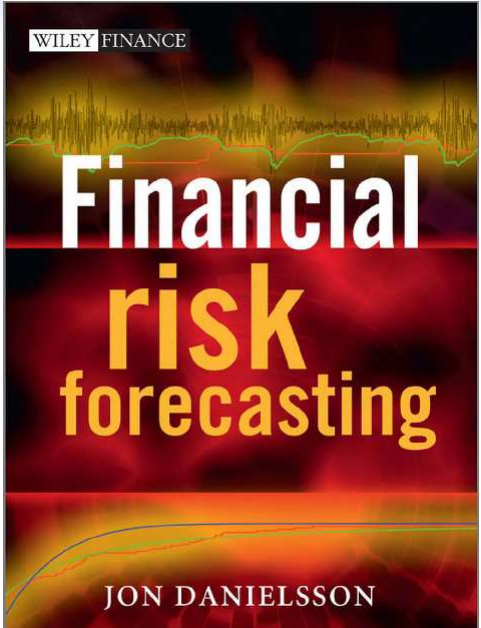
Financial Risk Forecasting
Chapter 2
Univariate Volatility Modelling

Jon Danielsson ©2023
London School of Economics

To accompany
Financial Risk Forecasting
www.financialriskforecasting.com
Published by Wiley 2011

Version 8.0, August 2023


WILEY FINANCE
Financial
risk
forecasting
JON DANIELSSON

Univariate Volatility Modelling

Volatility

- Volatility is the main measure of risk (see Chapter 4)
- Investment decisions
- Portfolio construction
- Derivative pricing

Estimation and Forecasting

- In this Chapter we focus on the estimation and forecasting of volatility for a single asset (univariate)
- The next Chapter does multivariate volatility
- The focus in this Chapter is on estimation, both of a model and in-sample volatilities
- Chapter 5 introduces (out-of-sample) forecasting of volatilities

Structure

- The theoretical specification of common volatility models
- The theory of estimating the models
- Practical implementation in estimation
- Diagnostics of estimated volatility models

Univariate Volatility Models

- Moving average (MA)
- Exponentially weighted moving average (EWMA)
- Generalized autoregressive conditional heteroskedasticity (GARCH) and its extension models
- Implied volatilities, like VIX volatility index
- Realised volatilities

Notation

W_E	Estimation window
λ	Decay factor in EWMA
ϵ_t	Residuals
ω, α, β	Main model parameters
ζ, δ	Other model parameters
L_1, L_2	Lags in volatility models

Simple Models

The Mean of Returns

- Recall from the last Chapter that the mean of daily returns is very low
- In one case mean is 0.0001007711
- While the standard error is more than 100 times larger, at 0.01245144
- With daily returns it is usually quite safe to assume the mean is zero
- It is convenient because it makes the mathematics simpler
- We will return to this issue several times later in the book

Volatility

- Volatility is the standard error (square root of variance) of returns
- The definition of variance is

$$\hat{\sigma}^2 = \frac{1}{T-1}\sum_{i=1}^{T}(x_i - \bar{x})^2$$

- And since we assume the mean is zero, this simplifies to

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{i=1}^{T}x_i^2$$

Moving Average

- The simplest volatility model is *moving average*
- Where the *conditional* volatility is the square root of the average of squared returns over the *estimation window*, W_E

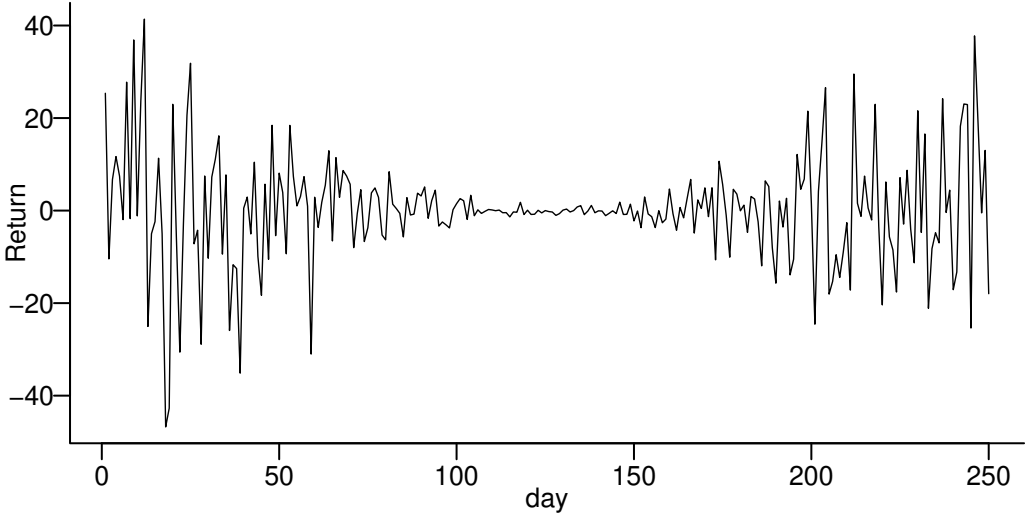
$$\hat{\sigma}_t = \sqrt{\frac{1}{W_E}\sum_{i=1}^{W_E} y_{t-i}^2}$$

- Note that this is a *one-day-ahead* forecast
- Because it uses information until time $t-1$ to do the calculation for time t

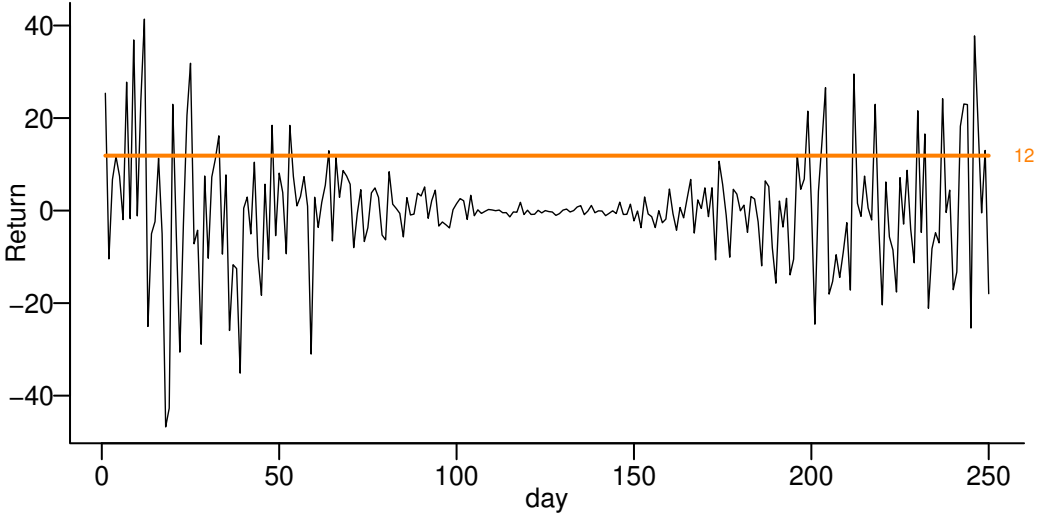
Moving Average (cont.)

- The moving average model will generally perform quite badly
- Because it is quite sensitive to the size of the estimation window (see the next plots)
- And because it does not weigh history

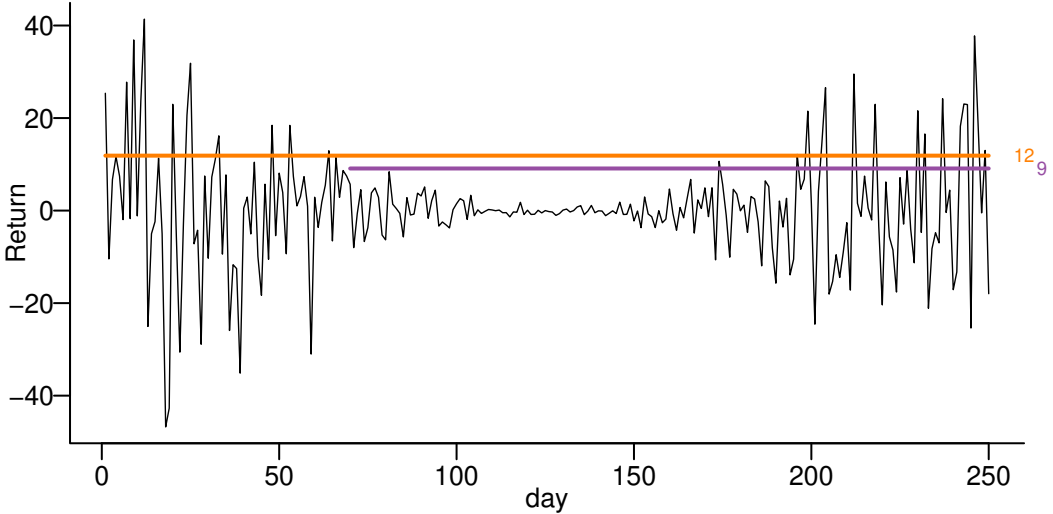
Hypothetical Returns



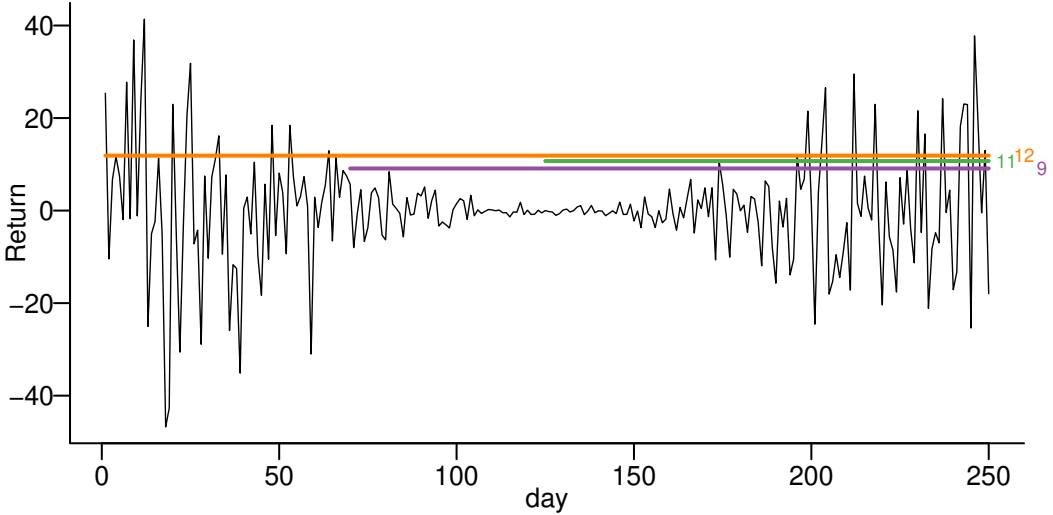
Hypothetical Returns (cont.)



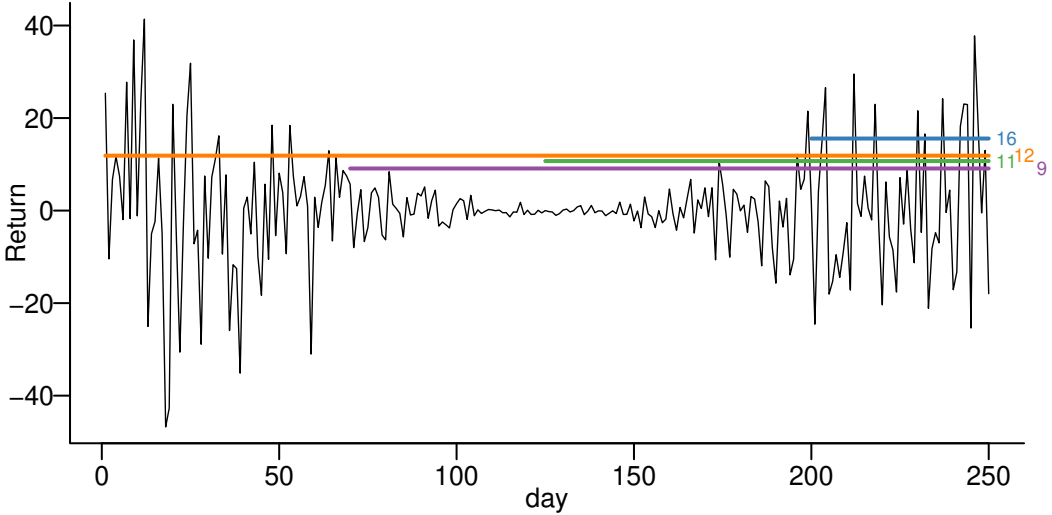
Hypothetical Returns (cont.)



