the constant rate of return that would give the same multiperiod return from period 1 to T

Log returns

- Simple returns are a natural way of measuring the variation of the value of an asset, however, it is more common to work with log returns
- Log returns are defined as

$$\epsilon_t = \log P_t - \log P_{t-1} = \log \frac{P_t}{P_{t-1}} = \log(1 + R_t)$$

- Note that $\epsilon_t \in (-\infty, +\infty)$, thus it is possible to have negative returns lower than -100%

- One of the advantages of the log returns is that the multiperiod log return

$$\epsilon_{t-k:t} = \log P_t - \log P_{t-k}$$

can be simply expressed as a sum of one period returns

$$\epsilon_{t-k:t} = \sum_{i=0}^{k-1} \epsilon_{t-i} = \sum_{i=0}^{k-1} \log P_t - \log P_{t-1-i} = \log P_t - \log P_{t-k}$$

- Are Log and Simple Returns that different?

- Not really. If price variations are small, then log and simple returns are very close to each other. This can be seen via a Taylor expansion. If R_t is close to 0, then

$$\epsilon_t = \log(1 + R_t) \approx \log 1 + \left.\frac{1}{1 + R_t}\right|_{R_t=0} R_t = R_t$$

- In practice, we will work almost exclusively with log returns
- It is a good habit to multiply returns by 100 to express them as a percentage. Some statistical packages are sensitive to the scale of the data. Since log differences can be small, sometimes this creates numerical difficulties
- Thus, we are going to work with returns defined as

$$\epsilon_t = 100 \times (\log P_t - \log P_{t-1})$$

Portfolios

- A portfolio p is defined as a collection of N assets and portfolio weights $w_1, ..., w_N$ reflecting the percentage of wealth invested N assets is defined as a weigheted average of the simple returns with weights equal to the percentage of

$$R_{p\,t} = \sum_{i=1}^{N} w_i R_{i\,t}$$

- Unfortunatelly, this doesn't hold for log returns. However, if the magnitude of the returns is small then the approximation error is moderate

$$\epsilon_{p\,t} \approx \sum_{i=1}^{N} w_i \epsilon_{i\,t}$$

In practice, in portfolio analysis this approximation is used often

Which kind of financial assets are we going to look at?

- (US) Common Stocks: IBM, Apple, General Motors, Goldman Sachs, ...
- Stock Indices: S&P 500 index, Dow Jones Industrial average, FTSE 100, CAC 40, DAX, ...
- ETFs: Index ETFs, Commodity ETFs, ...
  Exchange Traded Funds are investment fund which can be traded on an exchange. ETFs are designed to track the performance of specific types of investment such as investing on an index, a country, a commodity and so forth
- ...but the models considered are also useful for: exchange rates, inflation, interest rates

Which horizon?

- So far we haven't specified the length of a period
- In practice, the choice of the frequency depends on the context of the application
  From an industry perspective, it is reasonable assume that different agents are focused on different horizons
  - Fund Manager: Monthly
  - Risk Manager: Daily
  - Traders: Intra–Daily
- We are going to work mostly with daily and monthly data

Which price?

- For each time period, there are typically multiple prices available
- For instance, at a daily frequency, data providers typically report: Opening, Low (min), High (max) and Closing Prices
- Returns are computed using the last price available in each period (i.e. the closing)

- Stock prices are affected by Stock Splits and Dividends
- It is preferable to work with “Adjusted” series which take care of these features of the data
- Luckily, data providers take care of adjusting the series for us
  We only need to make sure we are working with an adjusted series

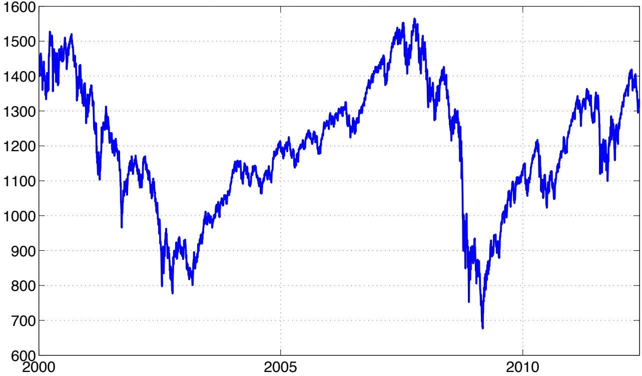
Databases

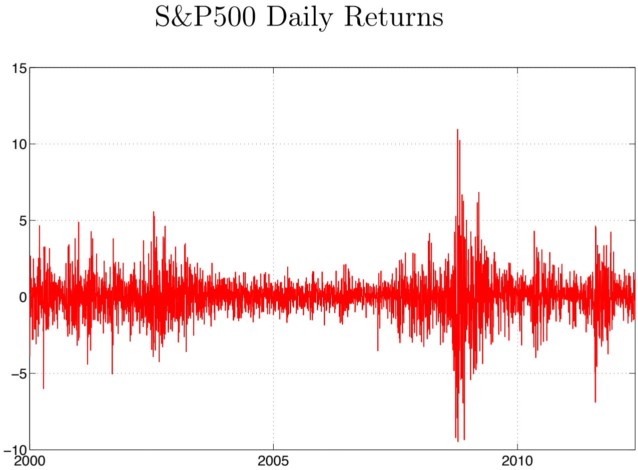
- For daily and monthly data, some of the most used (commercial) databases are the one of the Center of Research in Security Prices (aka CRSP) at U. Chicago and Datastream.
  These are commercial DBs.

- Google and Yahoo also provide daily and monthly series (for free!). We are going to rely on these sources to get some sample series for practice.

- For illustration purposes we make use of the Standard & Poor's 500 index
- The S&P500, is a stock market index based on the market capitalizations of 500 large companies having common stock listed on the NYSE or NASDAQ. The components and their weightings are determined by S&P Dow Jones Indices. It is one of the most commonly followed equity indices, and many consider it one of the best representations of the U.S. stock market, and a bellwether for the U.S. economy.
- Daily series: January 2000 - May 2012 (3123 Obs.)

S&P500 Daily Adj. Closing Prices



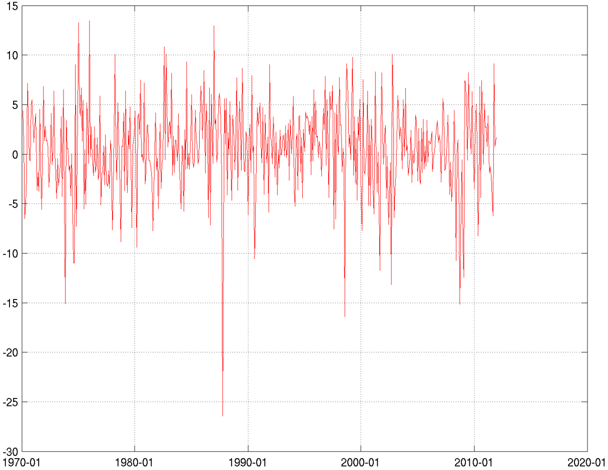

S&P500 Daily Returns
15
10
5
0
−5
−10
2000
2005
2010

A similar series worth looking at is also

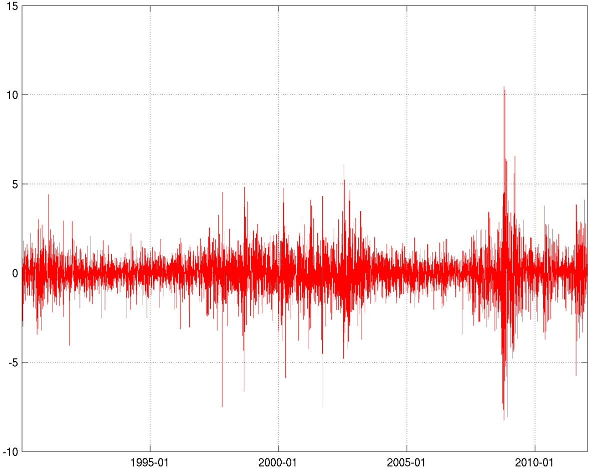
- The Dow Jones Industrial Average which is an average of 30 major US public companies.
  It is one of the oldest market indices

- Daily series: January 1990 - December 2011 (5555 Obs.)

Dow Jones Monthly Returns



Dow Jones Daily Returns



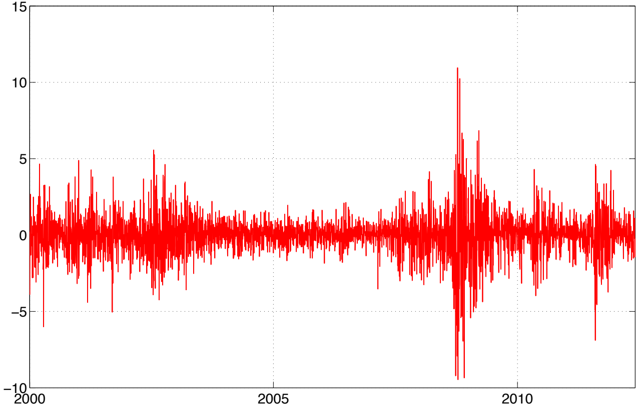
Outline

Mandelbrot (1956)

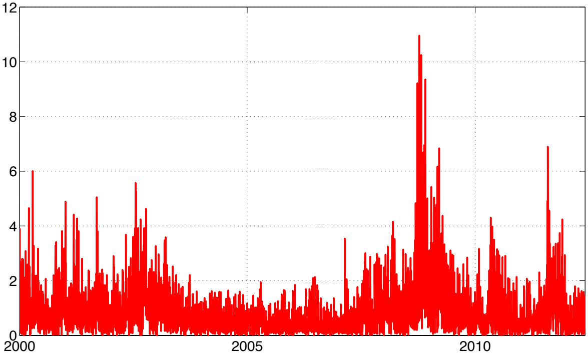
- Nonstationarity of prices P_t (random walk), stationarity of returns
- Absence of autocorrelation of returns ϵ_t (white noise)
- Autocorrelation of squared returns ϵ_t^2 or $|\epsilon_t|$ (weak white noise)
- Volatility clustering large returns (in absolute value) tend to be followed by large returns (in absolute value), and vice versa
- Fat–tailed distribution of returns, kurtosis $>> 3$ (leptokurtic), i.e. non–Gaussian
- Leverage effects, i.e. negative returns (decrease in prices) tend to increase volatility by a larger amount than positive returns

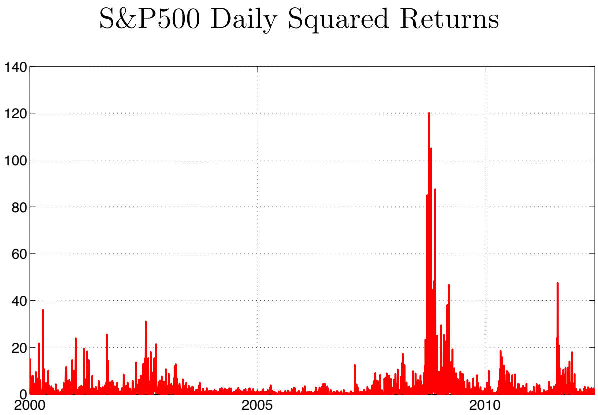
- These stylized facts have been documented starting from at least the 1960's but the first models able to capture volatility clustering were proposed starting from the 1980's
- We are going to analyse volatility clustering and introduce nonlinear dynamic models able to capture it