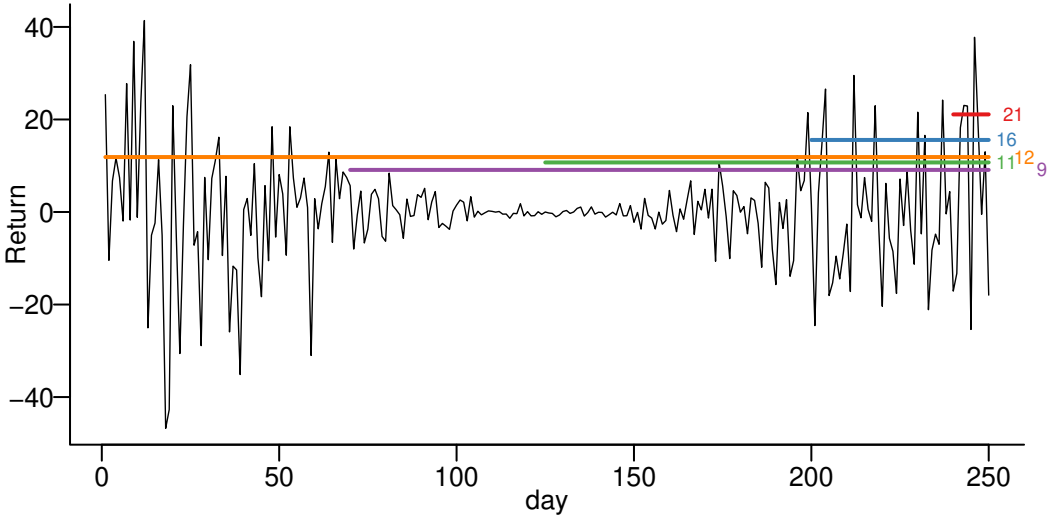
Hypothetical Returns (cont.)



Hypothetical Returns (cont.)



Hypothetical Returns (cont.)



Two Concepts of Volatility

- Unconditional volatility (σ)
  Volatility over an entire time period

- Conditional volatility (σ_t)
  Volatility conditional on a given time period, the past history, model and model parameters

Exponentially Weighted Moving Average (EWMA)

Exponentially Weighted Moving Average (EWMA)

- JP Morgan proposed a model called *RiskMetrics* in 1993
- Since they later used that name for a consulting company they spun off, we use the more pedestrian
- *Exponentially weighted moving average* (EWMA)

EWMA

- Volatility a *weighted* sum of past returns, with weights w_i

$$\hat{\sigma}_t^2 = w_1 y_{t-1}^2 + w_2 y_{t-2}^2 \ldots + w_{W_E} y_{t-W_E}^2$$

- Let the weights be *exponentially declining*, and denote them by λ^i, start by:

$$\hat{\sigma}_t^2 = \lambda y_{t-1}^2 + \lambda^2 y_{t-2}^2 \ldots + \lambda^{W_E} y_{t-W_E}^2$$

- $1 > \lambda > 0$
- If W_E is large enough, the terms λ^n are negligible for all $n \geq W_E$
- So set $W_E = \infty$
- However, we have to ensure the sum of the weights is one

Deriving the EWMA Model

- Sum of an infinite power series

$$\text{Sum} = \frac{\lambda}{1-\lambda} = \lambda + \lambda^2 + \lambda^3 + \ldots + \lambda^\infty$$

- Then since this is the sum of a geometric progression:

$$\hat{\sigma}_t^2 = \lambda y_{t-1}^2 + \lambda^2 y_{t-2}^2 \ldots + \lambda^\infty y_{t-\infty}^2$$

- We get

$$\hat{\sigma}_t^2 = \frac{1-\lambda}{\lambda} \sum_{i=1}^{\infty} \lambda^i y_{t-i}^2$$

Deriving the EWMA Model (cont.)

- Rewriting

$$\hat{\sigma}_t^2 = (1-\lambda)y_{t-1}^2 + \frac{1-\lambda}{\lambda}\sum_{i=2}^{\infty}\lambda^i y_{t-i}^2$$

- Since

$$(1-\lambda)\sum_{i=1}^{\infty}\lambda^i = (1-\lambda)(\lambda^1 + \cdots + \lambda^\infty) = \lambda$$

- We get the EWMA equation

$$\hat{\sigma}_t^2 = (1-\lambda)y_{t-1}^2 + \lambda\hat{\sigma}_{t-1}^2$$

What Is λ ?

$$\hat{\sigma}_t^2 = (1-\lambda)y_{t-1}^2 + \lambda\hat{\sigma}_{t-1}^2$$

- It is possible to estimate EWMA with the methods discussed later in this Chapter
  - Since the EWMA is a reduced GARCH model, we could simply estimate λ with maximum likelihood (see the GARCH discussion below)
- JP Morgan set for daily data

$$\lambda = 0.94$$

- And when we use the term EWMA we are assuming that value

Unconditional EWMA Volatility

- The EWMA is a reduced GARCH model
- So, using an equation shown later, we get that the unconditional volatility is

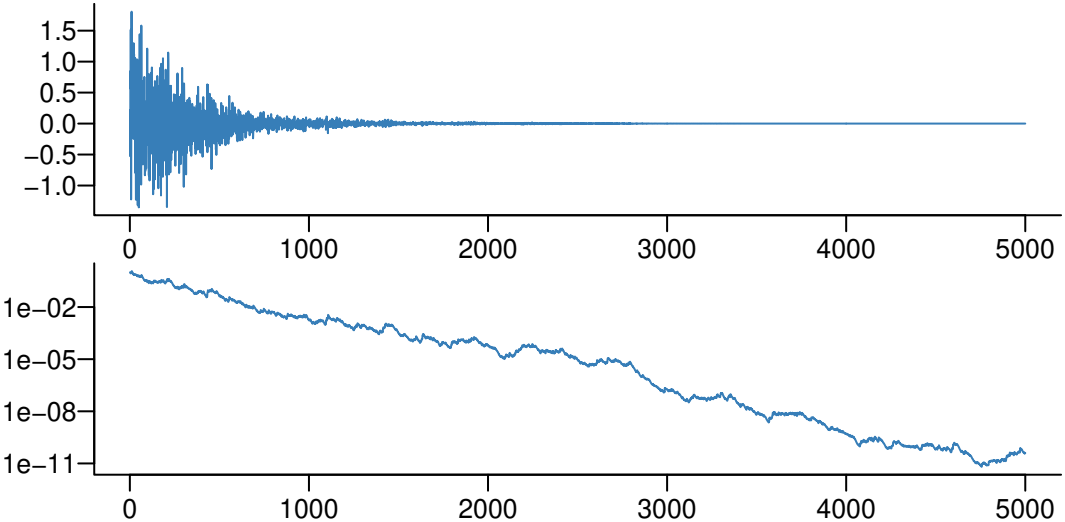
$$\sigma^2 = \frac{0}{0}$$

- In other words, undefined
- We can explore that more with simulations

Simulation Analysis

```
N=1000
y=rnorm(N)
lambda=0.94
sigma2=vector(length=N)
sigma2[1]=1
for(i in 2:N){
  sigma2[i]=lambda*sigma2[i-1]+(1-lambda)*y[i-1]^2
  y[i]=y[i]*sqrt(sigma2[i])
}
plot(y,type='l')
plot(sigma2,type='l',log='y')
```

Simulate EWMA



Conclusion From Simulation

- As we simulated longer and longer paths
- The σ become smaller and smaller
- Eventually zero
- So the unconditional volatility is undefined

EWMA Pros and Cons

- Pros
  - It is really simple to implement
  - And not that inaccurate compared to the more sophisticated GARCH
  - And multivatiate (see Chapter 3) versions are really easy
- Cons
  - By definition it is less accurate than GARCH
  - Which can become important in some cases
  - The unconditional volatility is not defined, which can be a problem

The ARCH Family

ARCH

- Robert Engle proposed a model in 1982 called autoregressive conditionally heteroscadastic (ARCH)
- Most volatility models derive from this
- Returns have a *conditional* distribution (here assumed to be normal)

$$Y_t \sim \mathcal{N}(0, \sigma_t^2)$$

- Or we can write:

$$Y_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, 1)$$

- Where ϵ_t is called *residual*

ARCH

- The volatility is weighted average of past returns
- ARCH(L_1)

$$\sigma_t^2 = \omega + \sum_{i=1}^{L_1} \alpha_i y_{t-i}^2$$

- The number of lags is L_1
- The most common form is *ARCH(1)*

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

- ω, α are parameters to be estimated with maximum likelihood

Moments

- Moments are the expected value of a random variable to some power
- So the mth moment is:

$$\mathrm{E}[X^m] = \int_{-\infty}^{\infty} x^m f(x) dx$$

- The mean is the first moment

$$\mu = \mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) dx$$

- The variance is a function of the first and the second moment

$$\mathrm{E}[(X-\mu)^2] = \mathrm{E}[X^2] + \mu^2 - 2\mu\,\mathrm{E}[X]$$

Some Mathematics of Moments

- The expected value (unconditionally) of a time dependent variable Y_t is:

$$\mathrm{E}(Y^m) = \mathrm{E}\left(\mathrm{E}_t(Y^m)\right) = \mathrm{E}(Y_t^m)$$

for all t. Therefore when $\mu = 0$:

$$E(Y^2) = \sigma^2 = \mathrm{E}\left(Y_t^2\right)$$

Some Mathematics of Moments (cont.)

- If we write that in terms of residuals

$$\mathrm{E}(Y^2) = \mathrm{E}\left(\sigma_t^2 \epsilon_t^2\right) = \mathrm{E}\left(\sigma_t^2\right)$$

- Then:

$$\sigma^2 = \mathrm{E}(\omega + \alpha Y_{t-1}^2) = \omega + \alpha\sigma^2$$

- Because the parameters are constant
- So, the *unconditional* volatility of the ARCH(1) model is:

$$\sigma^2 = \frac{\omega}{1-\alpha}$$

Fat Tails

- Recall the discussion in the last chapter which said that a distribution is fat-tailed if it has more extreme outcomes than a normal with the same mean and variance
- Kurtosis is normalised fourth moment

$$\text{Kurtosis} = \frac{\mathrm{E}(Y^4)}{(\mathrm{E}(Y^2))^2}$$

- Kurtosis is 3 for the normal
- Regardless of what the mean and variance are

ARCH(1) Fat Tails

- The most common distributional assumption for residuals ϵ is standard normality; that is:

$$\epsilon_t \sim N(0,1)$$

- Because ϵ_t is normal, so must conditional returns, $\sigma_t \epsilon_t$ because σ_t is a constant
- What about the *unconditional* distribution of the returns?

ARCH(1) Fat Tails (cont.)

- The 4th moment is
$$\mathrm{E}(Y^4) = \mathrm{E}(Y_t^4) = \mathrm{E}(\sigma_t^4 \epsilon_1^4)$$
- Because of the normality of ϵ_t and because it has variance one
$$\mathrm{E}[\epsilon_t^4] = 3$$
- Then the 4th moment of Y_t is
$$3\,\mathrm{E}[\sigma_t^4]$$

ARCH(1) Fat Tails (cont.)

- Recall the definition of kurtosis

$$\text{Kurtosis} = \frac{\mathrm{E}\left(Y^4\right)}{\left(\mathrm{E}\left(Y^2\right)\right)^2} = \frac{\mathrm{E}\left(Y^4\right)}{\sigma^4}$$

- Plug in the ARCH parameters

$$\begin{aligned}
\mathrm{E}(Y^4) &= 3\,\mathrm{E}\left(\left(\omega + \alpha Y_{t-1}^2\right)^2\right) \\
&= 3\omega^2 + 6\alpha\omega\,\mathrm{E}(Y^2) + 3\alpha^2\,\mathrm{E}(Y^4) \\
&= 3\omega^2 + 6\alpha\omega\frac{\omega}{1-\alpha} + 3\alpha^2\,\mathrm{E}(Y^4)
\end{aligned}$$

ARCH(1) Fat Tails (cont.)

- Then

$$
\begin{aligned}
\mathrm{E}(Y^4)(1-3\alpha^2) &= 3\omega^2 + 6\alpha\omega\frac{\omega}{1-\alpha} \\
&= 3\frac{\omega^2(1+\alpha)}{1-\alpha}
\end{aligned}
$$

ARCH(1) Fat Tails (cont.)