S&P500 Daily Returns



S&P500 Daily Absolute Returns

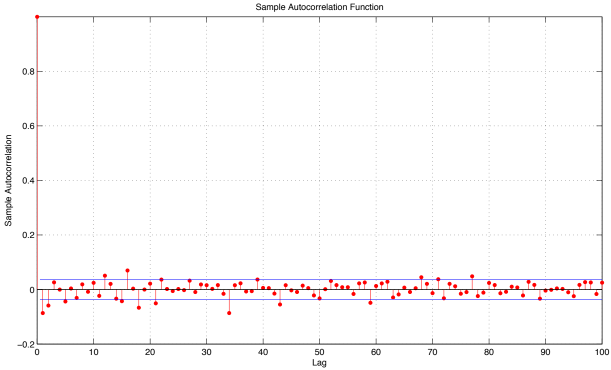



S&P500 Daily Squared Returns
140
120
100
80
60
40
20
0
2000
2005
2010

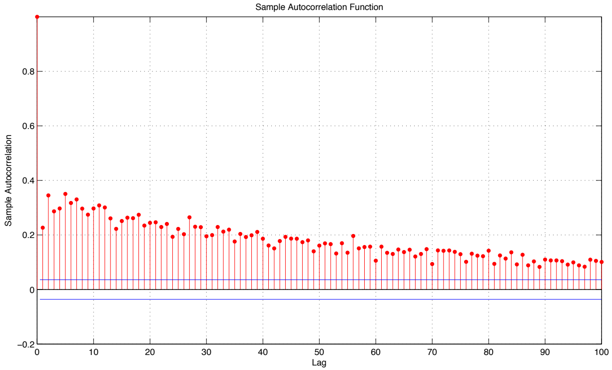
The inspection of the daily return time series plot suggests that:

- returns appear to have weak or no serial dependence
- absolute or squared returns appear to have strong serial dependence

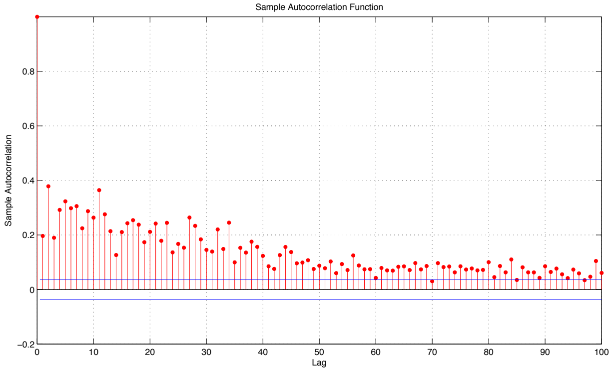
S&P500 Daily Returns ACF



S&P500 Daily Absolute Returns ACF



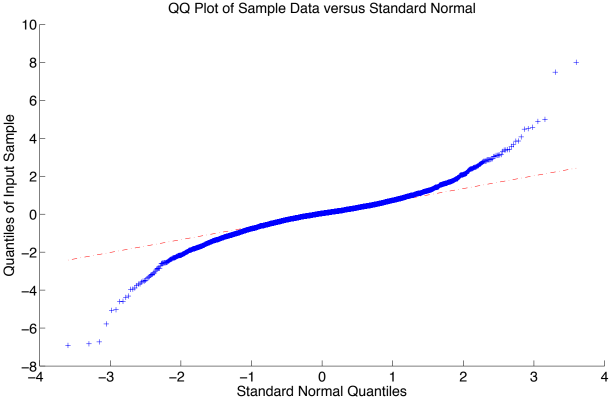
S&P500 Daily Squared Returns ACF



The inspection of the autcorrelograms suggests that:

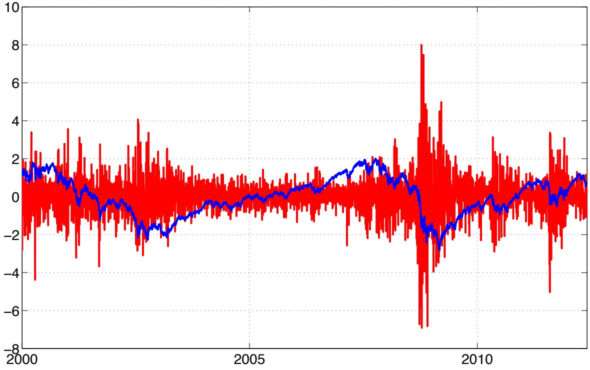
- returns appear to have weak or no serial dependence
- absolute and square returns appear to have strong serial dependence

S&P500 Daily quantiles vs. Gaussian quantiles



Evidence on non–Gaussianity and fat tails

S&P500 Daily returns and prices



Evidence of leverage, decreasing price implies higher volatility

Outline

- The strong evidence of serial dependence in absolute and square returns suggest that the scale of returns changes in time
- In other words, the variance of the process is time varying
- In order to capture volatility clustering, we need to introduce appropriate time series processes able to model this behavior

General model

- Consider the covariance stationary time series $\{y_t\}_{t=1}^T$, i.e. with $\mathsf{E}[y_t] = \mu$ and $\mathsf{Var}[y_t] = \sigma^2$ not depending on time

- We define conditional mean μ_t of the process as

$$\mu_t = \mathsf{E}[y_t | \mathcal{I}_{t-1}]$$

and conditional variance σ_t^2 as

$$\sigma_t^2 = \mathsf{Var}[y_t | \mathcal{I}_{t-1}] = \mathsf{E}[(y_t - \mu_t)^2 | \mathcal{I}_{t-1}]$$

where $\mathcal{I}_t = \{y_1, y_2, ..., y_t\}$ is the information set available at time t

- The shorthand notation $\mathsf{E}_{t-1}[\cdot]$ in place of $\mathsf{E}[\cdot | \mathcal{I}_{t-1}]$ is used sometime

Can we use ARMA models?

- Consider the simple ARMA(1,1) model

$$y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t - \theta_1 \epsilon_{t-1} \quad \epsilon_t \sim D(0, \sigma_\epsilon^2)$$

- For the simple ARMA(1,1) we have that

$$\mu_t = \phi_0 + \phi_1 y_{t-1} - \theta_1 \epsilon_{t-1}$$

and

$$\sigma_t^2 = \sigma_\epsilon^2$$

- Thus, while the conditional mean of an ARMA is time varying, the conditional variance of an ARMA is constant. An ARMA(p, q) is not able to capture time varying volatility

In general we have the following model (for simplicity consider the case $\mu_t = 0$, i.e. of zero mean and zero autocorrelation in y_t)

- conditional heteroskedasticity: $\sigma_t^2 \neq const.$ and

$$y_t \equiv \epsilon_t = \sigma_t z_t$$

- the volatility $\sigma_t > 0$ is a function of $\mathcal{I}_{t-1}$
- the innovations z_t are i.i.d. with $\mathsf{E}[z_t] = 0$ and $\mathsf{E}[z_t^2] = 1$
- innovations are independent of past returns, i.e. $\mathsf{Cov}(z_t \sigma_t) = 0$ and $\mathsf{Cov}(z_t \epsilon_{t-k}) = 0$ for $k > 0$

Moments of conditional heteroskedastic models

- mean of returns $\mathsf{E}[\epsilon_t] = \mathsf{E}[\sigma_t z_t] = \mathsf{E}[\sigma_t]\mathsf{E}[z_t] = 0$
- conditional mean of returns $\mathsf{E}_{t-1}[\epsilon_t] = 0$
- conditional variance of returns

$$\mathsf{E}_{t-1}[\epsilon_t^2] = \mathsf{E}_{t-1}[\sigma_t^2 z_t^2] = \mathsf{E}_{t-1}[\sigma_t^2]\mathsf{E}_{t-1}[z_t^2] = \sigma_t^2$$

- autocovariance of returns

$$\begin{aligned} \mathsf{Cov}(\epsilon_t \epsilon_{t-k}) &= \mathsf{E}[\epsilon_t \epsilon_{t-k}] - \mathsf{E}[\epsilon_t]\mathsf{E}[\epsilon_{t-k}] = \mathsf{E}[\epsilon_t \epsilon_{t-k}] = \\ &= \mathsf{E}[\sigma_t z_t \epsilon_{t-k}] = \mathsf{E}[z_t]\mathsf{E}[\sigma_t \epsilon_{t-k}] = 0 \end{aligned}$$

  thus ϵ_t is weak white noise but in general $\mathsf{E}[\epsilon_t^2 \epsilon_{t-k}^2] \neq 0$ thus not strong white noise
- kurtosis $\kappa_\epsilon = \kappa_z \left[1 + \frac{\mathsf{Var}(\sigma_t^2)}{(\mathsf{E}[\sigma_t^2])^2}\right]$, leptokurtic distribution of z_t

- In order to model volatility, the literature has proposed specific types of time series models
- There are two approaches in modelling the conditional variance σ_t^2:
  1. ARCH Approach: σ_t^2 is a deterministic equation
  2. Stochastic Volatility Approach: σ_t^2 is a stochastic equation
- In practice, the ARCH approach is more popular while Stochastic Volatility models are typically harder to work with

The ARCH Model

- In order to capture volatility clustering, in 1982 Robert Engle proposed the AutoRegressive Conditional Heteroskedasticity (ARCH) model
- This simple model has started the literature on nonlinear quantitative modelling of financial time series
- In 2003 Robert Engle won the Nobel prize for economics
  The ARCH model was mentioned as one of his most significant contributions

- The ARCH(1) model is

$$\epsilon_t = \sqrt{\sigma_t^2} z_t \quad z_t \sim D(0,1)$$

where D is a distribution with mean 0 and variance 1 and

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2$$

where $\omega > 0$ and $\alpha \geq 0$.

- The α coefficient needs to satisfy other regularities conditions to ensure that the process is “well behaved” (i.e. finite variance and stationarity)
- In words, the current conditional variance of returns is proportional to the past squared return

- The ARCH(q) model is

$$\epsilon_t = \sqrt{\sigma_t^2} z_t \quad z_t \sim D(0,1)$$

where D is a distribution with mean 0 and variance 1 and

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + ... + \alpha_q \epsilon_{t-q}^2$$

where $\omega > 0$ and $\alpha_i \geq 0$. (Again, the α_i coefficient needs to satisfy other regularities conditions to ensure that the process is "well behaved")

- In the early 1980's, the simple idea of making the current conditional variance of the process a deterministic function of the past history of the process opened the door to a new way of modelling time series
- Since the original contribution of Engle, a vast number of alternative specifications for the conditional variance of returns have been proposed

ARCH(q) is equivalent to an AR(q) for the squared returns (if the model is stationary)

- Define the innovations of squared returns

$$\nu_t = \epsilon_t^2 - \mathsf{E}_{t-1}[\epsilon_t^2] = \epsilon_t^2 - \sigma_t^2$$

- then $\sigma_t^2 = \epsilon_t^2 - \nu_t$ and the ARCH(q) becomes

$$\epsilon_t^2 - \nu_t = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2$$

- which is an AR(q)

$$\epsilon_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \nu_t$$

but ν_t is not i.i.d. as $\mathsf{Cov}(\epsilon_t^2 \epsilon_{t-k}^2) \neq 0$.

ACF of squares

- The AR representation allows for computing the ACF of ϵ_t^2
- For an AR(1) we have

$$y_t = ay_{t-1} + e_t \quad \Rightarrow \quad \rho(h) = \frac{\mathsf{Cov}(y_t, y_{t-h})}{\mathsf{Var}(y_t)} = a^h$$

where $e_t \sim N(0,1)$ and $\mathsf{E}[y_t] = 0$

- An ARCH(1) is equivalent to an AR(1) for the squared returns

$$\rho(h) = \frac{\mathsf{Cov}(\epsilon_t^2, \epsilon_{t-h}^2)}{\mathsf{Var}(\epsilon_t^2)} = \alpha^h$$

- Notice that $\rho(h) > 0$ for any h

- For higher order ARCH we have to solve Yule Walker difference equations

$$\rho(h) = \sum_{i=1}^{q} \alpha_i \rho(h-i)$$

- For an ARCH(2) we have

$$\begin{aligned} \rho(2) &= \alpha_1 \rho(1) + \alpha_2 \rho(0) \\ \rho(1) &= \alpha_1 \rho(0) + \alpha_2 \rho(-1) = \alpha_1 \rho(0) + \alpha_2 \rho(1) \end{aligned}$$

Then

$$\frac{\rho(2)}{\rho(1)} = \alpha_1 + \alpha_2 \frac{\rho(0)}{\rho(1)} = \alpha_1 + \alpha_2 \frac{1-\alpha_2}{\alpha_1}$$

- Thus $\rho(2) < \rho(1)$ implies a restriction on the coefficients and in general ACF do not decrease to zero monotonically

Detecting ARCH effects

- Engle (1982) introduces a simple test to detect the presence of ARCH effects using the AR representation

- The ARCH-LM test is constructed as follows
  1. Estimate the coefficients of the following autoregression by OLS
  $$\epsilon_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + ... + \alpha_q \epsilon_{t-q}^2 + u_t$$
  2. The null of no ARCH effects is formulated as $H_0 : \alpha_1 = 0, \alpha_2 = 0, ..., \alpha_q = 0$
  3. The test statistic for the ARCH-LM is
  $$T \cdot R^2$$
  where R^2 is the usual "R-squared" coefficient. Under the null of no arch effects the test statistic is asymptotically distributed as a χ_q^2

Estimating ARCH

- ARCH models are typically estimated by Maximum Likelihood
- The ML estimator has no closed form expression and needs to be found numerically
- This will be the topic of the next lectures

Residuals' diagnostics

- Model adequacy can be checked by inspection of the so called "standardized residuals" defined as

$$\hat{z}_t = \frac{\epsilon_t}{\hat{\sigma}_t} = \frac{\epsilon_t}{\sqrt{\hat{\omega} + \sum_{i=1}^{q} \hat{\alpha}_i \epsilon_{t-i}^2}}$$

  where $\hat{\omega}$ and $\hat{\alpha}_i$ are obtained by ML under distribution D
- If the specification is correct, standardised residuals should be
  1. Approximately distributed according to distribution D
  2. Should not exhibit dependence in levels, absolute levels, square levels, etc... i.e. the must be i.i.d.

Forecasting

- Forecast formulas of the variance of the ARCH(1) are analogous to those of the AR(1)

- 1–step ahead forecast of the variance is

$$\sigma^2_{t+1|t} = \mathsf{E}_t[\sigma^2_{t+1}] = \mathsf{E}_t[\omega + \alpha\epsilon^2_t] = \omega + \alpha\epsilon^2_t$$

- 2–steps ahead forecast of the variance is

$$\begin{aligned}\sigma^2_{t+2|t} &= \mathsf{E}_t[\omega + \alpha\epsilon^2_{t+1}] = \omega + \alpha\mathsf{E}_t[\epsilon^2_{t+1}] = \omega + \alpha\sigma^2_{t+1}\\ &= \omega + \alpha(\omega + \alpha\epsilon^2_t) = \omega(1+\alpha) + \alpha^2\epsilon^2_t\end{aligned}$$

- k–steps ahead forecast of the variance is

$$\begin{aligned}\sigma^2_{t+k|t} &= \mathsf{E}_t[\omega + \alpha\epsilon^2_{t+k-1}]\\ &= \omega\left(\sum_{i=0}^{k-1}\alpha^i\right) + \alpha^k\epsilon^2_t\end{aligned}$$

S&P500 Daily Volatility Analysis

- As an empirical illustration, we are going to use an ARCH(3) model (with intercept, i.e. $\mu_t = c$) to fit daily returns

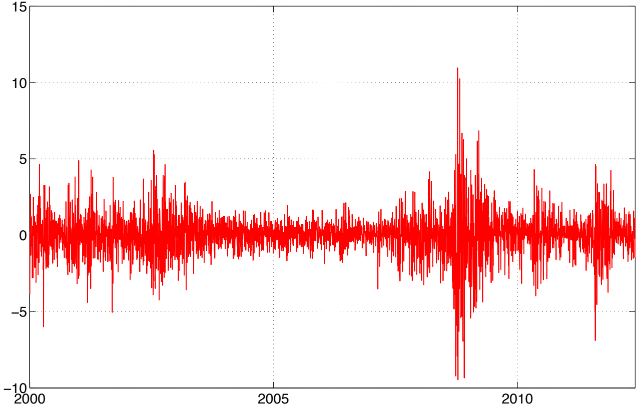
- The ARCH(3) with intercept is defined as

$$y_t = c + \epsilon_t = c + \sqrt{\sigma_t^2} z_t \quad z_t \sim N(0,1)$$

where the variance equation is defined as

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \alpha_3 \epsilon_{t-3}^2$$

S&P500 Daily Returns

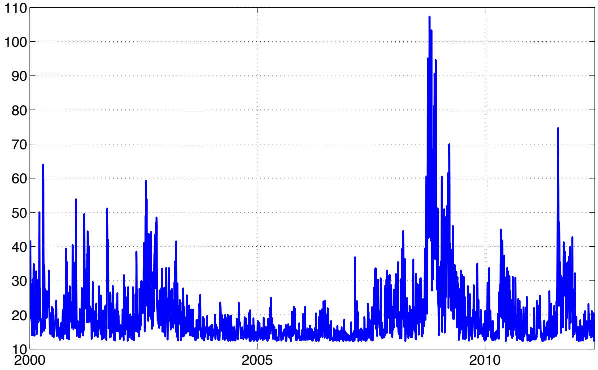


ARCH-LM Stat (3 lags): 516.694 - p–value = 0.000

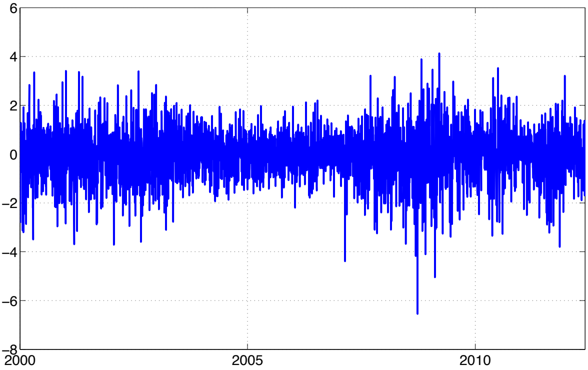
Estimates

	param	se	t-stat
c	0.01	0.020	0.5
ω	0.60	0.014	42.9
α_1	0.12	0.014	8.6
α_2	0.37	0.019	19.5
α_3	0.23	0.021	11.0

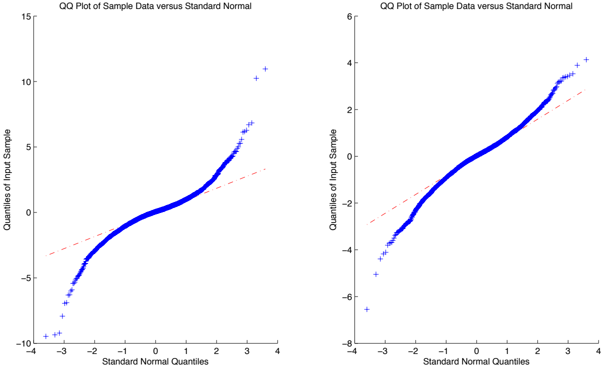
S&P500 Annualized Volatility $\sqrt{252\hat{\sigma}_t^2}$



SP&P500 Std Residuals

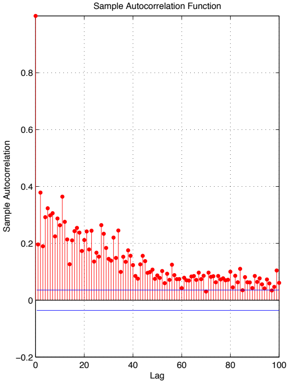


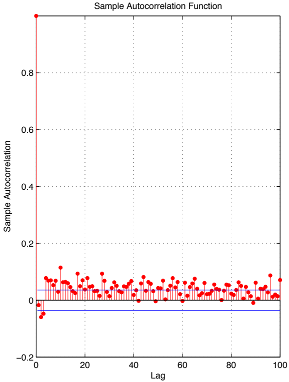
S&P500 Daily Returns (left) vs. Std. Residual (right) QQ-plot



Jarque-Bera Test Before & After: 6746 & 631

S&P500 Daily Squared Returns (left) & Squared Std. Residual (right) ACF





- S&P500 exhibits strong evidence of volatility clustering
- The ARCH(3) model captures a substantial portion of clustering
- However, residual diagnostic signal that the specification is not fully satisfactory

- In practice, only rather rich ARCH parameterizations are able to fit financial series adequately
- However, largely parameterized models can be unstable in forecasting and a hard to estimate
- In order to overcome the shortcomings of the ARCH, Tim Bollerslev proposed a generalisation of the ARCH model called GARCH (Bollerslev, 1986)
- The model allows to fit financial returns adequately while keeping the number of parameters small
- In practice, the GARCH model is one of the most successfully employed volatility models

- The GARCH(1,1) model is

$$\epsilon_t = \sqrt{\sigma_t^2} z_t \quad z_t \sim D(0,1)$$

where D is a distribution with mean 0 and variance 1 and

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

where $\omega > 0$, $\alpha \geq 0$ and $\beta > 0$ and $\alpha + \beta < 1$ (in order to have stationarity, see next)

- The GARCH(p, q) model is

$$\epsilon_t = \sqrt{\sigma_t^2} z_t \quad z_t \sim D(0, 1)$$

where D is a distribution with mean 0 and variance 1 and

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + ... + \alpha_q \epsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + ... + \beta_q \sigma_{t-q}^2$$

where $\omega > 0$, $\alpha_i \geq 0$ and $\beta_i > 0$ and $\sum \alpha_i + \sum \beta_i < 1$ (in order to have stationarity, see next)

Stationarity

- A necessary condition for weak stationarity of GARCH (1,1) is

$$\alpha + \beta < 1 \tag{1}$$

- We have $\mathsf{E}[\epsilon_t^2] = \mathsf{E}[\mathsf{E}_{t-1}[\epsilon_t^2]] = \mathsf{E}[\sigma_t^2]$ which because of stationarity does not depend on t

- Taking expectations in a GARCH(1,1) we have

$$\mathsf{E}[\epsilon_t^2] = \omega + \alpha\mathsf{E}[\epsilon_t^2] + \beta\mathsf{E}[\epsilon_t^2] \quad \Rightarrow \quad \mathsf{E}[\epsilon_t^2](1-\alpha-\beta) = \omega$$

- In order to have $\omega > 0$ we need condition (1)

- Condition (1) is also a sufficient condition for weak stationarity

- A necessary condition for strong stationarity of GARCH(1,1) is

$$-\infty \leq \mathsf{E}[\log(\alpha z_t^2 + \beta)] < 0$$

- This implies $\beta < 1$

- If $\alpha + \beta < 1$ we have strict stationarity, indeed

$$\mathsf{E}[\log(\alpha z_t^2 + \beta)] \leq \log \mathsf{E}[\alpha z_t^2 + \beta] = \log(\alpha + \beta) < 0$$

- Thus $\alpha + \beta < 1$ is a sufficient condition for weak and strong stationarity

- For a GARCH(p, q) condition (1) generalizes to

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

Unconditional variance

- If we have $\alpha + \beta < 1$ (sufficient condition) then we have a stationary GARCH(1,1)
- By the law of iterated expectations we have $\mathsf{E}[\epsilon_t^2] = \mathsf{E}[\mathsf{E}_{t-1}[\epsilon_t^2]] = \mathsf{E}[\sigma_t^2]$
- Taking expectations in a GARCH(1,1) we have

$$\mathsf{E}[\epsilon_t^2] = \omega + \alpha \mathsf{E}[\epsilon_t^2] + \beta \mathsf{E}[\epsilon_t^2]$$

- The unconditional variance is

$$\mathsf{E}[\epsilon_t^2] = \sigma^2 = \frac{\omega}{1 - \alpha - \beta} > 0 \quad \text{if } \omega > 0$$

- For a GARCH(p, q) we have

$$\mathsf{E}[\epsilon_t^2] = \sigma^2 = \frac{\omega}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j}$$