- Solve for

$$\begin{aligned} \mathrm{E}(Y^4) &= \frac{3\omega^2(1+\alpha)}{(1-\alpha)(1-3\alpha^2)} \\ &= \frac{3\sigma^4(1-\alpha^2)}{1-3\alpha^2} \end{aligned}$$

- If that exceeds three then Y_t must be fat-tailed

$$\text{Kurtosis} = \frac{3\left(1-\alpha^2\right)}{1-3\alpha^2} > 3 \quad \text{if } 3\alpha^2 < 1.$$

Extreme Value Analysis (See Chapter 9)

- The table below shows that the the higher the α, the fatter the tails (lower tail index ι implies fatter tails)

α	0.10	0.50	0.90	0.99
ι	26.48	4.73	2.30	2.02

Parameter Restrictions for ARCH(1)

- The ARCH(1) model has two parameters, ω and α
- Can we allow those two to take any value on the real line?
- No; there are two restrictions on the values the parameters can take
- One we always impose, and the other sometimes
- To ensure positive volatility, ensure both parameters are positive

$$\alpha > 0, \quad \omega > 0$$

Preventing Explosions – Stationarity

- Suppose $\alpha > 1$ then we expect σ_t to become bigger and bigger over time
- Which would mean that the unconditional variance is undefined

$$\sigma^2 = \frac{\omega}{1-\alpha}$$

- Because, something to the power 2 cannot be negative (these are not complex numbers)
- We might think therefore to restrict α to be less than one

$$0 < \alpha < 1$$

- This is not needed except in special circumstances (see discussion on GARCH below)

Generalised ARCH (GARCH)

Generalised ARCH (GARCH)

- As it turns out, ARCH is not a very good model and almost nobody uses it
- The reason is that it needs to use information from many days before t to calculate volatility on day t
- That is, it needs a lot of lags
- The solution is to write it as an ARMA model
- That is, add one component to the equation, $\beta\sigma_{t-1}^2$

GARCH Equation

GARCH(L_1,L_2)

$$\sigma_t^2 = \omega + \sum_{i=1}^{L_1} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{L_2} \beta_j \sigma_{t-j}^2$$

GARCH(1,1)

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

- GARCH(1,1) is the most common specification

GARCH(1,1) Unconditional Volatility

- The unconditional volatility is the unconditional expectation of volatility on a given day

$$\sigma^2 = \mathrm{E}[\sigma_t^2]$$

- So plug in the parammeters

$$\sigma^2 = \mathrm{E}(\omega + \alpha Y_{t-1}^2 + \beta \sigma_{t-1}^2) = \omega + \alpha \sigma^2 + \beta \sigma^2$$

- And solve

$$\sigma^2 = \frac{\omega}{1 - \alpha - \beta}$$

Parameter Restrictions

- To ensure positive volatility forecasts

$$\omega, \alpha, \beta \geq 0$$

- Because if any parameter is negative, σ_{t+1} may be negative

Stationarity

- Should we impose

$$\alpha + \beta < 1$$

- Because

$$\sigma^2 = \frac{\omega}{1 - \alpha - \beta}$$

- Not advisable except in special circumstances for two reasons
  1. Can lead to multiple parameter combinations satisfying the constraint so volatility forecasts can be non-unique
  2. Model is misspecified, and the non-restricted model could give more accurate forecasts

EWMA Unconditional Variance

- EWMA is
$$\sigma_t^2 = (1-\lambda)y_{t-1}^2 + \lambda\sigma_{t-1}^2$$
- The GARCH model is
$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2$$
- And GARCH becomes EWMA when $\omega = 0, \beta = \lambda, \alpha = 1-\lambda$
- So for EWMA
$$\sigma^2 = \frac{0}{0}$$

Meaning of The GARCH Parameters

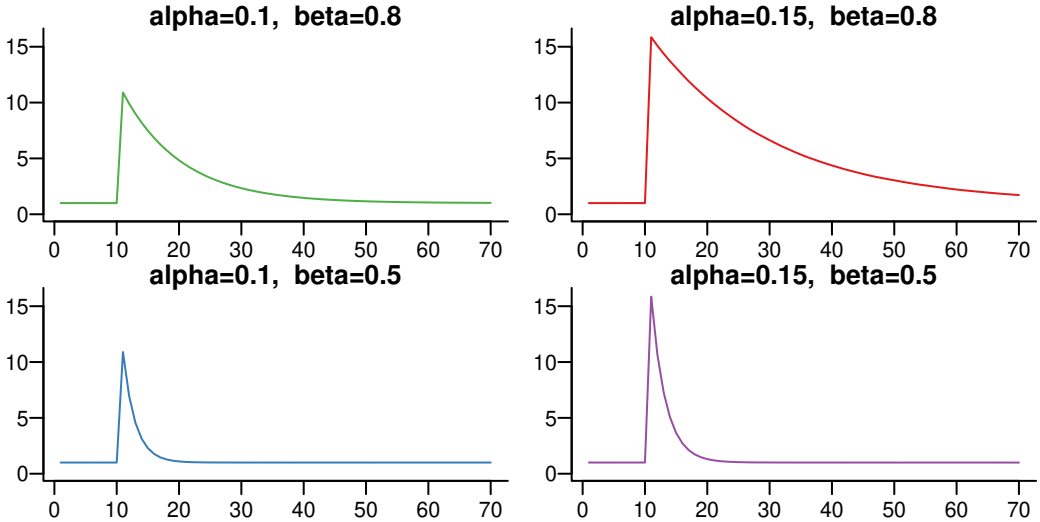
Meaning of Parameters

- α is news – how volatility reacts to new information
- β is memory – how much volatility remembers from the past
- While GARCH looks like an ARMA model, it is not, note:

$$\alpha y_{t-1}^2 = \alpha \sigma_{t-1}^2 \epsilon_{t-1}^2$$

- The size of $(\alpha + \beta)$ determines how quickly the predictability (memory) of the process dies out:
- If $(\alpha + \beta)$ is close to zero, predictability will die out very quickly
- If $(\alpha + \beta)$ is close to one, predictability will die out slowly

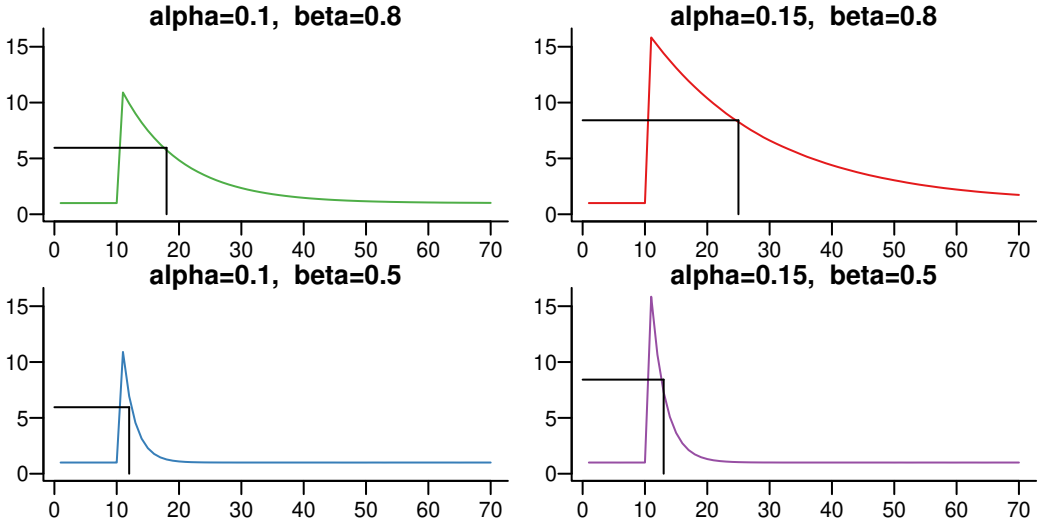
GARCH News and Memory



Half-Life

- In physics, half-life indicates how quickly a radioactive material decays to half the radiation
- The half-life of GARCH is how quickly a shock to volatility dies out to half the impact of the shock
- You can see that labelled on the next figure

GARCH News and Memory



GARCH Half-Life

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

As

$$E_{t-1}[y_t^2] = \sigma_t^2$$

Then given $t-1$ and expectation

$$y_t^2 = \omega + (\alpha + \beta) y_{t-1}^2$$

Subtract $\sigma^2 = \omega/(1-\alpha-\beta)$ from both sides

$$\begin{aligned} y_t^2 - \sigma^2 =& \omega + (\alpha+\beta) y_{t-1}^2 - \frac{\omega}{1-\alpha-\beta} \\ =& (\alpha+\beta)(y_{t-1}^2 - \sigma^2) \end{aligned}$$

GARCH Half-Life (cont.)

- Suppose we are interested in σ_{t+n} as a function of y_{t+1}
- Then

$$y_{t+n}^2 - \sigma^2 = (\alpha + \beta)^{n-1}(y_{t+1}^2 - \sigma^2)$$

- The number of periods, n^*, it takes for conditional variance to revert back halfway towards unconditional variance

$$\sigma_{t+n^*}^2 - \sigma^2 = \frac{1}{2}(\sigma_{t+1}^2 - \sigma^2)$$

- So we want to solve for n^* as a function of the parameters

GARCH Half-Life (cont.)

- So

$$(\alpha+\beta)^{n^*-1}(\sigma_{t+1}^2-\sigma^2)=\frac{1}{2}(\sigma_{t+1}^2-\sigma^2)$$

- So

$$n^*=1+\frac{\log\left(\frac{1}{2}\right)}{\log(\alpha+\beta)}$$

- And as $(\alpha+\beta)\longrightarrow 1$, $n^*\to\infty$, memory is infinite

Other GARCH-type models

Conditional Distributions

- The conditional distribution in the GARCH model in the previous section is the normal

$$\epsilon_t \sim \mathcal{N}(0,1)$$

- We showed in Chapter 1 that unconditional returns are fat and we also showed in a previous section that the unconditional returns in GARCH are fat even if the conditional distribution is normal
- That leaves the question of whether they are fat enough, which becomes important in the risk measure chapters later in the book
- We can make the GARCH model *fatter* by using a different conditional distribution

Student-t GARCH (tGARCH)

- As we discussed in the previous chapter, the Student-t distribution is fat, where the degrees of freedom parameter – ν – controls the *conditional fatness*
- If $\nu = \infty$ the Student becomes the normal
- The Student-t GARCH then replaces the normal innovation distribution with the Student

$$\epsilon_t \sim t_{(\nu)}$$

- When it comes to estimation, ν becomes yet another parameter to be estimated along with the three GARCH parameters

Standardised or Non-Standardised t

- An important complication is that by convention the Student-t in tGARCH uses the *standardised* Student-t density, that is, one where it is standardised to have *variance=1*
- While the Student-t density density has *variance*$=\nu/\nu-2$
- That becomes important later
- The second moment (variance) is not defined for $\nu \leq 2$, so the above needs $\nu > 2$

- There are two downsides to tGARCH
  1. The tGARCH needs more observations in estimation than the normal GARCH, usually at least several thousand
  2. The ν parameter is often estimated with high standard error so it is imprecise and can move around when we use estimation windows in risk forecasting

What About the Mean?

- In the standard GARCH model the mean is assumed to be zero

$$Y_t \sim (0, \sigma_t^2)$$

- What is often done in practice is to subtract the mean (de-mean) from the returns prior to estimation
- Usually best for risk applications
- However, in other applications, like price forecasting, it can be beneficial to forecast the mean
- While there are many complicated ways to do so, two are relatively simple and common

(G)ARCH in Mean

- Incorporating a mean

$$Y_t = \mu_t + \sigma_t \epsilon_t$$

- Return can be correlated with volatility

$$\mu_t = a\sigma_t$$

- Where a captures the impact of volatility on the mean
- And is estimated as an extra parameter

(G)ARCH in Mean With ARMA

- We can also make the mean follow some autoregressive process like ARMA (autoregressive–moving-average)

$$\mu_t = a + by_{t-1} + c\sigma_{t-1}^2$$

- This is build into the rugarch package

Leverage Effect

- If the price of equity falls, the company's *debt to equity* ratio increases and the company becomes riskier as a consequence
- We might, therefore, expect the volatility to increase
- That is known as the *leverage effect*
- Stock returns are negatively correlated with changes in volatility
- However, the standard GARCH model assumes symmetry

- So separate out the impact of the positive and negative returns with an extra parameter, ζ

$$\sigma_t^2 = \omega + \alpha \left(|Y_{t-1}| - \zeta Y_{t-1} \right)^2 + \beta \sigma_{t-1}^2$$

- If $\zeta = 0$ this model reduces to the standard GARCH model

Power Effect

- In the standard GARCH model, the power on lagged returns and volatility is 2

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

- However, there is no reason to believe that it should be 2, and it is sometimes found in estimation that the GARCH model is improved if the power is different
- In the *power GARCH* we also estimate the parameter for the power, δ

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

- If

$$\delta \neq 2$$

- The model has power effects

Asymmetric Power GARCH – apARCH

- These two effects come together in the *asymmetric power* GARCH, or apARCH, model

$$\sigma_t^2 = \omega + \alpha \left(|Y_{t-1}| - \zeta Y_{t-1} \right)^\delta + \beta \sigma_{t-1}^\delta$$

- The model allows for leverage effects when $\zeta \neq 0$ and power effects when $\delta \neq 2$
- This model can be difficult to estimate and typically requires thousands of observations