Kurtosis

- The kurtosis coefficient is defined as

$$\kappa_\epsilon = \frac{\mathsf{E}[\epsilon_t^4]}{\left\{\mathsf{E}[\epsilon_t^2]\right\}^2} = \frac{\mathsf{E}[\mathsf{E}_{t-1}[\epsilon_t^4]]}{\left\{\mathsf{E}[\mathsf{E}_{t-1}[\epsilon_t^2]]\right\}^2} = \frac{\mathsf{E}[\sigma_t^4]}{\left\{\mathsf{E}[\sigma_t^2]\right\}^2}\kappa_z$$

  where $\kappa_z = \mathsf{E}[z_t^4]$ and if $z_t \sim N(0,1)$ then $\kappa_z = 3$

- Since $\mathsf{E}[\sigma_t^4] = \mathsf{Var}[\sigma_t^2] + \left\{\mathsf{E}[\sigma_t^2]\right\}^2$ then $\kappa_\epsilon \geq \kappa_z$

- Thus even if $z_t \sim N(0,1)$ we have $\kappa_\epsilon \geq 3$

- If there are no ARCH effects, i.e. σ_t^2 is constant, then $\mathsf{Var}[\sigma_t^2] = 0$ and $\kappa_\epsilon = \kappa_z$

- Kurtosis for ARCH(1)

$$\begin{aligned}\mathsf{E}[\sigma_t^4] &= \omega^2 + \alpha^2\mathsf{E}[\epsilon_{t-1}^4] + 2\omega\alpha\mathsf{E}[\sigma_{t-1}^2] \\ &= \omega^2 + \alpha^2\mathsf{E}[\sigma_{t-1}^4]\kappa_z + 2\omega\alpha\mathsf{E}[\sigma_{t-1}^2]\end{aligned}$$

which implies

$$\mathsf{E}[\sigma_t^4] = \frac{\omega^2 + 2\omega\alpha\mathsf{E}[\sigma_{t-1}^2]}{1-\alpha^2\kappa_z}$$

- Then

$$\kappa_\epsilon = \frac{\mathsf{E}[\sigma_t^4]}{\left\{\mathsf{E}[\sigma_t^2]\right\}^2}\kappa_z = \frac{\omega^2 + 2\omega\alpha\mathsf{E}[\sigma_{t-1}^2]}{(1-\alpha^2\kappa_z)\left\{\mathsf{E}[\sigma_t^2]\right\}^2}\kappa_z$$

- Using $\mathsf{E}[\sigma_t^2] = \mathsf{E}[\epsilon_t^2] = \sigma^2 = \omega/(1-\alpha)$ we get

$$\kappa_\epsilon = \frac{1-\alpha^2}{1-\alpha^2\kappa_z}\kappa_z$$

which for the Gaussian case is defined for $0 \le \alpha^2 < 1/3$

- Kurtosis for GARCH(1,1) is

$$\kappa_\epsilon = \frac{1-(\alpha+\beta)^2}{1-(\alpha+\beta)^2-\alpha^2(\kappa_z-1)}\kappa_z$$

- It increases with κ_z and when the coefficients approach the zone of non–stationarity

- The excess kurtosis measures deviations from Gaussianity

$$\kappa_\epsilon^* = \kappa_\epsilon - 3$$

Alternative parameterization: Exponential Weighted Moving Average

- The GARCH model can be seen as an infinite ARCH models
- for GARCH(1,1) we have

$$\begin{aligned}\sigma_t^2 &= \omega + \alpha\epsilon_{t-1}^2 + \beta[\omega + \alpha\epsilon_{t-2}^2 + \beta(\omega + \alpha\epsilon_{t-3}^2 + \beta\sigma_{t-3}^2)] \\ &= \omega\left(\sum_{i=0}^{\infty}\beta^i\right) + \alpha\sum_{i=0}^{\infty}\beta^i\epsilon_{t-i-1}^2 \\ &= \frac{\omega}{1-\beta} + \alpha\sum_{i=0}^{\infty}\beta^i\epsilon_{t-i-1}^2\end{aligned}$$

- This representation shows how a GARCH(1,1) is a parsimonious way of characterizing ARCH dynamics

- The conditional variance of a GARCH(1,1) can be seen as a weighted average of recent returns such that the weight given to past information decreases exponentially fast ($\beta < 1$)

Alternative parameterization: ARMA

- GARCH models can also be represented as an ARMA models by defining the innovations $\nu_t = \epsilon_t^2 - \sigma_t^2$

- in case of the GARCH(1,1) we get:

$$\epsilon_t^2 = \omega + (\alpha + \beta)\epsilon_{t-1}^2 + \nu_t - \beta\nu_{t-1}$$

- in case of the GARCH(p, q) we get:

$$\epsilon_t^2 = \omega + \sum_{i=1}^{r}(\alpha_i + \beta_i)\epsilon_{t-i}^2 + \nu_t - \sum_{j=1}^{p}\beta_j\nu_{t-j}$$

  where $r = \max(p, q)$

- Thus, a GARCH can be seen as an ARMA model for squared returns but ν_t is not white noise

IGARCH

- If $\alpha + \beta = 1$ we have an Integrated GARCH(p, q) or IGARCH(p, q)
- Its ARMA representation is

$$\epsilon_t^2 = \omega + \epsilon_{t-1}^2 + \nu_t - \beta\nu_{t-1}$$

thus it has a unit root

- The term $\alpha + \beta$ measures the "persistence" of the process ϵ_t^2
- But while ARIMA have not a stationary solution an IGARCH can admit (under some conditions) a strictly stationary solution

ACF of squares

- The ARMA representation allows for computing the ACF of ϵ_t^2
- A GARCH(1,1) is equivalent to an ARMA(1,1) for the squared returns

$$\rho(h) = \frac{\mathsf{Cov}(\epsilon_t^2, \epsilon_{t-h}^2)}{\mathsf{Var}(\epsilon_t^2)} = \rho(1)(\alpha + \beta)^{h-1}$$

where

$$\rho(1) = \frac{\alpha[1 - \beta(\alpha + \beta)]}{1 - (\alpha + \beta)^2 + \alpha^2}$$

- Notice that $\rho(h) > 0$ always and in this case ACF are decreasing monotonically

Forecasting

- 1–step ahead forecast of the variance is

$$\sigma^2_{t+1|t} = \mathsf{E}_t[\sigma^2_{t+1}] = \omega + \alpha\epsilon^2_t + \beta\sigma^2_t$$

- 2–steps ahead forecast of the variance is

$$\begin{aligned} \sigma^2_{t+2|t} &= \mathsf{E}_t[\sigma^2_{t+2}] = \omega + \alpha\mathsf{E}_t[\epsilon^2_{t+1}] + \beta\sigma^2_{t+1|t} \\ &= \omega + \alpha\sigma^2_{t+1|t} + \beta(\omega + \alpha\epsilon^2_t + \beta\sigma^2_t) \\ &= \omega + (\alpha+\beta)\sigma^2_{t+1|t} = \sigma^2 + (\alpha+\beta)(\sigma^2_{t+1|t} - \sigma^2) \end{aligned}$$

since $\omega = \sigma^2 - \sigma^2(\alpha+\beta)$

- k–steps ahead forecast of the variance is

$$\sigma^2_{t+k|t} = \sigma^2 + (\alpha+\beta)^{k-1}(\sigma^2_{t+1|t} - \sigma^2)$$

- since $\alpha + \beta < 1$, as $k \to \infty$, we have $\sigma^2_{t+k|t} \to \sigma^2$

- As an empirical illustration, we are going to use a GARCH(1,1) model (with intercept i.e. $\mu_t = c$) to fit daily returns

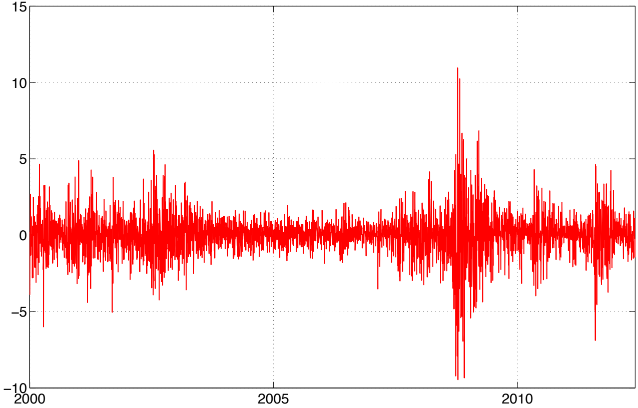
- The GARCH(1,1) with intercept is defined as

$$y_t = c + \epsilon_t = c + \sqrt{\sigma_t^2} z_t \quad z_t \sim N(0,1)$$

where the variance equation is defined as

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

S&P500 Daily Returns

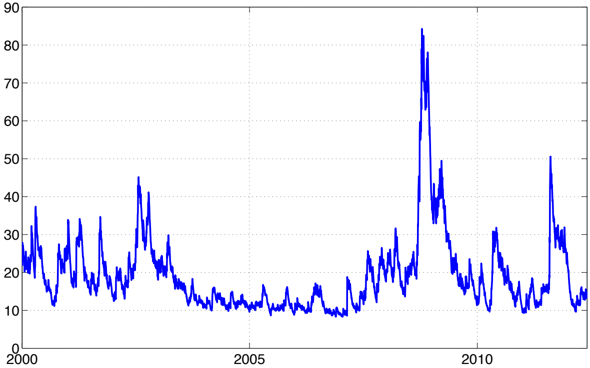


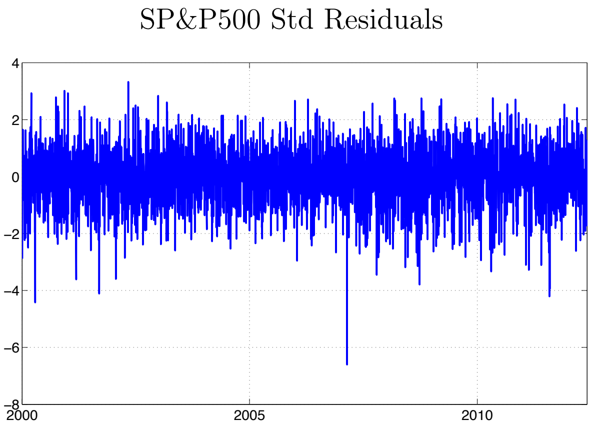
ARCH-LM Stat (3 lags): 516.694 - p–value = 0.000

Estimates

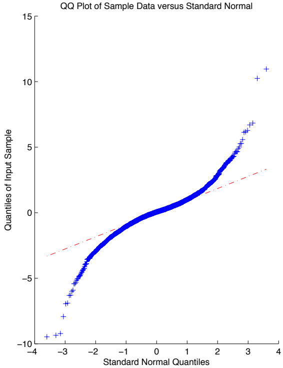
	param	se	t-stat
c	0.04	0.020	20.0
ω	0.02	0.002	10.0
α	0.09	0.008	11.3
β	0.90	0.008	112.5

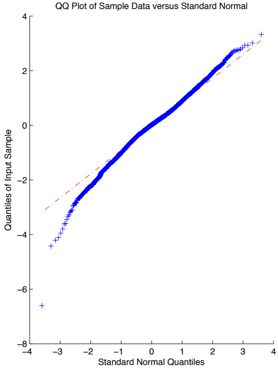
S&P500 Annualized Volatility $\sqrt{252\hat{\sigma}_t^2}$




SP&P500 Std Residuals
4
2
0
−2
−4
−6
−8
2000
2005
2010

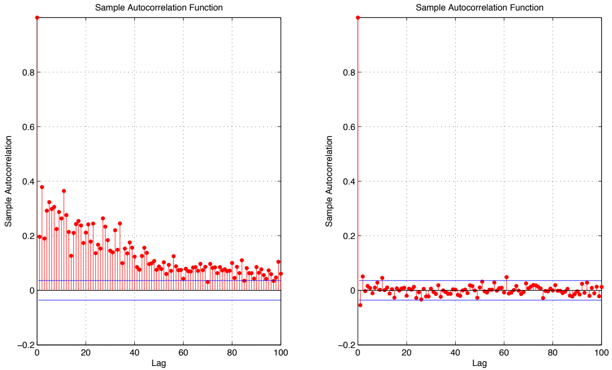
S&P500 Daily Returns (left) vs. Std. Residual (right) QQ-plot



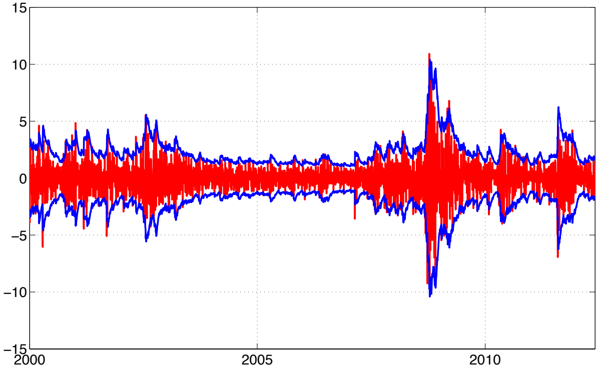


Jarque-Bera Test Before & After: 6746 & 241

S&P500 Daily Squared Returns (left) & Squared Std. Residual (right) ACF



S&P500 Daily Returns 95% conditional confidence intervals $\pm 1.96\sqrt{\hat{\sigma}_t^2}$(assuming $z_t \sim N(0,1)$)



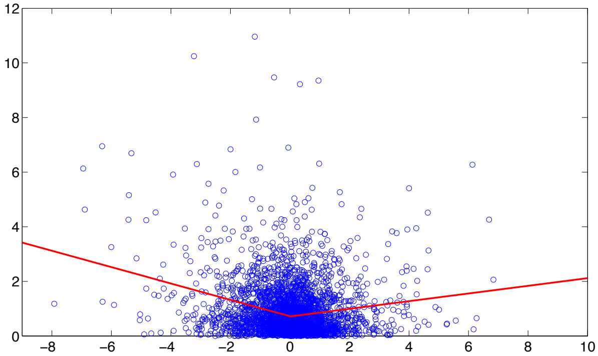
- S&P500 exhibits strong evidence of volatility clustering
- The simple GARCH(1,1) model captures adequately conditional heteroskedasticity in the data
- The hypothesis of normality of the standardized shocks however is still forcefully rejected by the data

The simple GARCH(p, q) has some limitations

The most important is that it cannot take into account the dependence between volatility and the sign of past returns

- Standard GARCH models assume that positive and negative error terms have a symmetric effect on the volatility, i.e. good and bad news have the same effect on the volatility
- In many real situations the volatility reacts asymmetrically to the sign of the shocks. In particular negative past returns have a bigger effect on σ_t^2 than positive returns of the same size
- This dependence is due to the leverage effect, i.e. a negative shock to returns would increase the debt to equity ratio which in turn will increase uncertainty of future returns

S&P500 absolute returns $|\epsilon_t|$ vs lagged returns ϵ_{t-1}



- Consider the model for returns

$$\begin{aligned}\epsilon_t &= \sigma_t z_t\\ \sigma_t^2 &= \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2\end{aligned}$$

- Then we can re-write the volatility as

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i z_{t-i}^2 \sigma_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

which is invariant to changes in the sign of z_t

- Empirically we find that

$$\mathsf{Corr}(\epsilon_t, \epsilon_{t-h}) \simeq 0, \quad \mathsf{Corr}(\epsilon_t^2, \epsilon_{t-h}^2) > 0, \quad \mathsf{Corr}(|\epsilon_t|, |\epsilon_{t-h}|) > 0,$$

which are properties reproduced by GARCH models

- But we also find $\mathsf{Corr}(\epsilon_t^+, \epsilon_{t-h}) < 0$. When $\epsilon_t = \sigma_t z_t$, with σ_t a positive function of ϵ_t as in GARCH, we have

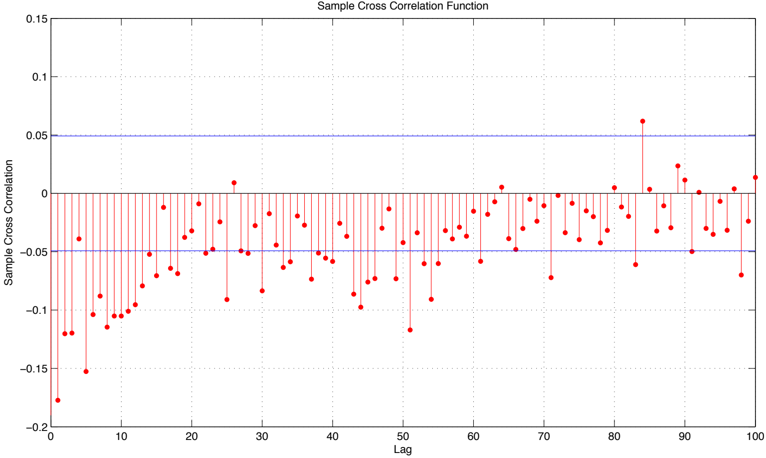
$$\mathsf{Corr}(\epsilon_t^+, \epsilon_{t-h}) = K\,\mathsf{Cov}(\sigma_t, \epsilon_{t-h}) < 0,$$

for some constant $K > 0$

- This is the leverage effect that GARCH models cannot reproduce

- GARCH models with leverage (asymmetric) effects are the subject of next lectures

$$\mathrm{Corr}(\epsilon_t^+, \epsilon_{t-h})$$



Outline

General model

- Consider a stochastic process $\{y_t\}$ such that

$$y_t = \mu_t + \epsilon_t$$

with $\epsilon_t \sim w.n.(0, \sigma^2)$

- The conditional mean of the process is

$$\mathsf{E}_{t-1}[y_t] = \mu_t$$

since $\mathsf{E}_{t-1}[\epsilon_t] = 0$

- For $\{\epsilon_t\}$ we assume a conditional heteroskedastic model

$$\epsilon_t = \sigma_t z_t$$

with $z_t \sim i.i.d.(0, 1)$

- Then the conditional variance of the process is

$$\mathsf{Var}_{t-1}[y_t] = \mathsf{E}_{t-1}[(y_t - \mathsf{E}_{t-1}[y_t])^2] = \mathsf{E}_{t-1}[\epsilon_t^2] = \sigma_t^2 \mathsf{E}_{t-1}[z_t^2] = \sigma_t^2$$

- Consider the ARMA(P, Q)-GARCH(p, q) process

$$\begin{aligned} y_t &= \mu_t + \epsilon_t = c_0 + \sum_{k=1}^{P} a_{0k} y_{t-k} + \sum_{h=1}^{Q} b_{0h} \epsilon_{t-h} + \epsilon_t, \\ \epsilon_t &= \sigma_t z_t, \\ \sigma_t^2 &= \omega_0 + \sum_{i=1}^{p} \alpha_{0i} \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_{0j} \sigma_{t-j}^2. \end{aligned}$$

  where $z_t \sim i.i.d.(0, 1)$, $\omega_0 > 0$, $\alpha_{0i} \geq 0$, $\beta_{0j} \geq 0$.

- We assume that the orders P, Q, p, q are known

- In the conditional mean we can also include exogenous variables

Notation

- We have the vectors of parameters

$$\boldsymbol{\phi} = (c, a_1, \ldots, a_P, b_1, \ldots, b_Q)', \qquad \boldsymbol{\theta} = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)',$$

- We collect all parameters in a single vector

$$\boldsymbol{\psi} = (\boldsymbol{\phi}', \boldsymbol{\theta}')' \in \boldsymbol{\Psi} \subset \mathbb{R}^{P+Q+1} \times (0, +\infty) \times [0, +\infty)^{p+q}.$$

- The true values of the parameters are:

$$\boldsymbol{\psi}_0 = (\boldsymbol{\phi}_0', \boldsymbol{\theta}_0')'$$

- We assume $\boldsymbol{\psi}_0 \in \boldsymbol{\Psi}$ with $\boldsymbol{\Psi}$ compact

- We want to estimate $\boldsymbol{\psi}_0$

- Assume to observe T realizations $(y_1 \dots y_T)$ of $\{y_t\}$. We know that

$$f(y_1 \dots y_T) = f(y_T | y_1 \dots y_{T-1}) f(y_1 \dots y_{T-1})$$

- By iterating we have

$$f(y_1 \dots y_T) = f(y_0) \prod_{t=1}^{T} f(y_t | y_1 \dots y_{t-1}) = f(y_0) \prod_{t=1}^{T} f(y_t | \mathcal{I}_{t-1})$$

- So we need to find $f(y_t | \mathcal{I}_{t-1})$

Distribution of the i.i.d. innovations z_t

- We call the pdf $f_z(z_t, \boldsymbol{\psi}, \boldsymbol{\eta})$ and it depends on
  - the conditional mean and variance parameters $\boldsymbol{\psi} = (\boldsymbol{\phi}', \boldsymbol{\theta}')'$
  - the shape parameters $\boldsymbol{\eta}$

- Since we do not observe $\{z_t\}$, we need the distribution of $\{y_t\}$ for which we observe T realizations

Conditional distribution of y_t

- If f_z is known then

$$f(y_t, \boldsymbol{\psi}, \boldsymbol{\eta}|\mathcal{I}_{t-1}) = f_z(g^{-1}(y_t, \boldsymbol{\psi}), \boldsymbol{\eta}) \left| \frac{\partial g^{-1}(y_t, \boldsymbol{\psi})}{\partial y_t} \right|,$$

  where g is such that

$$\begin{aligned} y_t &= g(z_t, \boldsymbol{\psi}) = \mu_t(\boldsymbol{\psi}) + \sigma_t(\boldsymbol{\psi}) z_t, \\ z_t &= g^{-1}(y_t, \boldsymbol{\psi}) = \frac{y_t - \mu_t(\boldsymbol{\psi})}{\sigma_t(\boldsymbol{\psi})} \end{aligned}$$

- It follows that

$$f(y_t, \boldsymbol{\psi}|\mathcal{I}_{t-1}) = f_z(z_t, \boldsymbol{\psi}, \boldsymbol{\eta}) \frac{1}{\sigma_t(\boldsymbol{\psi})}$$

- Hereafter, unless necessary we omit the dependence on $\boldsymbol{\eta}$

Likelihood

- By taking logs we have the log–likelihood for one observation

$$\ell_t(\boldsymbol{\psi}) \equiv \log f(y_t, \boldsymbol{\psi}|\mathcal{I}_{t-1}) = \log f_z(z_t, \boldsymbol{\psi}) - \frac{1}{2}\log \sigma_t^2(\boldsymbol{\psi})$$

- The sample log–likelihood is

$$L_T(\boldsymbol{\psi}) = \sum_{t=1}^{T} \ell_t(\boldsymbol{\psi})$$

- The Maximum Likelihood (ML) estimator of the parameters $\boldsymbol{\psi}_0$ is defined as

$$\widehat{\boldsymbol{\psi}} = \arg\max_{\boldsymbol{\psi}} \quad L_T(\boldsymbol{\psi})$$

Asymptotic properties of MLE

- If the distribution of z_t is correctly specified then under some regularity conditions and as $T \to \infty$

- $\widehat{\boldsymbol{\psi}}$ is consistent estimator of the true parameter vector $\boldsymbol{\psi}_0$

- $\widehat{\boldsymbol{\psi}}$ is asymptotically normally distributed

- $\widehat{\boldsymbol{\psi}}$ is efficient, i.e. the asymptotic variance covariance matrix of $\widehat{\boldsymbol{\psi}}$ is the inverse of the Fisher information matrix (lower bound)

$$\mathbf{B} = \mathsf{E}_0 \left[\frac{\partial \ell_t(\boldsymbol{\psi}_0)}{\partial \boldsymbol{\psi}} \; \frac{\partial \ell_t(\boldsymbol{\psi}_0)}{\partial \boldsymbol{\psi}'} \right]$$

where the expectation is with respect to the true density $f(y_t, \boldsymbol{\psi}_0 | \mathcal{I}_{t-1})$ and derivatives are computed in $\boldsymbol{\psi}_0$

Possible choices for f_z are

- Gaussian
- Student-t
- Generalized error distribution (GED)

Gaussian ML.