News Impact Curve

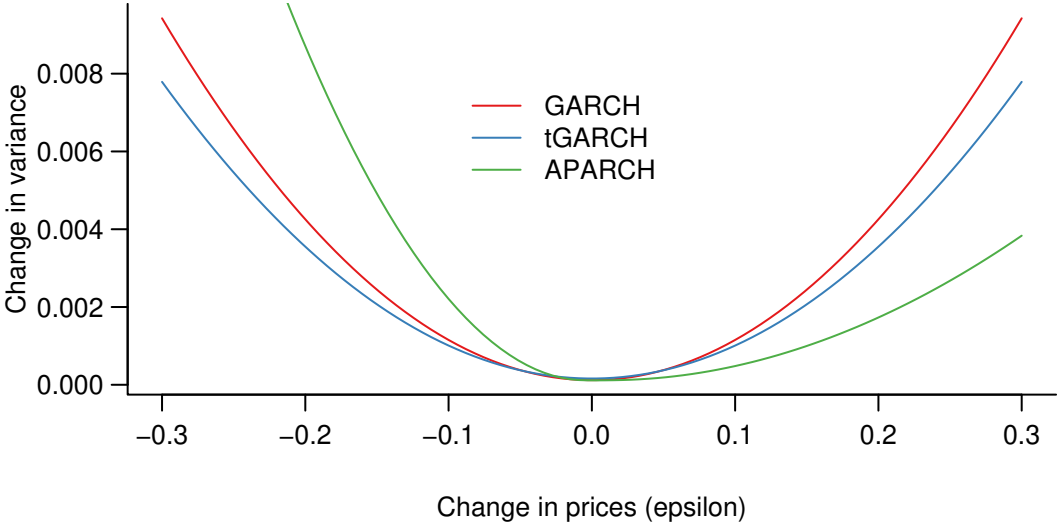
- The news impact curve captures the relationship between shocks at time $t-1$ (ϵ_{t-1}) to the variance at time t, σ_t^2

S&P-500

- 10,000 observations
- 12 November 1980 to 9 July 2020
- We discuss the significance a bit later

	ω	α	β	ζ	δ	ν	Half-life
1	0.0000018	0.09	0.89				48
2	0.0000010	0.08	0.91			5.9	110
3	0.0000001	0.06	0.89	0.35	2.6		47

GARCH News and Memory (same S&P-500)



Maximum Likelihood

Maximum Likelihood Estimation

- A linear model is

$$y = a + bx + \epsilon$$

- Where ϵ relates linearly to the dependent variable y
- Why we can do ordinary least squares (OLS)
- But GARCH is *non-linear*
- Do maximum likelihood
- Consistent parameter estimates even if true density is non-normal

What Is Maximum Likelihood?

- We ask the question: Which parameters most likely generated the data?
- Suppose we have a sample of

$$(-0.2, 3, 4, -1, 0.5)$$

- Of the three possibilities, which is most likely?

	μ	σ
1	1	5
2	-2	2
3	1	2

Normal Density

- If:

$$x \sim \mathcal{N}(\mu, \sigma^2)$$

- Then its density is:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$$

ARCH(1)

- ϵ_t in ARCH(1) is standard normally distributed

$$
\begin{aligned}
Y_t &= \sigma_t \epsilon_t \\
\sigma_t^2 &= \omega + \alpha Y_{t-1}^2 \\
\epsilon_t &\sim \mathcal{N}(0,1)
\end{aligned}
$$

ARCH(1) Density

- Presence of lagged returns implies the density function for $t = 1$ is unknown since y_0 is unknown
- The $t = 2$ density is:

$$f(y_2|y_1) = \frac{1}{\sqrt{2\pi(\omega + \alpha y_1^2)}} \exp\left(-\frac{1}{2}\frac{y_2^2}{\omega + \alpha y_1^2}\right)$$

- Joint density

$$\prod_{t=2}^{T} f(y_t|y_{t-1}) = \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi(\omega + \alpha y_{t-1}^2)}} \exp\left(-\frac{1}{2}\frac{y_t^2}{\omega + \alpha y_{t-1}^2}\right)$$

ARCH(1) Likelihood Function

- The likelihood function is then the density, except we condition the parameters $\theta = (\alpha, \beta)$ on the data (returns)

$$\log \mathcal{L}(\theta|\text{data})$$

- The log likelihood function is then

$$\log \mathcal{L} = \underbrace{-\frac{T-1}{2}\log(2\pi)}_{\text{Constant}} - \frac{1}{2}\sum_{t=2}^{T}\left(\log(\omega + \alpha y_{t-1}^2) + \frac{y_t^2}{\omega + \alpha y_{t-1}^2}\right)$$

GARCH(1,1)

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

$$f(y_2|y_1) = \frac{1}{\sqrt{2\pi(\omega + \alpha y_1^2 + \beta \hat{\sigma}_1^2)}} \exp\left(-\frac{1}{2}\frac{y_2^2}{\omega + \alpha y_1^2 + \beta \hat{\sigma}_1^2}\right)$$

$$\log \mathcal{L} = \underbrace{-\frac{T-1}{2}\log(2\pi)}_{\text{Constant}}$$

$$-\frac{1}{2}\sum_{t=2}^{T}\left((\log(\omega + \alpha y_{t-1}^2 + \beta \hat{\sigma}_{t-1}^2)) + \frac{y_t^2}{\omega + \alpha y_{t-1}^2 + \beta \hat{\sigma}_{t-1}^2}\right)$$

Importance of σ_1

- Value of σ_1 can make a large difference
- Especially when the sample size is small
- Typically set $\sigma_1 = \hat{\sigma}$

Volatility Targeting

- Since we know that

$$\sigma^2 = \frac{\omega}{1 - \alpha - \beta}$$

- We can set

$$\omega = \hat{\sigma}^2(1 - \alpha - \beta)$$

- Where $\hat{\sigma}^2$ is the sample variance
- And save one parameter in the estimation

Issues in Estimation

- Parameters are obtained by maximising the likelihood function

  θ The set of parameters

  Θ The space the parameters live in

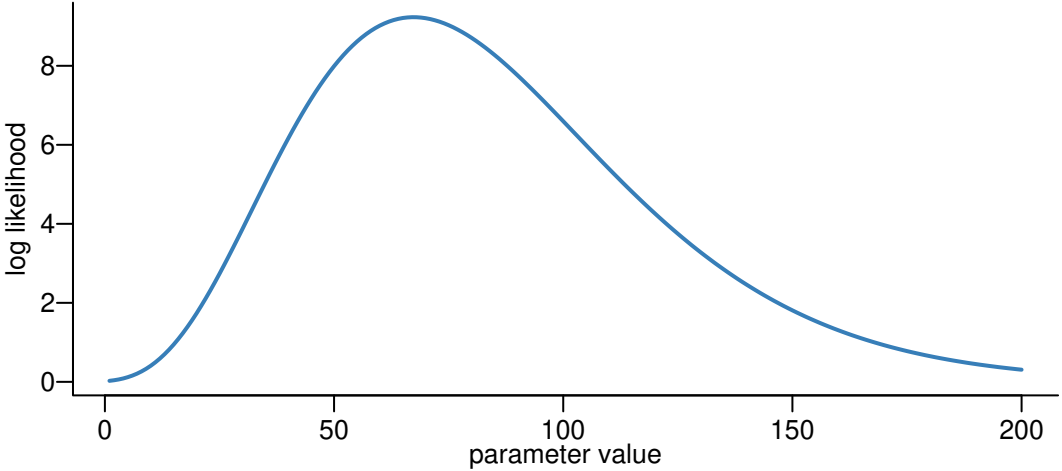
$$\max_{\theta \in \Theta} \log \mathcal{L}(\theta|y) = \hat{\theta}_{\text{ML}}$$

- Done with an algorithm called *optimiser*
- There often are *numerical problems*

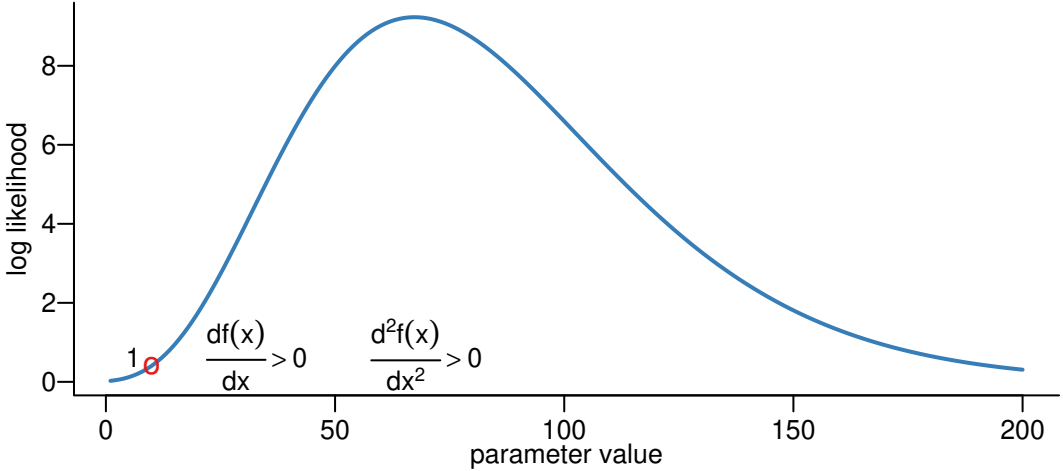
Optimising

- The computer uses an algorithm called optimiser to maximise the likelihood function
- There are many optimisers available
- A classical, and not the best, optimiser is *Newton-Raphson*