- We assume
$$z_t \sim N(0,1)$$
and since $E[z_t z_{t-k}] = 0$ for $k \neq 0$, then, because of normality, we have also independence.

- The pdf is
$$f_z(z_t, \boldsymbol{\psi}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z_t^2}{2}\right)$$

- Notice that the parameters $\boldsymbol{\psi}$ are implicit in $z_t = g^{-1}(y_t, \boldsymbol{\psi})$ and the shape parameters $\boldsymbol{\eta}$ are not present in this case

- The conditional pdf is

$$f(y_t, \boldsymbol{\psi}|\mathcal{I}_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2(\boldsymbol{\psi})}} \exp\left(-\frac{(y_t - \mu_t(\boldsymbol{\psi})^2}{2\sigma_t^2(\boldsymbol{\psi})}\right)$$

which is like saying $y_t|\mathcal{I}_{t-1} \sim N(\mu_t, \sigma_t^2)$

- The log–likelihood is

$$\ell_t(\boldsymbol{\psi}) = -\frac{1}{2}\log(2\pi) - \frac{(y_t - \mu_t(\boldsymbol{\psi}))^2}{2\sigma_t^2(\boldsymbol{\psi})} - \frac{1}{2}\log\sigma_t^2(\boldsymbol{\psi})$$

- The estimated parameters are solutions of (first order conditions)

$$\sum_{t=1}^{T} \mathbf{s}_t(\widehat{\boldsymbol{\psi}}) \equiv \sum_{t=1}^{T} \frac{\partial \ell_t(\widehat{\boldsymbol{\psi}})}{\partial \boldsymbol{\psi}} = \mathbf{0}$$

which requires a numerical solution

Student-t ML

- The unconditional distribution of z_t and of ϵ_t in financial time series usually display fatter tails than allowed by the Gaussian distribution.

- The Student-t can take into account this feature

$$f_z(z_t, \boldsymbol{\psi}, \nu) = \frac{\Gamma[(\nu+1)/2]}{\Gamma[\nu/2]} \frac{1}{\sqrt{(\nu-2)\pi}} \left(1 + \frac{z_t^2}{\nu-2}\right)^{-(\nu+1)/2} \quad \nu > 2$$

- Notice that the parameters $\boldsymbol{\psi}$ are implicit in $z_t = g^{-1}(y_t, \boldsymbol{\psi})$ and the shape parameters $\boldsymbol{\eta}$ is the number of degrees of freedom ν which determines the tail behavior. Indeed the kurtosis is

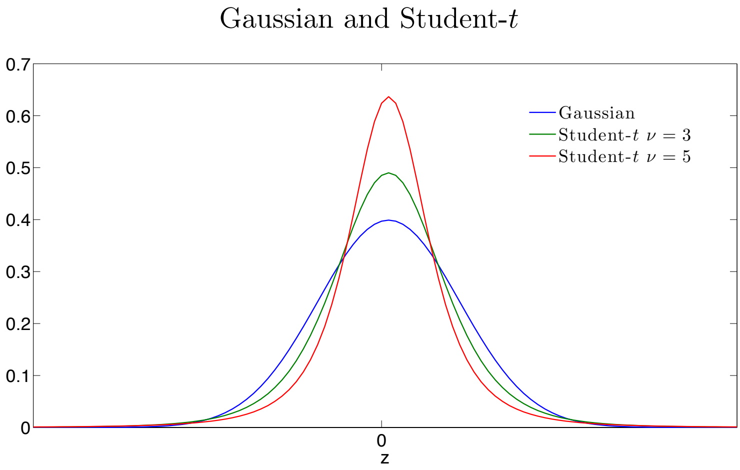
$$\kappa_z = 3\left(\frac{\nu-2}{\nu-4}\right) > 3$$

and it is defined only if $\nu > 4$

The conditional distribution of the observations is then

$$f(y_t, \boldsymbol{\psi}, \nu | \mathcal{I}_{t-1}) = \frac{\Gamma[(\nu+1)/2]}{\Gamma[\nu/2]} \frac{1}{\sqrt{(\nu-2)\pi\sigma_t^2(\boldsymbol{\psi})}} \left(1 + \frac{(y_t - \mu_t(\boldsymbol{\psi}))^2}{(\nu-2)\sigma_t^2(\boldsymbol{\psi})}\right)^{-\frac{\nu+1}{2}}$$

and ML can be used to estimate the parameters $(\boldsymbol{\psi}_0, \nu_0)$


Gaussian and Student-t
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
z
Gaussian
Student-t $\nu = 3$
Student-t $\nu = 5$

GED ML

- An even more flexible distribution is the Generalized Error Distribution

$$f_z(z_t, \boldsymbol{\psi}, \nu) = \frac{\nu}{\lambda 2^{1+1/\nu}\Gamma[1/\nu]} \exp\left(-\frac{1}{2}\left|\frac{z_t}{\lambda}\right|^{\nu}\right) \qquad \nu > 0$$

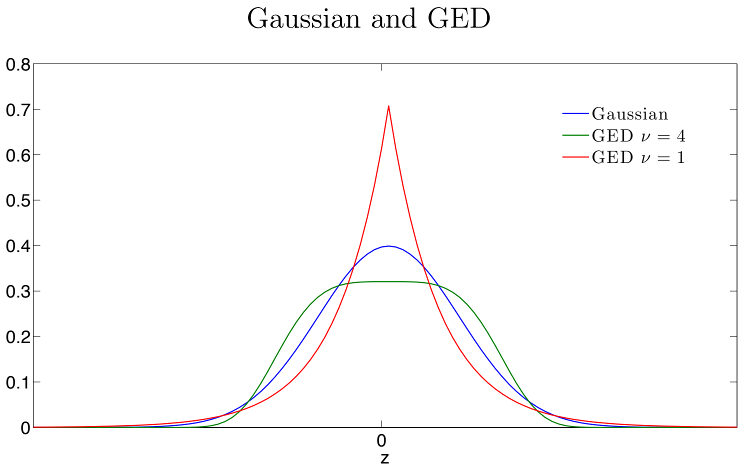
where $\lambda = [2^{-2/\nu}\Gamma[1/\nu]\Gamma[3/\nu]^{-1}]^{1/2}$

- When $\nu < 2$ (> 2) the GED has heavier (lighter) tails than the Gaussian. When $\nu = 2$ we have a Gaussian

- The conditional distribution of the observations is

$$f(y_t, \boldsymbol{\psi}, \nu | \mathcal{I}_{t-1}) = \frac{\nu}{\lambda 2^{1+1/\nu}\Gamma[1/\nu]\sigma_t^2(\boldsymbol{\psi})} \exp\left(-\frac{1}{2}\left|\frac{y_t - \mu_t(\boldsymbol{\psi})}{\lambda\sigma_t^2(\boldsymbol{\psi})}\right|^{\nu}\right)$$

and we can use ML again to estimate $(\boldsymbol{\psi}_0, \nu_0)$


Gaussian and GED
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
z
Gaussian
GED $\nu = 4$
GED $\nu = 1$

Outline

- Consider a family of parametric distributions $\mathcal{P} = \{P_{\boldsymbol{\psi}}, \boldsymbol{\psi} \in \Psi\}$, to which correspond densities $g(y, \boldsymbol{\psi})$. Define the true unknown distribution as P_0 defined through the true value $\boldsymbol{\psi}_0$ such that $P_0 = P_{\boldsymbol{\psi}_o}$ with density $f(y, \boldsymbol{\psi}_0)$

- We know that if the model is correctly specified ML estimators have all the "good" asymptotic properties

- However, in general there may be specification errors $P_0 \notin \mathcal{P}$ and the best we can do is to find $P_0^* \in \mathcal{P}$ that minimizes the distance with P_0. To this distribution corresponds the quasi true value of the parameters $\boldsymbol{\psi}_0^*$ such that $P_0^* = P_{\boldsymbol{\psi}_o^*}$

- The value $\boldsymbol{\psi}_0^*$ is such that it minimizes the Kullback-Leibler information criterion (a measure of the informational distance between distributions)

$$\min_{\boldsymbol{\psi}} I(P_{\boldsymbol{\psi}}, P_0) = \min_{\boldsymbol{\psi}} \mathsf{E}_0 \left[\log \frac{f(y, \boldsymbol{\psi}_0)}{g(y, \boldsymbol{\psi})}\right] = \int_{\mathcal{Y}} \log \frac{f(y, \boldsymbol{\psi}_0)}{g(y, \boldsymbol{\psi})} f(y, \boldsymbol{\psi}_0) \mathrm{d}y$$

- Which is equivalent to solve

$$\max_{\boldsymbol{\psi}} \mathsf{E}_0[\log g(y, \boldsymbol{\psi})]$$

i.e. $\boldsymbol{\psi}_0^*$ is the value that maximizes the expected log–likelihood

QML

- In the present context we have

$$\boldsymbol{\psi}_0^* = \arg\max_{\boldsymbol{\psi}} \mathsf{E}_{t-1}\mathsf{E}_0[\log g(y_t, \boldsymbol{\psi}|\mathcal{I}_{t-1})].$$

- Define $\ell_t(\boldsymbol{\psi}) = \log g(y_t, \boldsymbol{\psi}|\mathcal{I}_{t-1})$, then the estimator of $\boldsymbol{\psi}_0^*$ is defined as

$$\widehat{\boldsymbol{\psi}} = \arg\max_{\boldsymbol{\psi}} \sum_{t=1}^{T} \ell_t(\boldsymbol{\psi}).$$

- This is the Quasi Maximum Likelihood (QML) estimator

Asymptotic properties of QMLE

- We can prove that, under regularity conditions, and if $\mathsf{E}[z_t^4] = \kappa_z < \infty$,

$$\sqrt{T}(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0^*) \to \mathcal{N}(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1})$$

- In general $\widehat{\boldsymbol{\psi}}$ is a consistent estimator of $\boldsymbol{\psi}_0^*$ and not of $\boldsymbol{\psi}_0$

- However, if the first two conditional moments are correctly specified and $\mathcal{P}$ is the exponential family of distributions (e.g. the Gaussian), then $\widehat{\boldsymbol{\psi}}$ is a consistent estimator of the true value $\boldsymbol{\psi}_0$

- The QML estimator is then consistent, asymptotically normal, but it is no more efficient

- In general
$$\mathbf{B}^{-1} < \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$$
we have a larger asymptotic variance covariance matrix which accounts for the additional uncertainty due to possible specification errors

- In order to have the right confidence intervals and to do proper inference, we have to compute the matrices

$$\mathbf{A} = -\mathsf{E}_0\left[\frac{\partial^2 \ell_t(\boldsymbol{\psi}_0)}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'}\right]$$

  and

$$\mathbf{B} = \mathsf{E}_0\left[\frac{\partial \ell_t(\boldsymbol{\psi}_0)}{\partial\boldsymbol{\psi}}\frac{\partial \ell_t(\boldsymbol{\psi}_0)}{\partial\boldsymbol{\psi}'}\right]$$

- if the model is correctly specified, i.e. $\ell_t(\boldsymbol{\psi}) = \log f(y_t, \boldsymbol{\psi}|\mathcal{I}_{t-1})$, then $\mathbf{A} = \mathbf{B}$ and we are back to MLE

GARCH–QML

- We assume a Gaussian log–likelihood, then

$$\ell_t(\boldsymbol{\psi}) = -\frac{1}{2}\log\sigma_t^2(\boldsymbol{\psi}) - \frac{(y_t - \mu_t(\boldsymbol{\psi}))^2}{2\sigma_t^2(\boldsymbol{\psi})}$$

  and for simplicity let's start with the GARCH(p, q) model, i.e. when $\mu_t = 0$ and $\boldsymbol{\psi} \equiv \boldsymbol{\theta}$

$$\ell_t(\boldsymbol{\theta}) = -\frac{1}{2}\log\sigma_t^2(\boldsymbol{\theta}) - \frac{y_t^2}{2\sigma_t^2(\boldsymbol{\theta})}$$