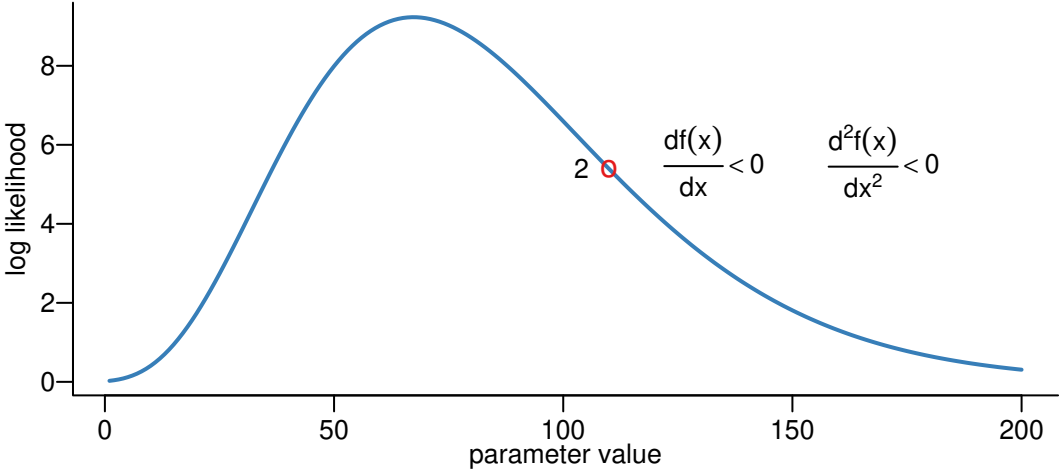
Newton-Raphson



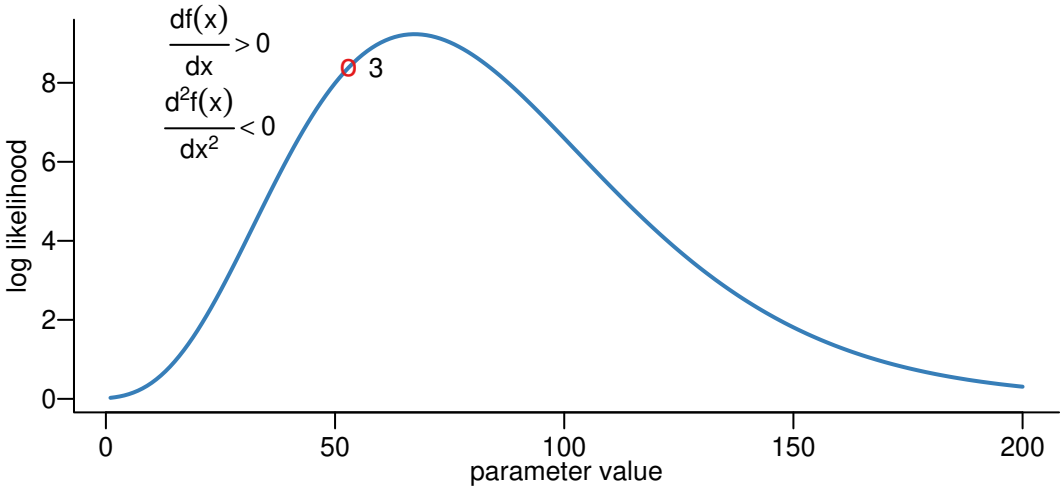
Newton-Raphson



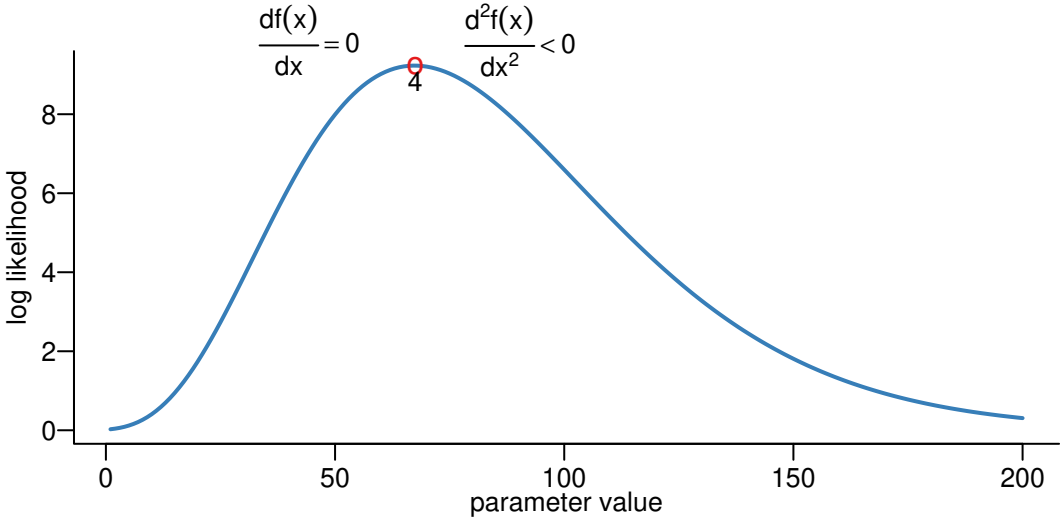
Newton-Raphson



Newton-Raphson



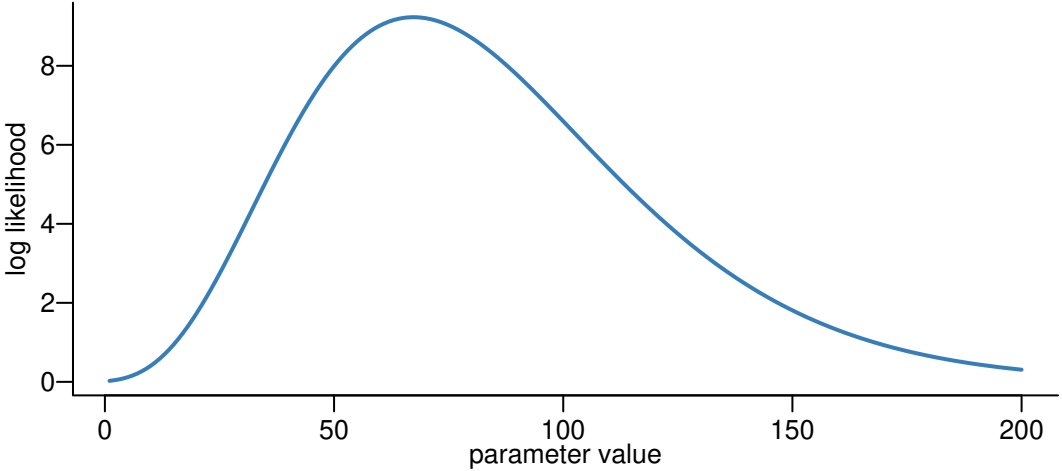
Newton-Raphson



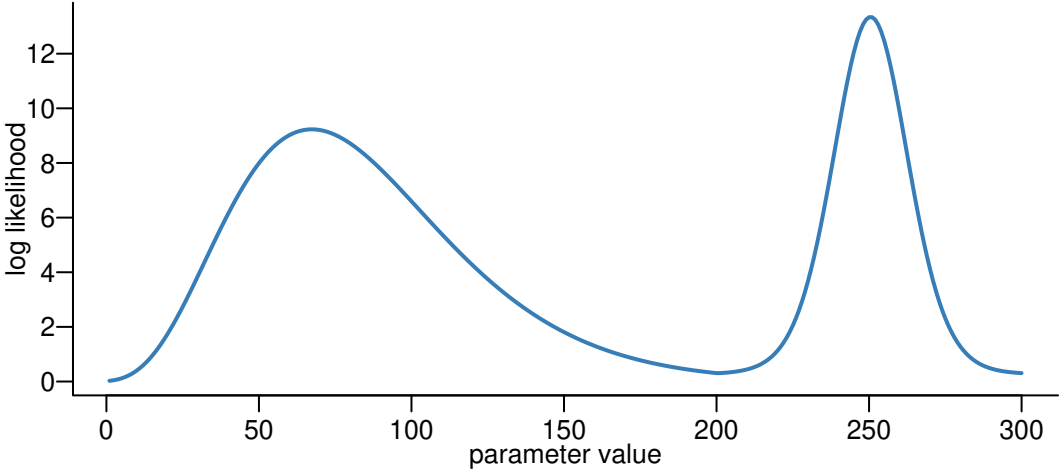
Pathologies

- It is straightforward to optimise the ARCH model
- And the GARCH
- For other models they can be a number of problems
- A common one is multiple local maxima

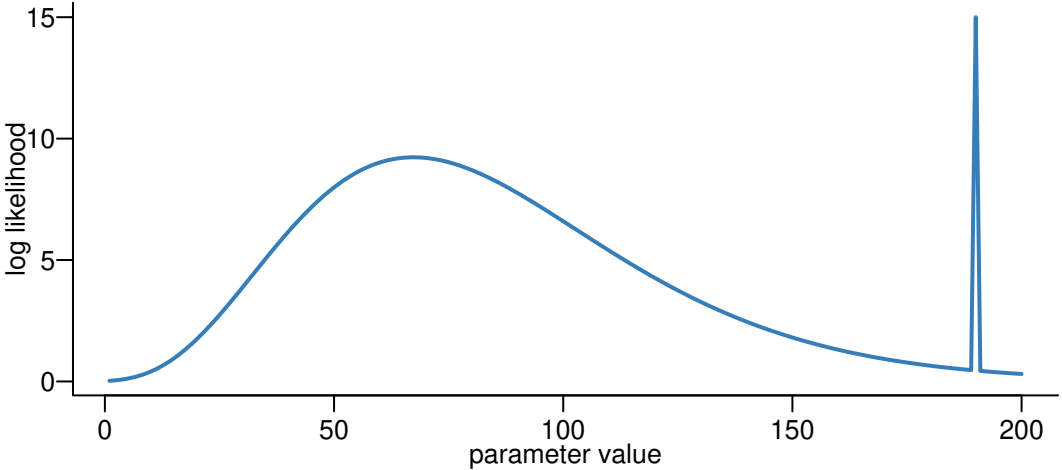
Pathologies



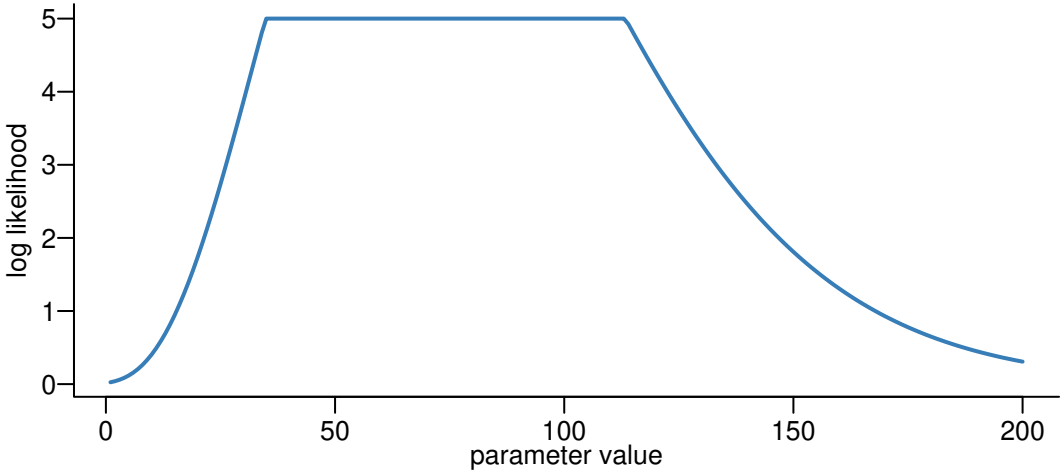
Pathologies



Pathologies



Pathologies



Issues

- Problems are rare for smaller models such as GARCH(1,1)
- The more parameters there are, the more likely problems are
- And the more data we need for estimation
- GARCH might need at least 500 observations, Student-t GARCH 3,000
- For the multivatiate models discussed in the next Chapter, estimation problems are common
- For example, multiple local minima

Estimation

Estimation Options

- At least two R packages support estimating GARCH-style models
- The best is `rugarch` by Alexios Ghalanos
- cran.r-project.org/web/packages/rugarch
- Provides useful plots (try plotting the result)
- And useful information when printing result

Specifying Models With `rugarch`

- There are a large number of GARCH-type models and specification
- `rugarch` separates out model specification and estimation
- `ugarchspec` creates model, some common options are
  - `variance.model`
  - `mean.model`
  - `distribution.model`
- Then model is estimated with `ugarchfit`
- It returns a structure with estimated parameters, fitted volatility, log likelihood, and other useful results

variance.model

- Input: list containing the variance model specification
- The standard GARCH is "sGARCH" (default). Another is "apARCH"
- garchOrder. The ARCH and GARCH orders (lags)
- external.regressors
- variance.targeting
- variance.model = list(garchOrder = c(1, 1))

mean.model

- List containing the mean model specification
- It is possible to make it a constant or use a richer model
- For now, we just tell the code not to do it
- `mean.model = list( armaOrder = c(0,0),include.mean = FALSE))`

distribution.model

- The conditional density to use for the innovations
- The default is "`norm`" for the normal distibution
- We also use "`std`" for the Student-t
- But there are many others that could be used
- `distribution.model = "std"`

R Estimation With rugarch

```
library(rugarch)

spec1 = ugarchspec(variance.model = list( garchOrder = c(1, 1)),
           mean.model = list( armaOrder = c(0,0),include.mean = FALSE))
res1 = ugarchfit(spec = spec1, data = y)

spec2 = ugarchspec(variance.model = list( garchOrder = c(1, 0)),
           mean.model = list( armaOrder = c(0,0),include.mean = FALSE))
res2 = ugarchfit(spec = spec2, data = y)

spec3 = ugarchspec(variance.model = list( garchOrder = c(1, 1)),
           mean.model = list( armaOrder = c(0,0),include.mean = FALSE),
 distribution.model = "std")
res3 = ugarchfit(spec = spec3, data = y)

plot(res3)
```

Choosing Optimisers

- `rugarch` allows us to choose optimisers, needed when we set 1,0 in `garchOrder`

```
spec4 = ugarchspec(variance.model = list( garchOrder = c(1, 0)),
mean.model = list( armaOrder = c(0,0),include.mean = FALSE),
distribution.model = "std")
res4 = ugarchfit(spec = spec4, data = y)
```

```
Warning message: In .sgarchfit(spec = spec, data = data, out.sample
= out.sample, : ugarchfit-->warning: solver failer to converge.
```

- But

```
res4 = ugarchfit(spec = spec4, data = y,solver="hybrid")
```

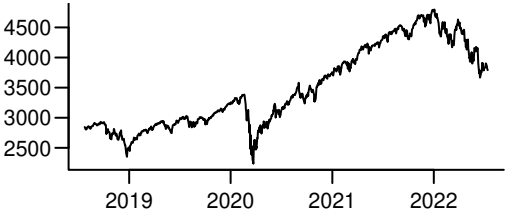
- Works
- The "hybrid" strategy solver first tries the "solnp" solver, in failing to converge tries "nlminb", then "gosolnp" and finally "nloptr" solvers

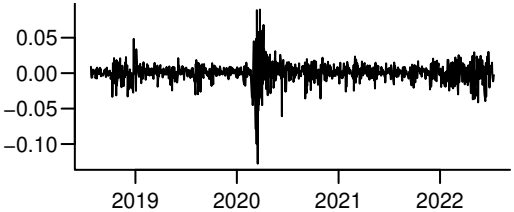
Simulation Analysis of GARCH Models

Simulation Analysis of GARCH Models

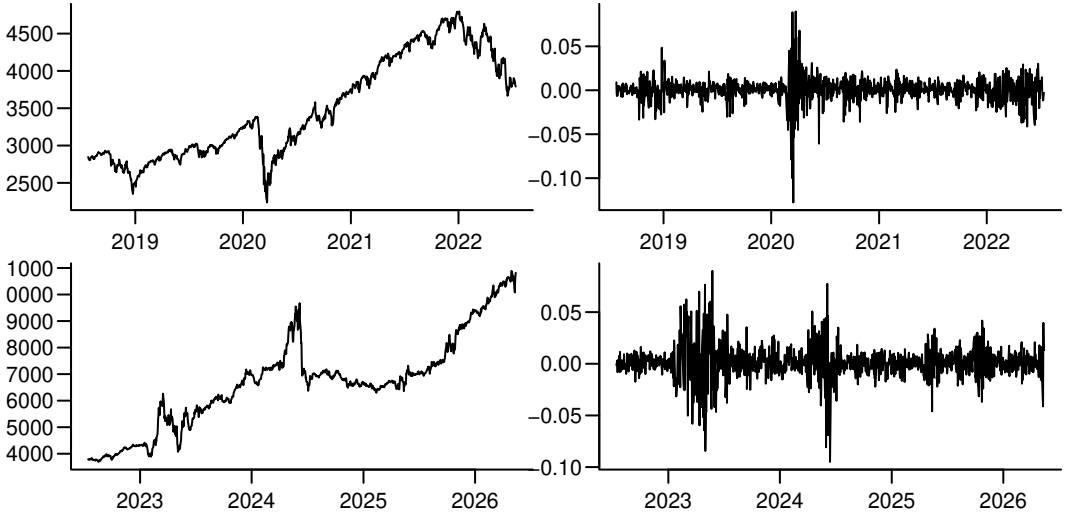
- We use thousand observations of the S&P-500
- First, estimate a normal GARCH model and include the mean effects