- The score is

$$
\begin{aligned}
\mathbf{s}_t(\boldsymbol{\theta}) \equiv \frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= -\frac{1}{2\sigma_t^2(\theta)}\frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{y_t^2}{2(\sigma_t^2(\boldsymbol{\theta}))^2}\frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \\
&= \frac{1}{2\sigma_t^2(\theta)}\frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\left(\frac{y_t^2}{\sigma_t^2(\boldsymbol{\theta})} - 1\right) = \\
&= \frac{1}{2\sigma_t^2(\theta)}\frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\left(z_t^2 - 1\right).
\end{aligned}
$$

- We know that $\mathsf{E}_0[\mathbf{s}_t(\boldsymbol{\theta}_0)] = \mathbf{0}$
- Therefore the variance of the score, i.e. the Fisher information $\mathbf{B}$, is

$$\mathbf{B} = \mathsf{E}_0\left[\mathbf{s}_t(\boldsymbol{\theta}_0)\mathbf{s}_t'(\boldsymbol{\theta}_0)\right] = (\kappa_z - 1)\mathsf{E}_0\left[\frac{1}{4(\sigma_t^2(\boldsymbol{\theta}_0))^2}\frac{\partial\sigma_t^2(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\right]$$

where $\kappa_z = \mathsf{E}[z_t^4]$ is the kurtosis coefficient of z_t (remember that $\mathsf{E}[z_t^2] = 1$)

- To compute $\mathbf{A}$ we need the second derivatives

$$
\begin{aligned}
\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} &= \frac{\partial \mathbf{s}_t'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{1}{2(\sigma_t^2(\boldsymbol{\theta}))^2} \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \left(\frac{y_t^2}{\sigma_t^2(\boldsymbol{\theta})} - 1 \right) \\
& -\frac{1}{2(\sigma_t^2(\boldsymbol{\theta}))^2} \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \frac{y_t^2}{\sigma_t^2(\boldsymbol{\theta})} \\
& +\frac{1}{2\sigma_t^2(\boldsymbol{\theta})} \frac{\partial^2 \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \left(\frac{y_t^2}{\sigma_t^2(\boldsymbol{\theta})} - 1 \right) = \\
&= \frac{1}{2(\sigma_t^2(\boldsymbol{\theta}))^2} \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} - \frac{z_t^2}{\sigma_t^2(\boldsymbol{\theta})} \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \\
& + \frac{z_t^2}{2} \frac{\partial^2 \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \frac{1}{2\sigma_t^2(\boldsymbol{\theta})} \frac{\partial^2 \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}
\end{aligned}
$$

with $z_t^2 = y_t^2 / \sigma_t^2(\boldsymbol{\theta})$

- By taking expectations and remembering that $\mathsf{E}[z_t^2] = 1$
- The Hessian

$$\mathbf{A} = -\mathsf{E}_0\left[\frac{\partial \mathbf{s}_t'(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right] = \mathsf{E}_0\left[\frac{1}{2(\sigma_t^2(\boldsymbol{\theta}_0))^2}\frac{\partial \sigma_t^2(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\frac{\partial \sigma_t^2(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right]$$

- The Fisher Information

$$\mathbf{B} = \mathsf{E}_0\left[\mathbf{s}_t(\boldsymbol{\theta}_0)\mathbf{s}_t'(\boldsymbol{\theta}_0)\right] = (\kappa_z - 1)\mathsf{E}_0\left[\frac{1}{4(\sigma_t^2(\boldsymbol{\theta}_0))^2}\frac{\partial \sigma_t^2(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\frac{\partial \sigma_t^2(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right].$$

- Notice that if the true distribution of z_t were a Gaussian, then $\kappa_z = 3$ and $\mathbf{A} = \mathbf{B}$

- The QML asymptotic variance covariance matrix for a GARCH model is then $\mathbf{V} = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$

$$\begin{aligned}
\mathbf{V} = \; = \;& \left(\mathsf{E}_0\left[\frac{1}{2(\sigma_t^2(\boldsymbol{\theta}_0))^2}\frac{\partial\sigma_t^2(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\right]\right)^{-1}\\
& (\kappa_z - 1)\mathsf{E}_0\left[\frac{1}{4(\sigma_t^2(\boldsymbol{\theta}_0))^2}\frac{\partial\sigma_t^2(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\right]\\
& \left(\mathsf{E}_0\left[\frac{1}{2(\sigma_t^2(\boldsymbol{\theta}_0))^2}\frac{\partial\sigma_t^2(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\right]\right)^{-1} =\\
= \;& (\kappa_z - 1)\left(\mathsf{E}_0\left[\frac{1}{(\sigma_t^2(\boldsymbol{\theta}_0))^2}\frac{\partial\sigma_t^2(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\right]\right)^{-1}
\end{aligned}$$

- A consistent estimator of $\mathbf{V}$ is obtained by replacing expectations with sums and the true value of the parameters with the QML estimates:

$$\widehat{\mathbf{V}} = \frac{1}{T}\sum_{t=1}^{T}(\widehat{\kappa_z} - 1)\left(\frac{1}{(\sigma_t^2(\widehat{\boldsymbol{\theta}}))^2}\frac{\partial \sigma_t^2(\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\frac{\partial \sigma_t^2(\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}\right)^{-1}$$

- Confidence intervals for $\widehat{\boldsymbol{\theta}}$ are built using the square–root of the diagonal elements of $T^{-1}\widehat{\mathbf{V}}$

- If we have a model also for the conditional mean then derivatives must be computed also with respect to the parameters $\boldsymbol{\phi}$
- For an ARMA(P, Q)-GARCH(p, q) the conditional variance depends both on $\boldsymbol{\theta}$ and on $\boldsymbol{\psi}$

$$
\begin{aligned}
\sigma_t^2 &= \omega_0 + \sum_{i=1}^{p} \alpha_{0i} \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_{0j} \sigma_{t-j}^2 = \\
&= \omega_0 + \sum_{i=1}^{p} \alpha_{0i} \left(y_{t-i} - c_0 - \sum_{k=1}^{P} a_{0k} y_{t-k-i} - \sum_{h=1}^{Q} b_{0h} \epsilon_{t-h-i} \right)^2 \\
&\quad + \sum_{j=1}^{q} \beta_{0j} \sigma_{t-j}^2
\end{aligned}
$$

which depends also on the conditional mean parameters $\boldsymbol{\phi} = (c, a_1, \ldots, a_P, b_1, \ldots, b_Q)'$

ARMA–GARCH QML

- If we assume a Gaussian log–likelihood we have

$$\ell_t(\boldsymbol{\phi}, \boldsymbol{\theta}) = -\frac{1}{2} \log \sigma_t^2(\boldsymbol{\phi}, \boldsymbol{\theta}) - \frac{\epsilon_t^2(\boldsymbol{\phi})}{2\sigma_t^2(\boldsymbol{\phi}, \boldsymbol{\theta})}$$

which has to be maximized jointly with respect to $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$.

- The score is

$$\mathbf{s}_t(\boldsymbol{\phi},\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial \ell_t(\boldsymbol{\phi},\boldsymbol{\theta})}{\partial \boldsymbol{\phi}} \\ \frac{\partial \ell_t(\boldsymbol{\phi},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \end{pmatrix} =$$

$$= \begin{pmatrix} \frac{1}{2\sigma_t^2(\boldsymbol{\phi},\boldsymbol{\theta})} \frac{\partial \sigma_t^2(\boldsymbol{\phi},\boldsymbol{\theta})}{\partial \boldsymbol{\phi}} \left(\frac{\epsilon_t^2(\boldsymbol{\phi})}{\sigma_t^2(\boldsymbol{\phi},\boldsymbol{\theta})} - 1 \right) - \frac{\epsilon_t(\boldsymbol{\phi})}{\sigma_t^2(\boldsymbol{\phi},\boldsymbol{\theta})} \frac{\partial \epsilon_t(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}} \\ \frac{1}{2\sigma_t^2(\boldsymbol{\phi},\boldsymbol{\theta})} \frac{\partial \sigma_t^2(\boldsymbol{\phi},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left(\frac{\epsilon_t^2(\boldsymbol{\phi})}{\sigma_t^2(\boldsymbol{\phi},\boldsymbol{\theta})} - 1 \right) \end{pmatrix}$$

- The second row is the same we had for the simple GARCH model
- Using the score the matrices **A** and **B** can be computed

- If z_t has a symmetric distribution, i.e. $\mathsf{E}[z_t^3] = 0$
- The Fisher Information is

$$\mathbf{B} = \mathsf{E}_0[\mathbf{s}_t(\boldsymbol{\phi}_0, \boldsymbol{\theta}_0)\mathbf{s}_t'(\boldsymbol{\phi}_0, \boldsymbol{\theta}_0)] = \begin{pmatrix} \mathbf{I}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 \end{pmatrix}$$

- with

$$\begin{aligned} \mathbf{I}_1 &= (\kappa_z - 1)\mathsf{E}_0\left[\frac{1}{4(\sigma_t^2(\boldsymbol{\phi}_0, \boldsymbol{\theta}_0))^2}\frac{\partial \sigma_t^2(\boldsymbol{\phi}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\phi}}\frac{\partial \sigma_t^2(\boldsymbol{\phi}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\phi}'}\right] + \\ &+\mathsf{E}_0\left[\frac{1}{\sigma_t^2(\boldsymbol{\phi}_0, \boldsymbol{\theta}_0)}\frac{\partial \epsilon_t(\boldsymbol{\phi}_0)}{\partial \boldsymbol{\phi}}\frac{\partial \epsilon_t(\boldsymbol{\phi}_0)}{\partial \boldsymbol{\phi}'}\right] \\ \mathbf{I}_2 &= (\kappa_z - 1)\mathsf{E}_0\left[\frac{1}{4(\sigma_t^2(\boldsymbol{\phi}_0, \boldsymbol{\theta}_0))^2}\frac{\partial \sigma_t^2(\boldsymbol{\phi}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\frac{\partial \sigma_t^2(\boldsymbol{\phi}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right] \end{aligned}$$

- The Hessian is

$$\mathbf{A} = \begin{pmatrix} \mathbf{J}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_2 \end{pmatrix}$$

- with

$$\begin{aligned} \mathbf{J}_1 &= \mathsf{E}_0\left[\frac{1}{2(\sigma_t^2(\boldsymbol{\phi}_0,\boldsymbol{\theta}_0))^2}\frac{\partial\sigma_t^2(\boldsymbol{\phi}_0,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\phi}}\frac{\partial\sigma_t^2(\boldsymbol{\phi}_0,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\phi}'}\right]+ \\ &\quad +\mathsf{E}_0\left[\frac{1}{\sigma_t^2(\boldsymbol{\phi}_0,\boldsymbol{\theta}_0)}\frac{\partial\epsilon_t(\boldsymbol{\phi}_0)}{\partial\boldsymbol{\phi}}\frac{\partial\epsilon_t(\boldsymbol{\phi}_0)}{\partial\boldsymbol{\phi}'}\right] \\ \mathbf{J}_2 &= \mathsf{E}_0\left[\frac{1}{2(\sigma_t^2(\boldsymbol{\phi}_0,\boldsymbol{\theta}_0))^2}\frac{\partial\sigma_t^2(\boldsymbol{\phi}_0,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2(\boldsymbol{\phi}_0,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\right] \end{aligned}$$

- In case of z_t Gaussian, $\kappa_z = 3$ and $\mathbf{J}_1 = \mathbf{I}_1$ and $\mathbf{J}_2 = \mathbf{I}_2$

- Therefore, if z_t has a symmetric distribution, the asymptotic variance covariance matrix is block diagonal

$$\mathbf{V} = \begin{pmatrix} \mathbf{J}_1^{-1}\mathbf{I}_1\mathbf{J}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_2^{-1}\mathbf{I}_2\mathbf{J}_2^{-1} \end{pmatrix}$$