S&P-500


4500
4000
3500
3000
2500
2019
2020
2021
2022


0.05
0.00
−0.05
−0.10
2019
2020
2021
2022

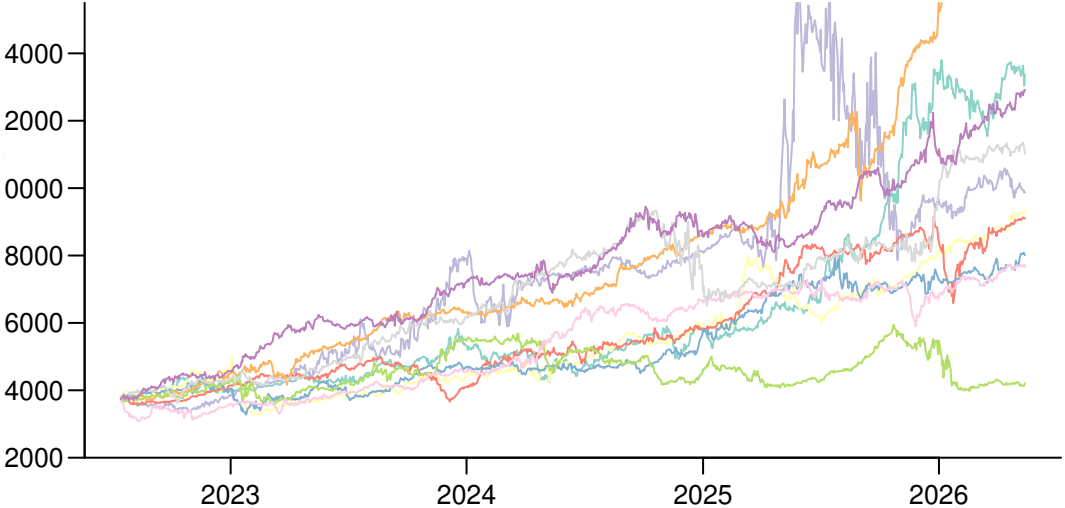
S&P-500 With One Simulation



Summary

- The simulated data looks quite different from the true data
- Of course, it should be different but what we're looking for is the dynamic patterns in it
  - Booms and busts
  - Volatility clusters
- The following plot shows ten simulations

S&P-500 With Ten Simulations



Ten Parameter Estimates From True and Simulated Data, Normal GARCH

	ω	α	β
True	3.8e-06	0.218	0.752
2	3.9e-06	0.272	0.727
3	4e-06	0.273	0.701
4	4.1e-06	0.281	0.718
5	2.8e-06	0.224	0.753
6	4.2e-06	0.246	0.716
7	3.1e-06	0.26	0.725
8	4.6e-06	0.271	0.703
9	2.4e-06	0.18	0.798
10	4.9e-06	0.257	0.712
11	6.9e-06	0.242	0.687

10 Parameter Estimates From True and Simulated Data, tGARCH

	ω	α	β	ν
True	2.2e-06	0.2	0.799	4.118
2	1.5e-06	0.19	0.809	3.901
3	2.2e-06	0.231	0.768	4.773
4	3.2e-06	0.252	0.735	4.317
5	2e-06	0.204	0.777	5.128
6	2.8e-06	0.192	0.807	4.139
7	2.1e-06	0.18	0.796	5.185
8	9e-07	0.117	0.881	3.928
9	2e-06	0.179	0.81	4.693
10	2.1e-06	0.237	0.762	4.32
11	2.1e-06	0.261	0.738	4.743

Diagnosing Volatility Models

Diagnosing Volatility Models

- Ideally we would want to know how a model performs operationally
- But usually we have to make do with an in-sample comparison
- In out-of-sample forecast comparisons, more parsimonious models perform better even if a more flexible model is significantly better in-sample
- If more flexible model performs better in-sample, it is unlikely to perform better out-of-sample

Alternatives

1. Likelihood ratio tests
2. Residual analysis

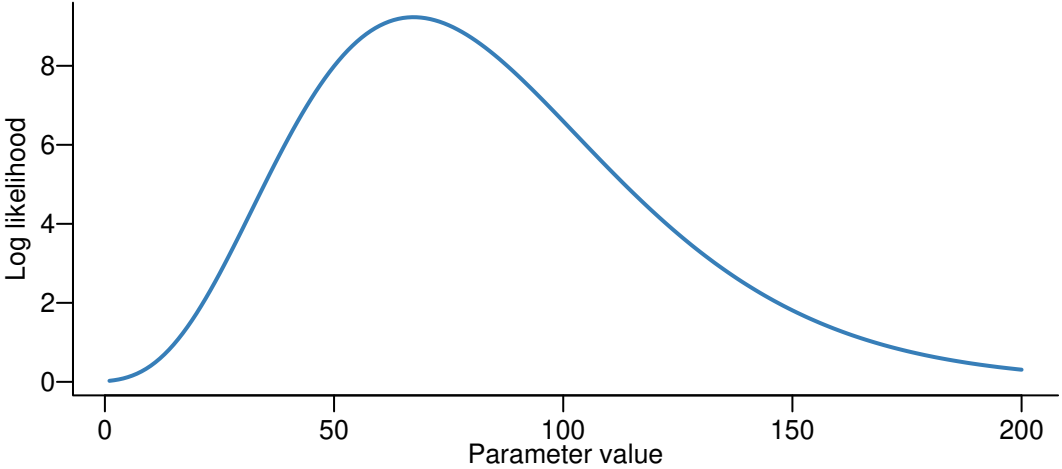
- Do not confuse the discussion here with backtesting as presented later

Likelihood Ratio Test

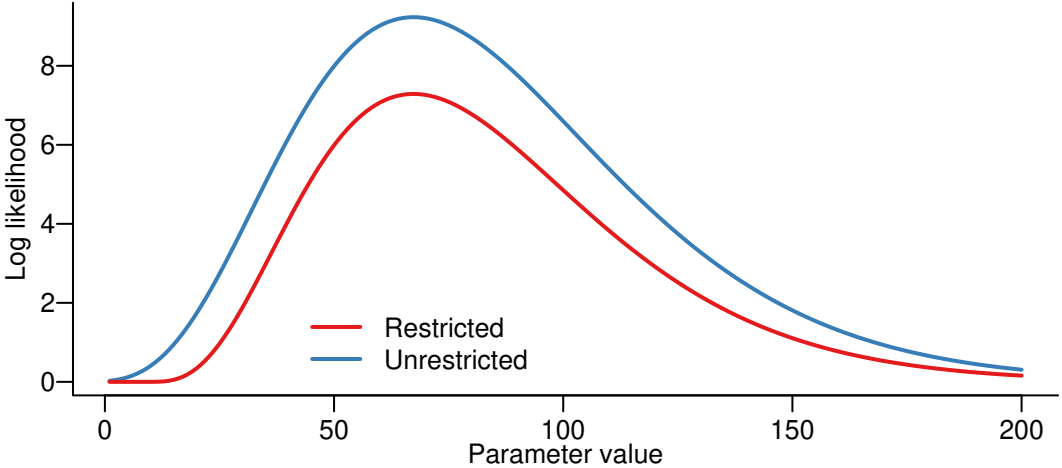
- Consider two models, where one is *nested* inside the other
- That is, one is a strict subset of the other
- Like ARCH(1) is a subset of GARCH(1,1)
- Is the same as a standard t-test for one parameter
- But also allows testing of multiple parameters
- Call the nested model *restricted* (R) and the other *unrestricted* (U)
- By definition

$$\log \mathcal{L}_R \leq \log \mathcal{L}_U$$

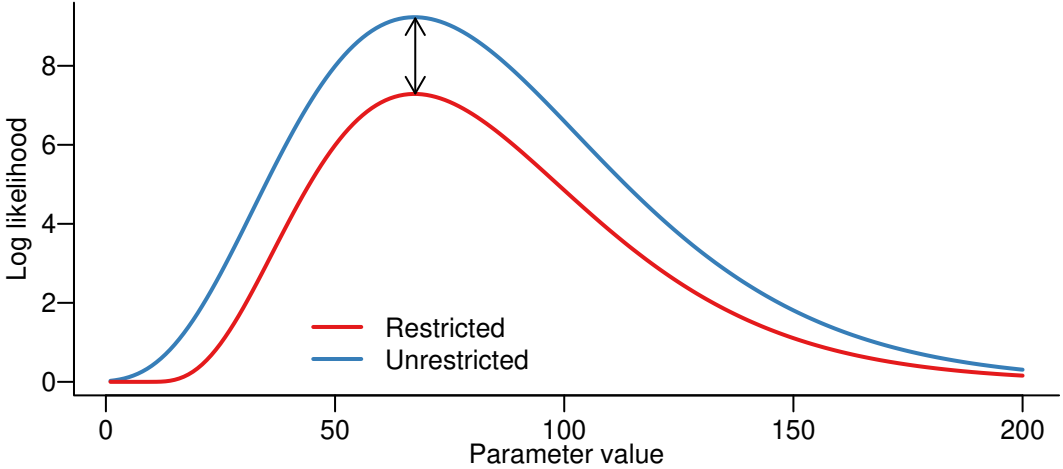
Likelihood Ratios



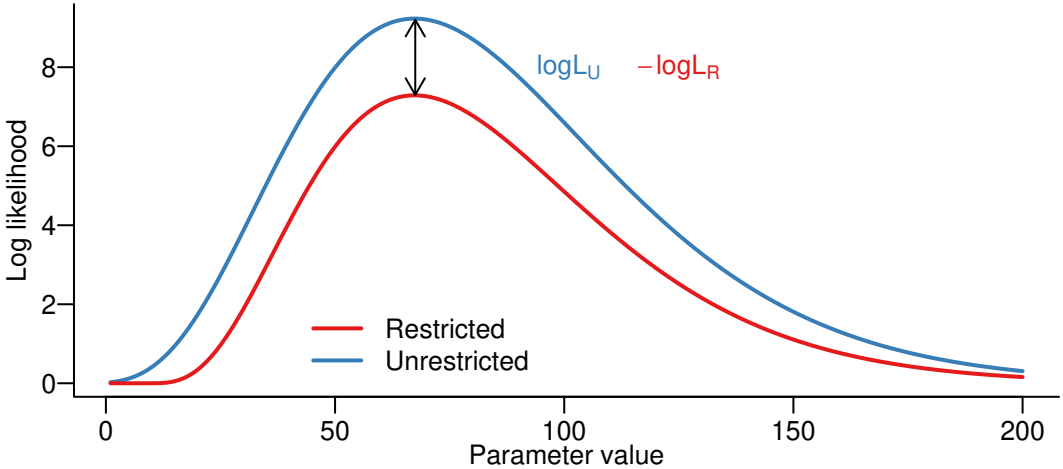
Likelihood Ratios



Likelihood Ratios



Likelihood Ratios



Likelihood Ratio Test

- Two times the difference between the unrestricted and restricted likelihoods is χ^2 distributed
- The degrees of freedom is equal to the number of restrictions

$$\text{LR} = 2(\log \mathcal{L}_U - \log \mathcal{L}_R) \sim \chi^2_{(\text{number of restrictions})}$$

```
qchisq(p=1-0.05,df=1)
 3.841459
```

χ^2 in R

```
x=seq(0.0,9,by=0.1)
plot(x,dchisq(x,3),type='l')
plot(x,pchisq(x,3),type='l')
plot(x,dchisq(x,1),type='l')
plot(x,pchisq(x,1),type='l')
```

Compare GARCH(1,1) to ARCH(1)

S&P-500 10,000 observations 12 November 1980 to 9 July 2020

```
spec0 = ugarchspec(variance.model = list( garchOrder = c(1, 0)),
  mean.model = list( armaOrder = c(0,0),include.mean = FALSE))
res0 = ugarchfit(spec = spec0, data = y,solver="hybrid")
likelihood(res0)

spec1 = ugarchspec(variance.model = list( garchOrder = c(1, 1)),
  mean.model = list( armaOrder = c(0,0),include.mean = FALSE))
res1 = ugarchfit(spec = spec1, data = y,solver="hybrid")
likelihood(res1)

LR=2 * (likelihood(res1)-likelihood(res0))
LR
740.9976
pvalue=1-pchisq(LR,1)
pvalue
0
```

Sample Tests

Unrestricted	Restricted	Test
ARCH(4)	ARCH(1)	$H_0 : \alpha_2 = \alpha_3 = \alpha_4 = 0$
GARCH(1,1)	ARCH(1)	$H_0 : \beta = 0$
GARCH(2,2)	GARCH(1,1)	$H_0 : \beta_2 = \alpha_2 = 0$
APARCH	GARCH	$H_0 : \delta = 2, \zeta = 0$

Sample Tests (cont.)

S&P-500 10,000 observations 12 November 1980 to 9 July 2020

Model	ω	α	β	ζ	δ	ν	lik
ARCH(1)	0.0000861	0.325					31347
GARCH(1,1)	0.0000018	0.095	0.891				32693
tGARCH(1,1)	0.0000010	0.079	0.915			5.926	33012
APGARCH(1,1)	0.0000001	0.06	0.889	0.352	2.573		32794

The tGARCH has the highest likelihood, followed by the APGARCH. It might be worthwhile to estimate a tAPGARCH, which might be significantly better than the others.

Residual Analysis

- Test if a single model is correctly specified
- Is based on the residuals (innovations) of the estimated model being what was assumed in the model specification
- For example, when estimating a normal GARCH, are the residuals IID normal?

Residual Analysis of GARCH(1,1)

- Returns have a distribution (here assumed to be normal)

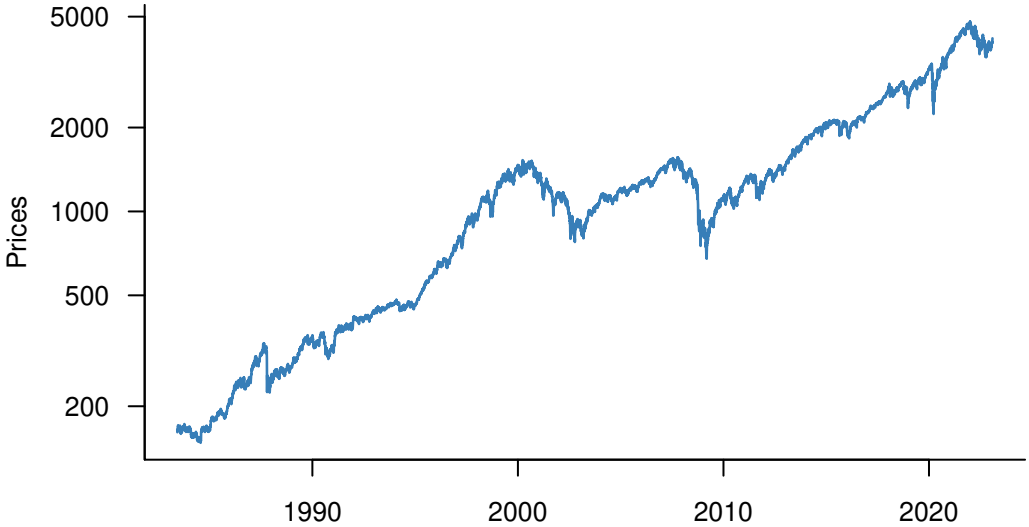
$$
\begin{aligned}
Y_t &\sim \mathcal{N}(0, \sigma_t^2) \\
Y_t &= \sigma_t \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0,1) \\
\sigma_t^2 &= \omega + \alpha Y_{t-1}^2 + \beta \sigma_{t-1}^2 \\
\hat{\epsilon}_t &= \frac{y_t}{\hat{\sigma}_t}
\end{aligned}
$$

- Can test the estimated residuals $\hat{\epsilon}$ with a Jarque-Bera test for normality and Ljung-Box test for autocorrelation of $\hat{\epsilon}$ and $\hat{\epsilon}^2$
- Can also do a QQ plot for $\hat{\epsilon}$

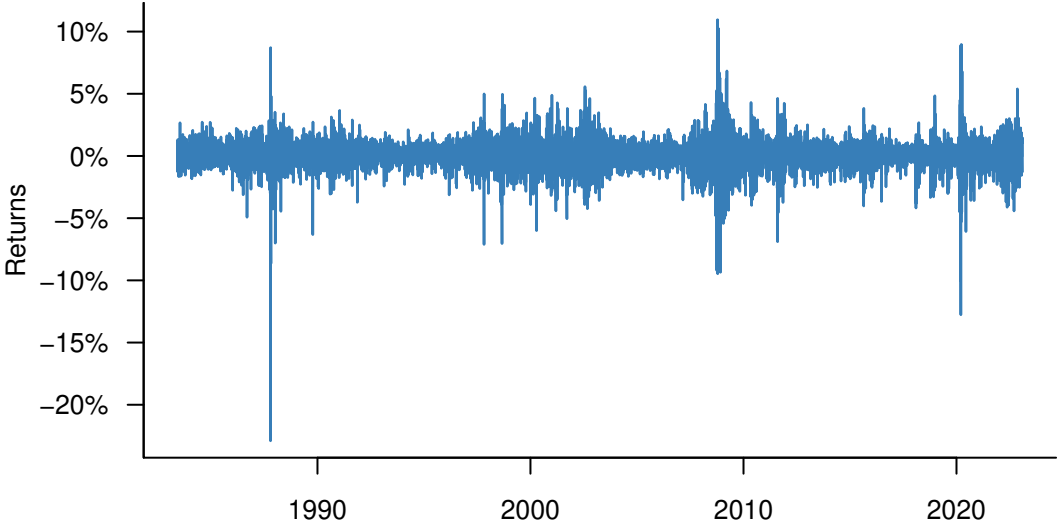
Residual Tests in R

```
residuals=coredata(y)/coredata(sigma(res1))
Box.test(residuals, lag = 20,
         type = c("Ljung-Box"))
library(tseries)
jarque.bera.test(residuals)
library(car)
qqPlot(residuals)
qqPlot(residuals,distribution="t",df=4)
```

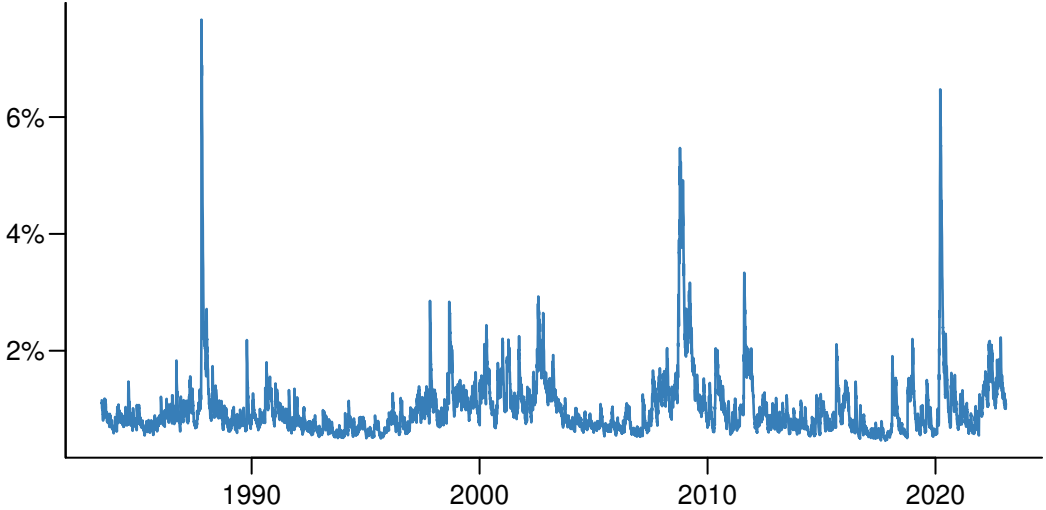
Price



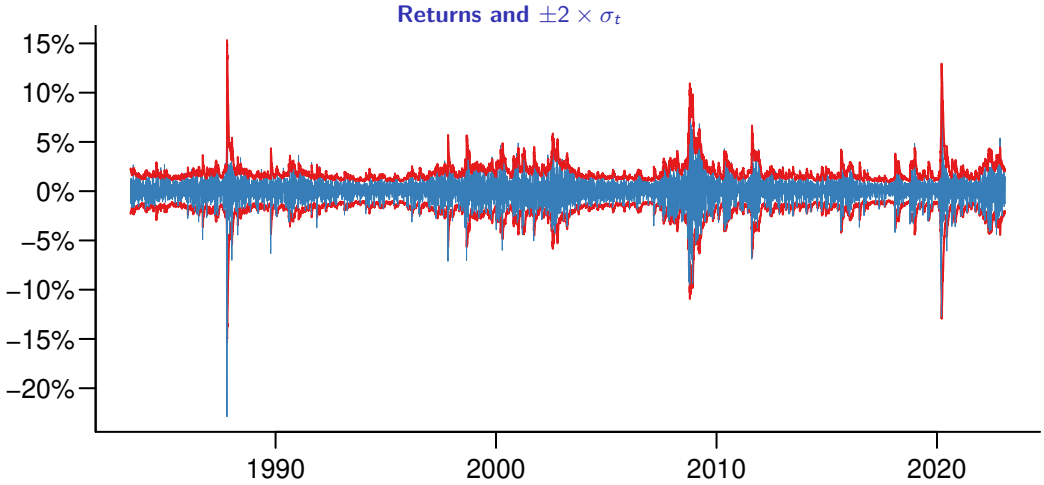
Return



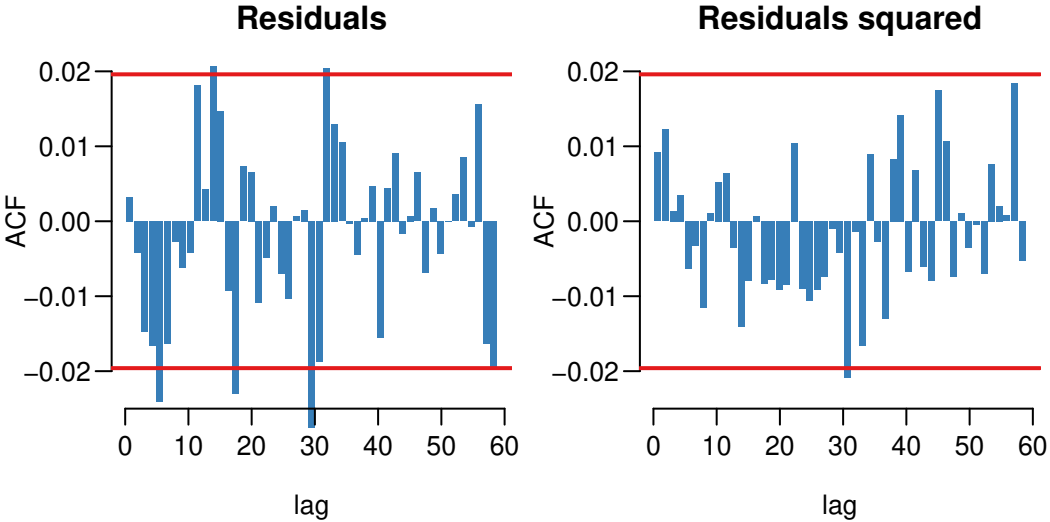
Conditional volatility (in-sample)



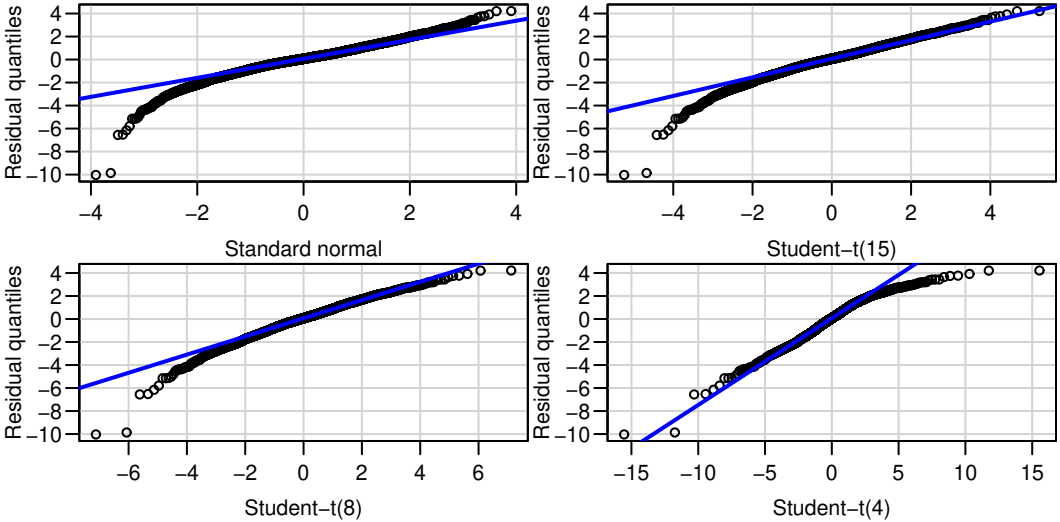
Conditional Volatility and Returns



ACF of Squared Residuals



QQ Plot of Residuals



Residual Analysis

$$\hat{\epsilon}_t = \frac{y_t}{\hat{\sigma}_t}$$

- Jarque-Bera test for normality and Ljung-Box test for autocorrelation

Model	Jarque-Bera test	Ljung-Box test (20 squared lags)
ARCH(1)	0.00	0.00
ARCH(4)	0.00	0.00
GARCH(4,1)	0.00	0.99
GARCH(1,1)	0.00	0.53