- We have asymptotic independence between estimated ARMA and GARCH coefficients

- However notice that the distribution of the estimators of the ARMA coefficients depends on the GARCH coefficients, while on the other hand the asymptotic accuracy in the estimated GARCH coefficients is not affected by the ARMA part

- We can first estimate an ARMA with heteroskedastic errors and then take the residuals and estimate a GARCH

Example

- Let us consider a simpler model, an AR(1)–ARCH(1) process:

$$\begin{aligned} y_t &= a_0 y_{t-1} + \epsilon_t \\ \epsilon_t &= \sigma_t z_t \\ \sigma_t^2 &= \omega_0 + \alpha_0 \epsilon_{t-1}^2 = \omega_0 + \alpha_0 (y_{t-1} - a_0 y_{t-2})^2. \end{aligned}$$

- We have the parameters

$$\boldsymbol{\phi} = a, \quad \boldsymbol{\theta} = (\omega, \alpha)'.$$

- Then

$$
\begin{aligned}
\frac{\partial \epsilon_t(a)}{\partial a} &= -y_{t-1}, \\
\frac{\partial \sigma_t^2(a, \boldsymbol{\theta})}{\partial a} &= -2\alpha y_{t-2}(y_{t-1} - a y_{t-2}) = -2\alpha y_{t-2}\epsilon_{t-1}, \\
\frac{\partial \sigma_t^2(a, \boldsymbol{\theta})}{\partial \omega} &= 1, \\
\frac{\partial \sigma_t^2(a, \boldsymbol{\theta})}{\partial \alpha} &= (y_{t-1} - a y_{t-2})^2 = \epsilon_{t-1}^2.
\end{aligned}
$$

- For the ARCH parameters we have

$$\mathbf{I}_2 = \frac{(\kappa_z - 1)}{4} \mathsf{E}_0 \begin{pmatrix} \epsilon^4_{t-1}/\sigma^4_t & \epsilon^2_{t-1}/\sigma^4_t \\ \epsilon^2_{t-1}/\sigma^4_t & 1/\sigma^4_t \end{pmatrix}$$

which does not depend on the AR parameters

- For the AR parameter we have

$$\mathbf{I}_1 = (\kappa_z - 1)\mathsf{E}_0 \left(\frac{y^2_{t-2}\epsilon^2_{t-1}\alpha^2}{\sigma^4_t} \right) + \mathsf{E}_0 \left(\frac{y^2_{t-1}}{\sigma^2_t} \right)$$

- If we had homoskedasticity we'd had only the second term which would be the usual term from OLS estimation but due to the ARCH effects the asymptotic variance depends also on the ARCH parameters

- Notice that, by exploiting the fact that

$$\mathsf{E}[\epsilon_t^2] \equiv \sigma_\epsilon^2 = \frac{\omega}{1-\alpha}$$

we have one parameter less in the ARCH which is replaced by σ_ϵ^2 that is easily estimated as

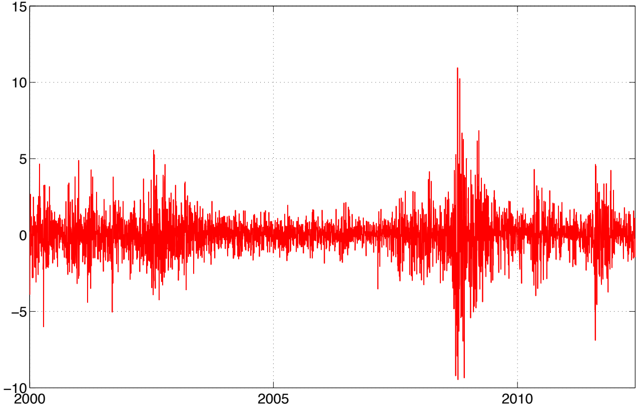
$$\widehat{\sigma_\epsilon^2} = \frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2$$

and finally

$$\widehat{\omega} = \widehat{\sigma_\epsilon^2}(1-\widehat{\alpha})$$

- This method is called variance targeting and is useful when we have many parameters to estimate

S&P500 Daily Returns



MLE of a GARCH(1,1) model with Gaussian innovations:

$$\omega = 0.0149 \quad \alpha = 0.0870 \quad \beta = 0.9042$$

If we believe that the innovations $z_t = \epsilon_t/\sigma_t$ are truly Gaussian then the asymptotic variance covariance matrix is given by the inverse Fisher information which is equivalent to the inverse Hessian:

$$\mathbf{J}^{-1} = \begin{pmatrix} 1.0755 & 1.1689 & -2.0443 \\ 1.1689 & 7.9883 & -7.5113 \\ -2.0443 & -7.5113 & 8.5347 \end{pmatrix} \cdot 10^{-05}$$

and the standard errors are the squared root of the diagonal elements:

$$se_\omega = 0.0033 \quad se_\alpha = 0.0089 \quad se_\beta = 0.0092$$

Actually the matrices $\mathbf{I}$ and $\mathbf{J}$ are not equal and the QML asymptotic variance covariance matrix is given by:

$$\mathbf{J}\mathbf{I}^{-1}\mathbf{J} = \begin{pmatrix} 2.8104 & 1.0996 & -3.3160 \\ 1.0996 & 11.9537 & -10.4660 \\ -3.3160 & -10.4660 & 11.8708 \end{pmatrix} \cdot 10^{-05}$$

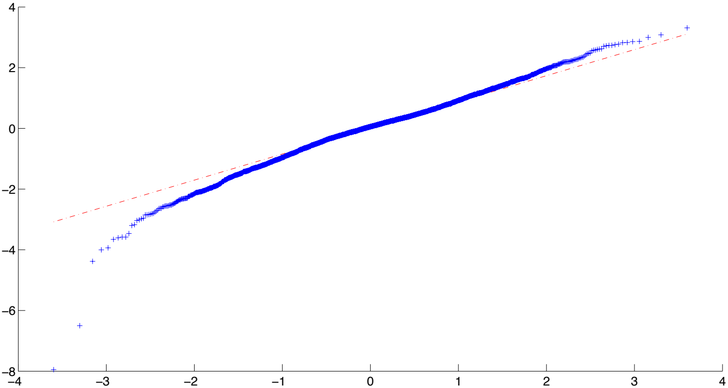
and the QML (robust) standard errors are

$$se^*_\omega = 0.0053 \quad se^*_\alpha = 0.0109 \quad se^*_\beta = 0.0109$$

and as expected

$$se^*_\omega > se_\omega \quad se^*_\alpha > se_\alpha \quad se^*_\beta > se_\beta$$

Is normality the right assumption? Let's look at the empirical quantiles of z_t (vertical axis) vs. the quantiles of a standard normal (horizontal axis)



Fat tails?

We can use ML with Student-t or GED distribution. Here are the estimated coefficients with their errors.

Student-t:

$$\omega = 0.0106(0.0035) \quad \alpha = 0.0834(0.0104) \quad \beta = 0.9119(0.0102)$$

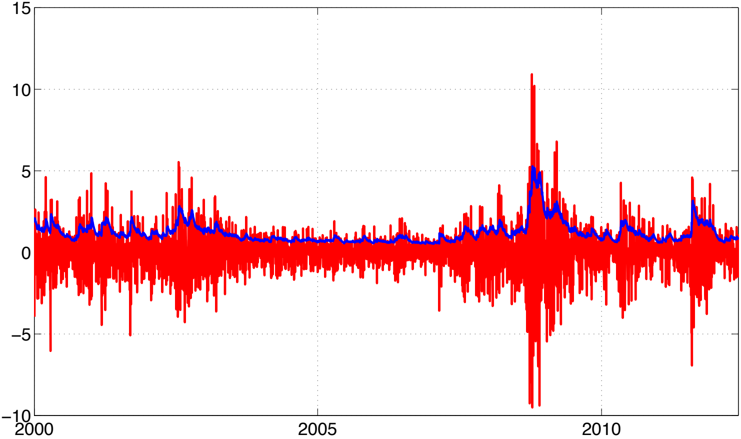
and $\nu = 8.8296(1.4565)$

GED:

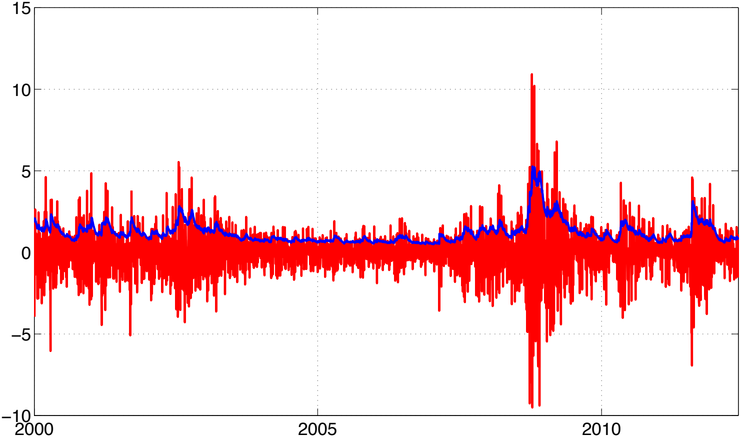
$$\omega = 0.0123(0.0041) \quad \alpha = 0.0846(0.0105) \quad \beta = 0.9089(0.0103)$$

and $\nu = 1.4489(0.0612)$

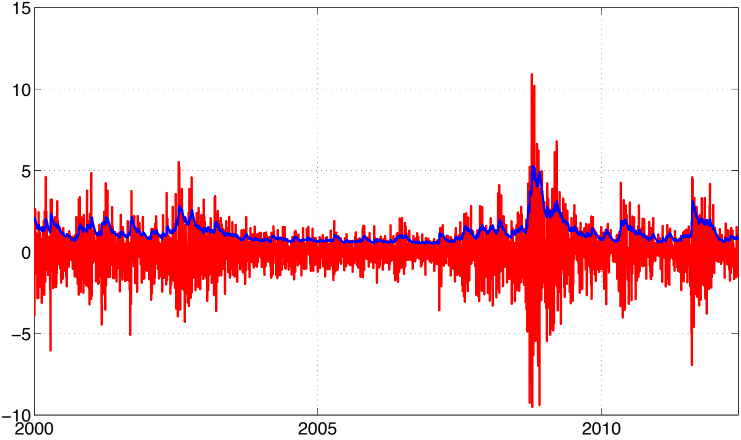
Estimated volatility with Gaussian innovations



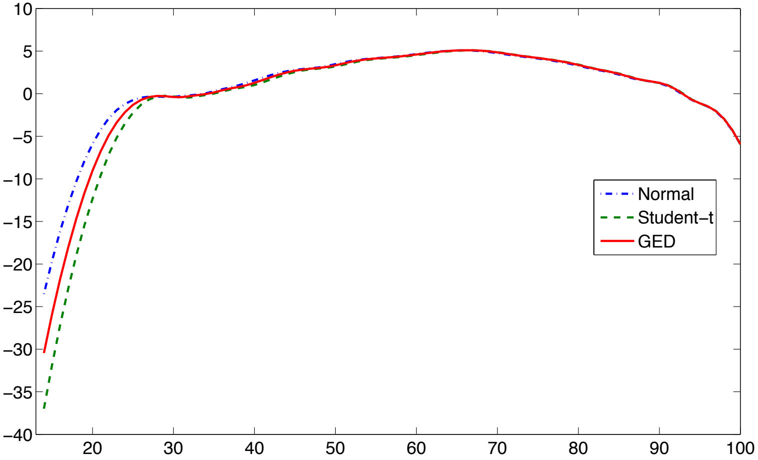
Estimated volatility with Student-t innovations



Estimated volatility with GED innovations



Kernel density estimations of Std. Residuals (log–scale)



Textbook references:

Hamilton, J. D. "Time Series Analysis", Chapter 21, Princeton University Press, 1994.

Francq, C. and Zakoian, J.M. "GARCH Models", Wiley, 2010.