Alternative Volatility Models

Realised Volatility

- Based on taking intraday data, sampled at regular intervals and using them to obtain covariance matrix
- Pros
  - Purely data driven and no reliance on parametric methods
- Cons
  - Intraday data must be available, data is difficult to obtain, hard to use, and not very clean, and expensive
  - Intraday data has complicated patterns that have to be addressed

Implied Volatility

- Black-Scholes (BS) equation

$$\text{Option price} = BS(T, r, S, X, \sigma)$$

- Volatility implied by the price

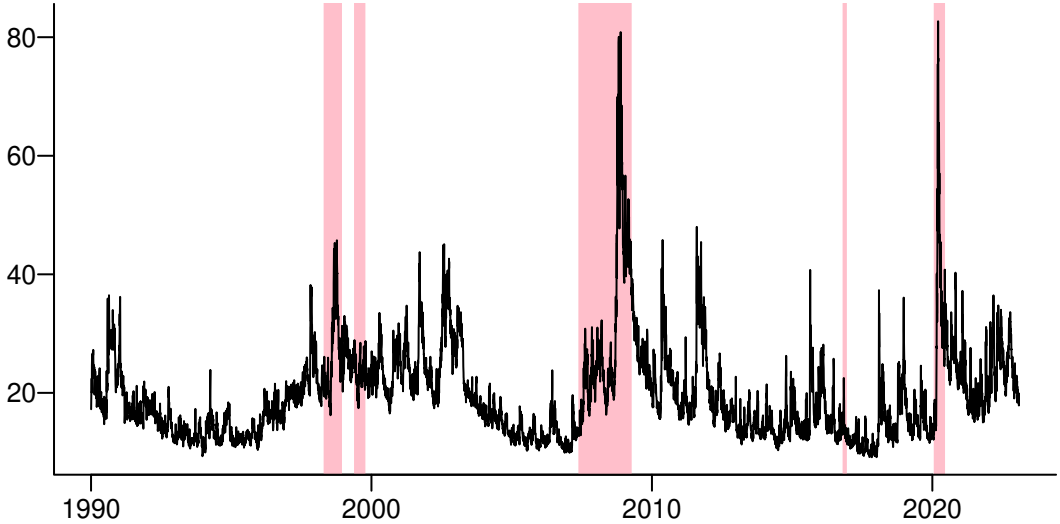
$$\text{Implied volatility} = BS^{-1}(T, r, S, X, \text{option price})$$

- Based on current market prices and not historical data
- Depends critically on accuracy of BS model (constant volatility and normal innovations)

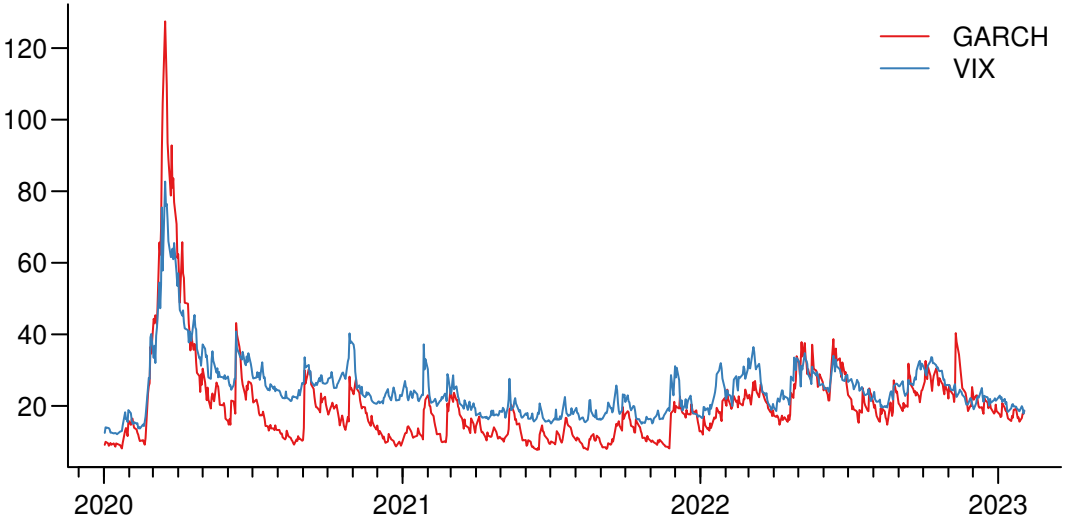
VIX

- From the Chicago Board Options Exchange (CBOE)
- Volatility index – VIX
- It captures the annualised one month volatility of the S&P-500
- Similar to implied volatilities
- It is known as the *"fear index"* of the financial markets
- And even has inspired a thriller by Robert Harris with the same title
- The following figure shows the VIX along with key market events

VIX



S&P-500 GARCH and VIX



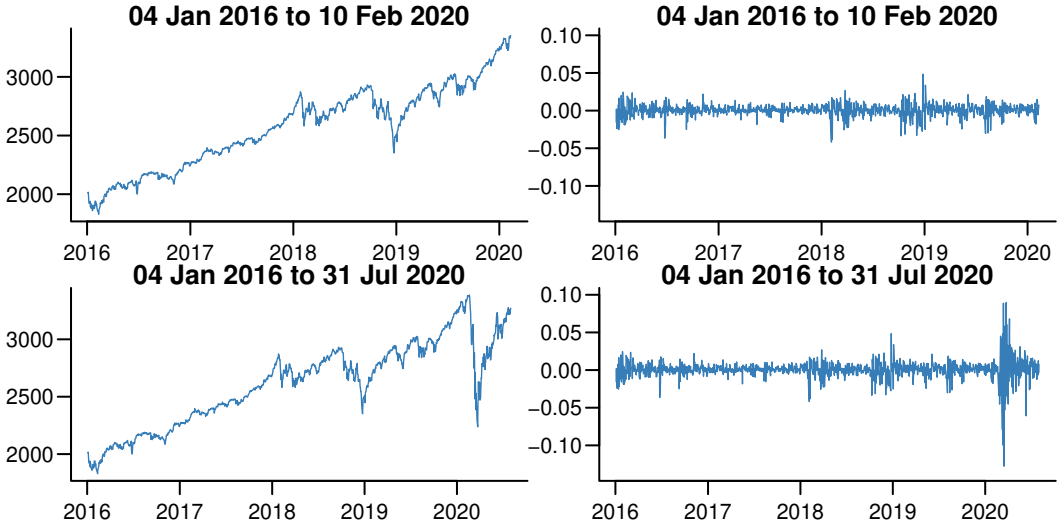
Summary

Summary

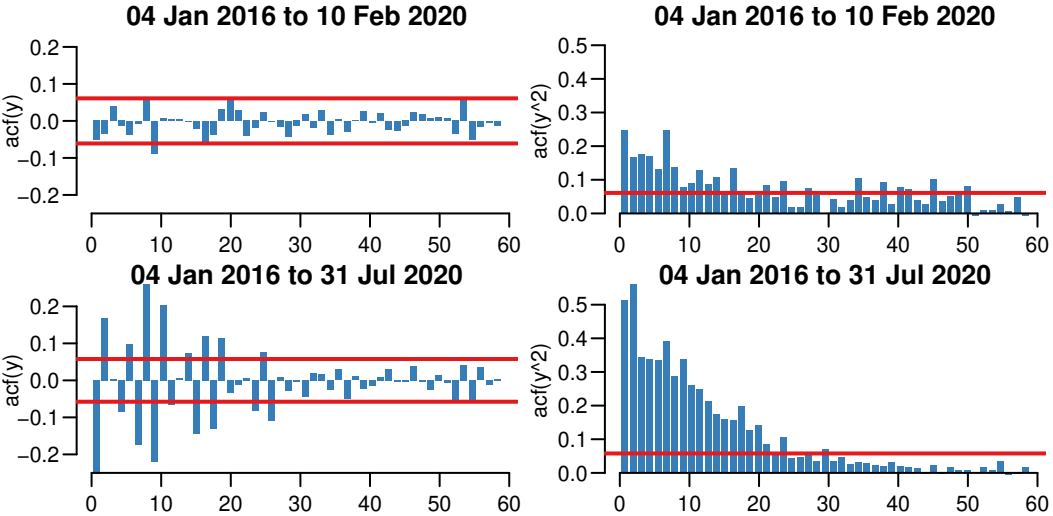
- It is not possible to measure volatility
- Instead, it has to be inferred from observed data
- Using some sort of model
- And that means there are multiple alternative ways to measure volatility
- And it can be very hard to discriminate between them

Cases: Covid-19

S&P-500 With and Without 2020



S&P-500 ACF With and Without 2020



GARCH Analysis

- Consider the normal and student to GARCH
- When we include and exclude the crisis period
- The outcome is exactly what we would expect

S&P-500

2016-02-22 to 2020-02-10

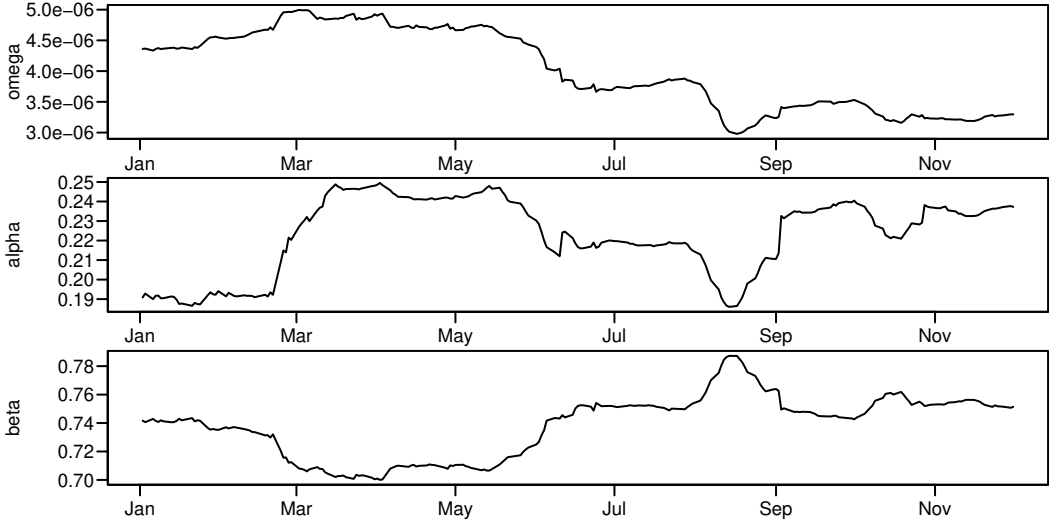
ω	α	β	ν	$\alpha+\beta$	Half
4.6e-06	0.19	0.73		0.93	9
2.7e-06	0.19	0.79	4.24	0.97	27

2016-07-14 to 2020-07-02

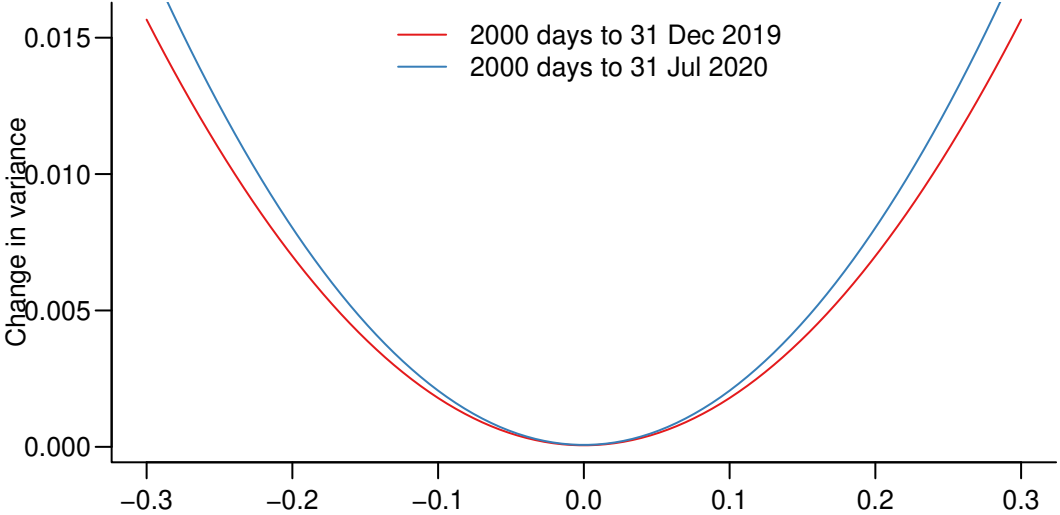
ω	α	β	ν	$\alpha+\beta$	Half
3.7e-06	0.22	0.75		0.97	23
2.1e-06	0.22	0.80	3.86	1.02	NA

Why do you think the half-life is NA for the Student-t?

S&P-500 GARCH Coefficients in 2020



S&P-500 News Impact (GARCH)



S&P-500 News Impact (APGARCH)